Universität Hamburg

Dissertation

Kernel-Based Generalized Interpolation and its Application to Computerized Tomography

Dissertation zur Erlangung des Doktorgrades
an der
Fakultät für Mathematik, Informatik und Naturwissenschaften

Fachbereich Mathematik
der Universität Hamburg

vorgelegt von
Kristof Albrecht

Hamburg, 2024

Gutachter
Prof. Dr. Armin Iske (Erstgutachter)
Prof. Dr. Marko Lindner (Zweitgutachter)

Datum der Disputation: 06.12.2024

Abstract

Throughout recent decades, positive definite kernel functions have turned out to be powerful and flexible approximation tools for several mathematical problems and their associated real-world applications. Besides the interpolation of Lagrangian data, kernel functions are well-suited to solve more general interpolation problems concerning the evaluation of arbitrary functionals, also known as generalized interpolation. Although the treatment of the generalized case is straightforward in many aspects, it has not gained the same attention as the standard interpolation case yet. So far, the research on this topic has mainly focused on the solution of partial differential equations, and further applications such as medical imaging are rather exotic.

In 2018, the authors De Marchi, Iske and Santin proposed the application of generalized interpolation to the field of computerized tomography in combination with weighted kernel functions. Inspired by this work, the objective of this thesis is to further elaborate the concept of generalized interpolation and investigate its utility for the reconstruction of images from scattered Radon data. To this end, we derive an extensive framework for treating generalized interpolation problems in the first part of this thesis, which includes data-dependent orthonormal systems, greedy data selection algorithms and suitable regularization methods. We provide convergence results under mild assumptions on the considered kernel and translate several results from standard Lagrangian interpolation to the generalized case.

The derived framework is then applied to the problem of computerized tomography in the second part, where we derive useful properties of weighted kernel functions. By choosing suitable kernels, we can guarantee the well-posedness of the reconstruction method and therefore make use of our general theoretical results from the first part. Moreover, we provide theoretical and numerical comparisons to other well-established reconstruction methods to demonstrate the advantages of kernel-based generalized interpolation in the field of computerized tomography.

Kurzfassung

Im Laufe der letzten Jahrzehnte haben sich positiv-definite Kernfunktionen als überaus nützlich herausgestellt für verschiedenste mathematische Probleme und die zugehörigen praktischen Anwendungen. Neben der Interpolation von gewöhnlichen Lagrange-Daten sind Kernfunktionen ebenfalls prädestiniert für die Interpolation bezüglich der Auswertung von beliebigen Funktionalen, auch bekannt als verallgemeinerte Interpolation. Auch wenn sich die Analyse des verallgemeinerten Problems in vielen Punkten nicht vom Standardfall unterscheidet, hat dieses Forschungsthema bisher noch keine große Aufmerksamkeit erlangt. Während sich der Großteil der zugehörigen Forschung mit der Lösung von partiellen Differentialgleichungen beschäftigt, existieren nur wenige Arbeiten über andere Anwendungsfelder wie die medizinische Bildgebung.

Eines dieser seltenen Anwendungsbeispiele ist Computertomographie, genauer gesagt die Rekonstruktion von Bildern mittels gewichteter Kernfunktionen, welches aus einer Publikation von den Autoren De Marchi, Iske und Santin von 2018 hervorging. Ausgehend von diesem Forschungsartikel ist es das Ziel dieser Doktorarbeit, das Konzept der Kern-basierten verallgemeinerten Interpolation weiter zu analysieren und die Nützlichkeit bezüglich der Rekonstruktion anhand verstreuter Radon-Daten weiter zu untersuchen. Dazu leiten wir im ersten Teil dieser Arbeit ein umfangreiches Grundgerüst für die Lösung von verallgemeinerten Interpolationsproblemen mittels positiv-definiter Kernfunktionen her, bestehend aus Daten-abhängigen Orthonormalsystemen, Greedy-Algorithmen zur Auswahl von Interpolationspunkten und geeigneten Regularisierungsmethoden. Wir beweisen die Konvergenz der Interpolationsmethode unter milden Voraussetzungen an die Kernfunktion und verallgemeinern diverse Resultate aus der Standardtheorie.

Im zweiten Teil dieser Arbeit wenden wir dieses Grundgerüst auf das zugrundeliegende mathematische Problem der Computertomographie an. Wir beweisen essenzielle Eigenschaften von gewichteten Kernfunktionen, welche die Wohldefiniertheit der Rekonstruktionsmethode sicherstellen. Insbesondere wird sichergestellt, dass wir die Resultate aus dem ersten Teil dieser Arbeit anwenden können. Darüber hinaus stellen wir einen theoretischen und numerischen Vergleich mit anderen etablierten Rekonstruktionsmethoden an, um die Vorteile der Kern-basierten verallgemeinerten Interpolation zu demonstrieren.

Acknowledgements

First and foremost, I want to thank my family Waltraud, Thomas and Jan-Hendrik for their unconditional support throughout my whole studies. Without them, none of this would have been possible. In the same way, I want to thank my partner Helen for always keeping me motivated and enduring the struggles of my PhD time.

Of course, I want to thank my supervisors Armin Iske and Marko Lindner for their support throughout my PhD studies. In particular, I want to thank Armin for encouraging me to continue my academic path and Marko for being one of the kindest bosses one can probably meet in working life.

Special thanks goes to Juliane Entzian, who has been an excellent research partner in the last few years. I really enjoyed the numerous research meetings and lunch breaks. Moreover, I would like to thank Nico Albers, Janek Becker, Matthias Beckmann and Dennis Schmeckpeper for proofreading the initial draft of my thesis. Your comments and suggestions notably helped to improve the final version.

Generally, I want to thank my colleagues at the Institute of Mathematics (TUHH) and the members of the RTG 2583 for welcoming me to their team with open arms. I always appreciated the friendly and supportive atmosphere.

Many thanks also go to my friends who always manage to make life more enjoyable.

Lastly, I acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG) within the Research Training Group GRK2583 "Modeling, Simulation and Optimization of Fluid Dynamic Applications".

Contents

1. Introduction

Positive (semi-)definite kernel functions (in short: kernel functions or kernels) are powerful and flexible tools for multivariate approximation problems. Regarding their mathematical theory, kernel functions were already investigated in the early 1900s in the context of integral operators. A prominent example of this time is *Mercer's theorem* (cf. [94]), which guarantees a representation of a kernel in terms of the eigenvalues and eigenfunctions of its associated integral operator. Nearly three decades later, the construction of translation-invariant kernels was investigated by Salomon Bochner in the 1930s (cf. [23], [24]), followed by a characterization of radial symmetric kernels due to the work of Isaac J. Schoenberg (cf. [121]). Another milestone in kernel-based theory is given by the *Moore-Aronszajn theorem*, which established the one-to-one correspondence between kernel functions and *reproducing kernel Hilbert spaces* that have useful additional properties in comparison to general Hilbert spaces. Based on the work of Eliakim H. Moore (cf. [96], [97]), the theorem was stated and proven by Nachman Aronszajn in his pioneering papers [11], [12]. The aforementioned theoretical breakthroughs have led to the development of various kernel-based approximation methods over the last decades. As a consequence, kernel functions have gained popularity in numerous mathematical fields, e.g. partial differential equations (cf. [47], [79], [136], [146]), statistics (cf. [22], [137]) and machine learning (cf. [26], [123], [132]).

One of the best-known applications of kernel functions is the interpolation of multivariate functions on scattered interpolation point sets. Here, the term *scattered* means that we do not assume any structure on the set of interpolation points, e.g. the points do not necessarily lie on a grid. The traditional idea of interpolation can be described in the following way: Given samples $f(x)$ of a function $f \in F$ at points $x \in X \subset \Omega$, where F is a normed space of real-valued functions on $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, we wish to approximately reconstruct the function f on its whole domain Ω. To treat this problem, it is common to fix a suitable space $S \subset F$ as the search space for the reconstruction and solve for an element $s \in S$ that satisfies

$$f(x) = s(x) \qquad \text{for all } x \in X. \tag{1.1}$$

Kernel-based solution methods for the interpolation problem (1.1) have been studied extensively in the past decades, and there is plenty of literature available (see, e.g., [27], [32], [70], [143]). However, there is a generalization to this interpolation problem that has not gained the same attention yet. If δ_x denotes the *point evaluation functional* with respect to a point $x \in \Omega$ on F, i.e.

$$\delta_x : F \to \mathbb{R},\ f \mapsto f(x) \qquad \text{for } x \in \Omega,$$

the problem (1.1) can be rewritten as

$$\delta_x(f) = \delta_x(s) \qquad \text{for all } x \in X.$$

Hence, the previously stated interpolation problem, which we call *standard interpolation problem*, consists of interpolating the evaluations of f with respect to the point evaluation functionals $\{\delta_x \mid x \in X\}$. With that, we can think about substituting the point evaluation functionals with other linear functionals

$$\lambda_i : F \to \mathbb{R} \qquad i = 1, ..., n, \quad n \in \mathbb{N},$$

leading to the main subject of this thesis, the *generalized interpolation problem*

$$\lambda_i(f) = \lambda_i(s) \qquad \text{for } i = 1, ..., n.$$

There has already been some research on this topic, where most of it is motivated by solving partial differential equations via collocation (cf. [47], [53], [54], [118], [146]). To this end, the involved differential operator is discretized, leading to a finite collection of functionals in the dual space. However, this generalized interpolation approach can also be applied to other types of linear operators. Only recently,

there have been some advances in the field of tomography (cf. [38], [39], [126], [148]). In particular, the authors of [38] proposed a novel kernel-based reconstruction method for *computerized tomography*. Similar to other kernel-based methods, this reconstruction method, which can be classified as an *algebraic reconstruction method*, is characterized by its high flexibility, as it is well-suited to process *scattered Radon data*. This could be advantageous in situations with severe data limitations that occur in practical applications of computerized tomography (cf. [38, page 4]).

The paper [38] has mainly motivated the content of this thesis and its three superordinate goals, which can be summarized as follows:

(1) **Generalization:** The book chapter [143, Chapter 16] provides a good initial discussion on kernel-based generalized interpolation, where it is pointed out that the generalized interpolation case can be treated in nearly the same way as the standard interpolation case (cf. [143, page 306]). Despite that, many theoretical results from standard interpolation have not been „translated“ to the generalized case yet. This mainly involves convergence criteria and stability estimates, where we want to work with generalized versions of the well-known *fill distance* and *power function* (cf. [116]). We aim to close as many of these gaps as possible so that standard interpolation can indeed be interpreted as just a special case of generalized interpolation. Moreover, we aim at providing a generalized discussion on data-dependent bases such as the *Newton basis* (cf. [98], [99], [104], [105] for the standard interpolation case) and *greedy data selection algorithms* (cf. [41], [43], [119], [145] for the standard case), which has already been done in [118], [146] to some degree. Here, *data-dependent* means that the basis depends on the considered set of functionals for the interpolation problem. Note that the generalization of data-dependent bases has been teased in [104, page 34, line 5-6], representing the main idea behind the first part of this thesis:

„There is a generalization of this chapter to interpolation of this form,
but we leave this for future work.“

We want to verify this claim, which applies to many other parts of standard interpolation theory as well, and thus take on the declared future work.

(2) **Unification:** In addition to the idea of generalization, the first part of this thesis aims to provide a broad list of tools for kernel-based interpolation that have been developed in the last two decades, such as the aforementioned data-dependent bases and greedy selection algorithms. For the analysis of these tools, we do not want to restrict to a special type of kernel or setting and rather discuss the generalized interpolation method on a meta-level with very general assumptions on the kernels and the interpolation domains. Of course, these assumptions later have to be verified for a given application in order to apply our developed theory. The ultimate goal of the first part is to provide an extensive framework for kernel-based generalized interpolation, which can be applied directly to many different problems.

(3) **Application to computerized tomography:** The content of the first part is primarily motivated by the main application in this thesis, which is the application of kernel-based generalized interpolation to *computerized tomography* as proposed in [38]. Although the main focus of this reconstruction method was rather laid on its flexibility, it yielded competitive results in comparison with the well-known *method of filtered back projection* on regular line distributions, see [38, Section 5.2]. But, in contrast to the promising numerical part, the theoretical part of this paper did not properly explain how this method fits into the setting of kernel-based generalized interpolation. In our work, we highlight and try to correct these critical points to show the well-posedness of this reconstruction method. Moreover, we want to extend this method by the framework provided in the first part of this thesis, see point (2) above. For example, the incorporation of greedy data selection algorithms would allow us to filter out the most relevant Radon data for the reconstruction process. We want to show that our derived theory is applicable to this reconstruction method and provide an extended numerical investigation so that we can evaluate its practical use in computerized tomography.

Outline

To follow along these goals, the thesis is divided into two parts. While the first part explains the general framework of kernel-based generalized interpolation, the second part describes the application of the derived framework to computerized tomography and provides numerical examples. The separation of the two parts is crucial in our eyes since it underlines the fact that the derived framework is not only applicable to computerized tomography but also to other linear operator equations. We stress this fact throughout the first part by providing other short examples of applications.

Within these two parts, the thesis is organized as follows:

In Chapter 2, we introduce positive (semi-)definite kernel functions, which build the core of kernel-based approximation theory. We list relevant theoretical properties and, most importantly, describe how these functions can be constructed. In addition to the characterizations of *Bochner* and *Schoenberg*, we discuss two special examples that demonstrate the wide range of kernel functions, i.e. *Wendland's compactly supported kernels* and *product kernels*.

As we have already pointed out in the introductory text, one important component of kernel-based approximation theory is the fact that each kernel function is associated with a unique *reproducing kernel Hilbert space*, also called the *native space* of the considered kernel (cf. [117]). Therefore, we explain the construction of the native space and highlight its special structural properties in Chapter 3. Concerning generalized interpolation problems, we show how to determine the Riesz representers of linear bounded functionals and explain how properties of the considered kernel are inherited by the functions of its native space.

With these theoretical insights, we can properly introduce the concept of *kernel-based generalized interpolation* in Chapter 4, where we highlight the main advantages over the general Hilbert space setting. To ensure the well-posedness of these interpolation problems, it is required that the considered functionals are linearly independent. Hence, we discuss the *linear independence* for some specific settings that have already occurred in the literature. Under the assumption that the considered set of functionals is linearly independent, we proceed to prove the most relevant properties of the resulting interpolation operator.

In Chapter 5, we analyze the convergence of the generalized interpolation method when successively adding new functionals to the interpolation problem. To this end, we introduce the generalized version of the *power function* and fill distance, which are well-established error measurements in standard interpolation theory. We derive convergence criteria regarding these two measurements and, by introducing the concept of *parametrizations*, link these results to basic convergence results from standard interpolation theory.

In the standard interpolation setting, it is well-known that naively using the standard basis, which consists of the Riesz representers of the considered functionals, brings along numerical problems, i.e. ill-conditioning and a lack of update formulas for the interpolation process. Therefore, we discuss the construction of alternative bases in Chapter 6 that can diminish these weaknesses. Most importantly, we analyze a generalized version of the *Newton basis*, which was initially developed for the standard interpolation case (cf. [99], [105]) and yields a more stable and efficient interpolation process.

Another critical aspect of kernel-based interpolation is the choice of functionals, also referred to as *data points*, which we discuss in Chapter 7. In [145], the authors were able to unify several known *greedy data selection algorithms* from standard interpolation with the concept of *β-greedy algorithms*, which was then translated to the generalized case in [146]. We prove the convergence of these algorithms under rather mild assumptions and analyze generalized *geometric* approaches for the selection of new data points. One practical application of these greedy algorithms is to *thin out* large, redundant sets of functionals. We show how this idea can be combined with regularization tools from kernel-based approximation.

The second part of this thesis starts with an introduction to *computerized tomography* in Chapter 8, where we formulate the underlying mathematical problem and its main mathematical operator, the *Radon transform*. We explain the explicit reconstruction formula, known as the *filtered back projection (FBP) formula*, and highlight its practical limitations. Regarding the stability, we briefly discuss the *ill-posedness* of the reconstruction problem as well.

Due to its limitations, the FBP formula cannot be applied directly in practice, which means that

approximation methods are required. In Chapter 9, we discuss suitable methods for the reconstruction from finitely many Radon values. Our focus mainly lies on the *method of filtered back projection* and *algebraic reconstruction methods*, which are two well-established methods in the field of computerized tomography. To round out this chapter, we provide a list of further reconstruction methods that have been developed in the more recent past.

In Chapter 10, we finally discuss the interpolation of Radon data. As proposed in [38], we introduce and analyze *weighted kernel functions*, which enable us to apply the framework of kernel-based generalized interpolation. Our analysis results in conditions on the weight function that guarantee the well-posedness of the reconstruction method. In particular, we can apply the convergence results and interpolation tools to this particular reconstruction problem for suitable weight functions. As a further reference for our numerical tests, we briefly describe a non-symmetric reconstruction approach using standard translation-invariant kernels.

To get an impression of its performance, we describe and evaluate numerical tests regarding the kernel-based reconstruction method in Chapter 11. On the one hand, we compare the different greedy selection strategies in terms of approximation quality and stability. This comparison gives an insight into which selection algorithms are well-suited for the selection of significant data points for the reconstruction. On the other hand, we compare the generalized interpolation method to the FBP method, where we mainly focus on the number of data measurements required to obtain a suitable reconstruction. In addition to the numerical comparisons, we demonstrate the effect of regularization tools on the reconstruction quality.

Lastly, we summarize the ideas and results of this thesis in Chapter 12. This results in an outlook on real-world applications of the derived kernel-based reconstruction method for computerized tomography and possible applications to other linear inverse problems. We end the final chapter by listing the remaining open problems of our investigations.

As a supplement to the main parts of this thesis, we provide further details on mathematical tools in Appendix A that we used in our analysis. These include the theory of *Fourier transforms*, *distributions* and *Sobolev spaces*. Since most of the presented content is well-known, we only state relevant theorems and refer to other works for respective proofs.

Conventions and notations

For this thesis, we expect the reader to have fundamental knowledge in *real and complex analysis*, *linear algebra*, *numerical analysis* and *functional analysis*. Before we start our analysis, we want to declare some basic notations and comment on conventions that we follow in our work:

- Let (Z, d_Z) be a metric space. For given center $a \in Z$ and radius $R > 0$, we define the corresponding `open ball` as

  $$B_R(a) := \{z \in Z \mid d_Z(z, a) < R\}.$$

  We denote the `closure` of a subset $A \subset Z$ as $\overline{A}$ and its `interior` as $\mathring{A}$.

- For the canonical space $\mathbb{R}^d$ with dimension $d \in \mathbb{N}$, we denote the `standard (Euclidean) norm` as

  $$\|x\|_2 := \left(\sum_{i=1}^{d} x_i^2 \right)^{1/2} \qquad \text{for } x = (x_1, ..., x_d) \in \mathbb{R}^d,$$

  which is induced by the `standard inner product`

  $$\langle x, y \rangle_2 := \sum_{i=1}^{d} x_i \cdot y_i \qquad \text{for } x, y \in \mathbb{R}^d.$$

- Let $\Omega \subset \mathbb{R}^d$ be open. For $k \in \mathbb{N}_0$, the `space of` k`-times continuously differentiable functions` on Ω is defined as

  $$\mathscr{C}^k(\Omega) := \left\{ f : \Omega \to \mathbb{R} \;\middle|\; \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} \text{ exists and is continuous for all } \alpha \in \mathbb{N}_0^d \text{ with } |\alpha| \leq k \right\},$$

where $|\alpha| := \sum_{i=1}^{d} \alpha_i$ for $\alpha \in \mathbb{N}_0^d$. For $d = 1$, we use $f^{(k)}$ to denote the k-th derivative of f. The space of infinitely differentiable functions is denoted as

$$\mathscr{C}^\infty(\Omega) := \bigcap_{k \in \mathbb{N}_0} \mathscr{C}^k(\Omega).$$

- Let $\Omega \subset \mathbb{R}^d$ and $p \in [1, \infty]$. The `Lebesgue space of p-integrable functions` is given by

$$L^p(\Omega) = \left\{ [f : \Omega \to \mathbb{R}] \;\middle|\; f \text{ is measurable on } \Omega \text{ and } \|f\|_p < \infty \right\}$$

Formally, this space consists of equivalence classes, where two functions are equivalent if they coincide almost everywhere on Ω and the norm is given by

$$\|f\|_p := \begin{cases} \left(\int\limits_\Omega |f(x)|^p \, dx \right)^{1/p}, & \text{for } p \in [1, \infty) \\ \inf\limits_{\substack{N \subset \Omega \\ \mu(N)=0}} \sup\limits_{x \in \Omega \setminus N} |f(x)|, & \text{for } p = \infty, \end{cases}$$

where the integral/infimum is taken with respect to the Lebesgue measure μ. The `essential support` of a function $f \in L^p(\Omega)$ is given by

$$\operatorname{ess\,supp}(f) := \Omega \setminus \bigcup \left\{ N \subset \Omega \;\middle|\; N \text{ is open and } f(x) = 0 \text{ for almost every } x \in N \right\}.$$

If f is continuous, this coincides with the usual definition of the support, i.e.

$$\operatorname{supp}(f) := \overline{\left\{ x \in \Omega \;\middle|\; f(x) \neq 0 \right\}},$$

so that we will only write $\operatorname{supp}(f)$ instead of $\operatorname{ess\,supp}(f)$ for $f \in L^p(\Omega)$ as well.

- In the literature, one can find different notations for the native space of a positive semi-definite kernel function (cf. Chapter 3). For example, the book [143] uses the notation $\mathcal{N}_\Phi(\mathbb{R}^d)$, whereas [70] uses $\mathcal{F}$. In our work, we follow the Hilbert space notation of [22] and denote the native space as $\mathcal{H}_K$ for a given kernel function K. Moreover, there is no complete agreement on the terminology of positive (semi-)definite functions. While our definition (cf. Definition 2.5) aligns with the respective definitions for matrices, other authors refer to these functions as *positive type functions* (cf. [22, Definition 2]) or (strictly) positive definite functions [32, Chapter 12 & 13]. In the latter case, one has to be cautious when comparing the results of different works.

- Throughout our work, we deal with interpolation point sets such as

$$X = \{x_1, ..., x_n\} \subset \mathbb{R}^d.$$

The numbering of the elements indicates that we care about the ordering within the sets, e.g. when defining the *interpolation matrix* with respect to a given point set (cf. Definition 2.5). Moreover, we care about duplicates in the sets, i.e. whether the points are pairwise distinct. These considerations are rather untypical for sets, as the order of the elements or the number of occurrences usually does not matter in set theory. Instead, one should rather interpret the sets as tuples, i.e.

$$X = (x_1, ..., x_n) \in \mathbb{R}^d \times ... \times \mathbb{R}^d.$$

However, since this slight abuse of notation should not cause any confusion and we want to use common operations for sets, we stick to the set notation as it is done in the literature.

- Based on the book chapter [143, Chapter 16], we call interpolation with respect to arbitrary functionals from the dual space *generalized interpolation*. In other works, this problem is also referred to as *generalized Hermite-Birkhoff interpolation*, e.g. in [68]. This name is derived from the term *Hermite-Birkhoff interpolation*, which usually stands for the interpolation with respect to point evaluations and partial derivatives (see, e.g., [151]).

Contributions to the research field

Before we start with the main part, we want to highlight the results of this thesis that represent new contributions to the research field of kernel-based approximation. In this way, we avoid meta-discussions on these results within our analysis that would interrupt the flow of reading in many cases.

- **Product kernels:** Initially, we wish to point out that the work on product kernels is a collaboration with Juliane Entzian and Armin Iske. The main new result is Theorem 2.20, where we show that the product of positive definite functions on lower-dimensional spaces is again positive definite. As we remark in the text before the theorem, this result was already proven for translation-invariant kernels using Bochner's characterization in [143, Proposition 6.25]. In this case, one can simply use that the Fourier transform of the generating function is the product of lower-dimensional Fourier transforms. However, the general case of arbitrary positive definite kernels is more complicated. To solve this problem, we make use of *grid-like data point sets*. This concept provides a workaround to prove the positive definiteness of the resulting product kernel (cf. Theorem 2.18, Corollary 2.19), which is essential for the well-posedness of interpolation problems. In addition, we provide a more up-to-date analysis of the native spaces of product kernels in Section 3.3. A comparison with the earlier works [12, Chapter I.8] of Aronszajn and [103, Chapter VI] of Neveu is given in Remark 3.21. Note that the joint work on product kernels has already been uploaded as the preprint

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

  which also elaborates further computational advantages in comparison to standard kernels when dealing with grid-like data sets.

- **Linear independence:** Our discussion on the linear independence of functionals in Section 4.1 is based on the works [68] and [151], where the authors derived useful integral representations of the inner product in the native dual space for translation-invariant kernels. We provide a more detailed verification of this integral representation for compactly supported distributions in this thesis. In particular, we demonstrate how this representation links to the setting of generalized interpolation as introduced in Chapter 4 under suitable conditions, see Lemma 4.4. Thus, we can shift the analysis to the well-known space of distributions (cf. Theorem 4.5), which was already stated in [68, Theorem 6.2]. We provide a new linear independence result for *cell average functionals* under suitable conditions in Theorem 4.7. In particular, our unified analysis covers the positive definiteness results of [129, Chapter 7]. Due to our result, linear independence assumptions as in [1, Section 3] can be omitted.

- **Convergence criteria:** For our convergence analysis, we focus on generalized versions of the *power function* and the *fill distance*. Note that the generalized version of the power function was already introduced and analyzed in [53], [54]. Given a fixed set of considered functionals, we prove in Theorem 5.6 that the generalized interpolation method converges on the whole induced subspace if and only if the power function converges pointwise to zero. Moreover, we show that the generalized interpolation method converges if the fill distance with respect to the dual space norm converges to zero, see Theorem 5.9. In order to build a connection to the convergence analysis for the standard interpolation case, we introduce the notion of *parametrizations* for subsets of functionals in Section 5.3. With that, standard convergence results such as [70, Theorem 8.37] can easily be derived from our theory. Parametrizations facilitate the convergence analysis and therefore play a key role in our analysis of the kernel-based reconstruction method for computerized tomography in Subsection 10.2.2. Large parts of the aforementioned results are included in the preprint

  [8] K. Albrecht and A. Iske. On the convergence of generalized kernel-based interpolation by greedy data selection algorithms. arXiv preprint: 2407.03840, 2024.

- **Data-dependent bases:** The content of Chapter 6 is basically a translation of [70, Subsection 8.1.2 & 8.2.1], [99] and [105] to the generalized interpolation case. Although the generalization is straightforward and, in our eyes, does not deserve a lot of credit, it has to be verified that well-known data-dependent bases can be used in the generalized case as well. In particular, the analysis of the *(generalized) Newton basis* rounds out the framework for generalized interpolation derived in

the first part of this thesis. We want to point out here that the generalized version of the Newton basis was already used in [118] and [146], but the authors only provided brief explanations.

- **Greedy algorithms:** In Subsection 7.2.1, we introduce the *geometric greedy algorithm* in the dual and parameter space, which are generalizations of the geometric approach in [37]. The convergence of the two algorithms is proven in Theorem 7.11 and Theorem 7.15 for totally bounded domains. Note that this assumption on the interpolation domain is motivated by Proposition 7.6. At this point, we have to admit that identifying totally bounded sets of functionals is not a trivial task in general. However, auxiliary results can simplify the identification (cf. Lemma A.4). It turns out that we mostly deal with totally bounded domains in applications, see e.g. Example 5.15 or Subsection 10.2.2. In addition to the geometric approaches, we prove convergence results for the *β-greedy algorithms* that were introduced in [145] and generalized in [146]. The normwise convergence of these algorithms and their vectorial versions (cf. Subsection 7.2.4) is proven in Theorem 7.19 and Theorem 7.21 for totally bounded domains. We remark that, in contrast to our analysis, the works [145], [146] provide convergence rates for the single-target β-greedy algorithms, but in a more restrictive setting. Since the underlying mathematical problem of our main application *computerized tomography* is ill-posed (cf. Section 8.3), we add some basic regularization tools to our framework for generalized interpolation. More precisely, we provide a generalization of [70, Section 8.6], where we substitute the native space norm regularization with more general regularization functionals from the dual space. Initial applications of greedy algorithms to computerized tomography without further analysis have been published in

  [7] K. Albrecht and A. Iske. Greedy algorithms for image approximation from scattered Radon data. PAMM, 21(1):e202100223, 2021.

  Further numerical examples as well as the convergence proofs for the greedy algorithms are included in the preprint [8]. Moreover, the present author has contributed to the proceedings paper

  [5] K. Albrecht, J. Entzian, and A. Iske. Anisotropic Kernels for Particle Flow Simulation. In A. Iske and T. Rung, editors, Modeling, Simulation and Optimization of Fluid Dynamic Applications, Lecture Notes in Computational Science and Engineering: Volume 148, pages 57–76. Springer, 2023

  by applying greedy algorithms to the respective standard interpolation problems (cf. [5, Section 4.6.2]) for the numerical comparison in [5, Section 4.6.4].

- **Kernel-based computerized tomography:** The most notable contributions are given in Chapter 10, where we analyze the kernel-based reconstruction method for computerized tomography proposed in [38]. To establish the well-posedness of the method, the authors introduced *weighted kernel functions*, which were further analyzed in the master thesis [57]. In our analysis of weighted kernel functions, we add to the results of [57]. For example, we improve the characterization of the native space (cf. [57, Theorem 3.4], Theorem 10.7) and provide conversion formulas for the Newton basis in the standard interpolation case (cf. Proposition 10.11). Due to our results on weighted kernels, we can prove that the functionals associated with the image reconstruction problem are indeed elements of the native dual space, see Proposition 10.15. Note that this has not been proven properly in [38] (cf. introduction of Chapter 10). Throughout Section 10.2, we demonstrate that the reconstruction method benefits from our derived framework for generalized interpolation, in particular from the convergence analysis and the interpolation tools. Furthermore, we add some ideas in Section 10.3 that are specifically adapted to computerized tomography.

Finally, we remark that the present author has already investigated data-dependent bases and greedy selection algorithms for the standard interpolation case in the master thesis

[4] K. Albrecht. Orthogonalsysteme in kernbasierten Approximationsräumen. Master's thesis, Universität Hamburg, 2019,

which has inspired the content of Chapter 5, 6 & 7.

Part I.

Generalized Interpolation in Reproducing Kernel Hilbert Spaces

In the first part, we explain and analyze the concept of generalized interpolation using positive semi-definite kernel functions, which bring along several advantages in comparison to the general Hilbert space setting due to their native reproducing kernel Hilbert space. For the introduction, we want to provide a timeline of previous research on this topic that summarizes the developments. Of course, this timeline is not complete.

One of the earlier works on partial differential equations in the context of kernel-based approximation is given by the paper [79] of Kansa, which introduced an unsymmetric approach by modeling the interpolant via the usual basis functions known from standard interpolation, also known as *Kansa's method*. In [151], Wu analyzed the well-posedness of *Hermite-Birkhoff interpolation* using kernel functions, where the basis functions were chosen as the Riesz representers of the considered functionals, and thus adapted to the given problem. The dependence of the basis functions on the differential operator leads to symmetric, positive definite interpolation matrices that ensure unique solutions to the interpolation problem. Based on this symmetric approach, Fasshauer proposed a solution method for partial differential equations via collocation in [47], which was later analyzed by Franke and Schaback (cf. [53], [54]). The analysis relies on a transformed version of the kernel due to the application of the considered differential operator that allows one to interpret the problem as a standard interpolation problem.

Besides the particular treatment of differential operators, there are some works on the interpolation with respect to more general subspaces of *compactly supported distributions*. In [100], Narcowich and Ward analyzed the interpolation setting for matrix-valued kernel functions to reconstruct vector fields, where the considered functionals were induced by certain multi-component distributions with compact support. The last part of this paper is mainly focused on differential operators again. Moreover, Iske discussed the topologization of the space of compactly supported distributions with order less or equal to a fixed $l \in \mathbb{N}$ in the kernel setting (cf. [68]), resulting in an integral representation of the native inner product that simplifies the identification of linearly independent functionals. This integral representation was already stated in [151] and is also found in [143, Theorem 16.4] for the special case of partial differential operators. Generally, the chapter [143, Chapter 16] of Wendland's prominent book *Scattered Data Approximation* provides a good initial discussion on the generalized interpolation problem and further elaborates the analysis of [53].

Another application of kernel-based generalized interpolation is given by the interpolation with respect to *cell average functionals*. In [128], Sonar explained how this type of functional occurs in *finite volume methods* for hyperbolic conservation laws and proposed the use of thin-plate splines for optimal recovery within an *ENO approach*. As a follow-up, the utility of kernel functions in this specific setting was further elaborated in [71] by Iske and Sonar as well as in the paper [130] by Sonar. Missing convergence rates were later provided by Wendland in the paper [142] for Sobolev kernels and thin-plate splines. Moreover, the authors Aboiyar, Georgoulis and Iske discussed a kernel-based *WENO approach* in [1], which yields a weighted overall reconstruction from cell average values.

Although the research on generalized interpolation has mainly focused on partial differential equations so far, there have been a few advances in the field of *tomography*. In his master thesis [126], Sironi discussed the kernel-based reconstruction from *scattered Radon data*. There, it was already observed that a naive application of the Radon transform to a Gaussian kernel leads to non-finite diagonal entries in the interpolation matrix. To circumvent this problem, the second application of the Radon transform was modified by multiplication with a weight function to ensure the integrability, see also [39]. The

modification leads to unsymmetric interpolation matrices, similar to the aforementioned method of Kansa, and therefore does not fit into the usual concept of kernel-based generalized interpolation. The problems were ultimately fixed in [38] with the introduction of *weighted kernel functions* by De Marchi, Iske and Santin, which is a symmetric application of the weight function to both sides of a given kernel. In addition to this application in *computerized tomography*, the master thesis [148] discussed the kernel-based reconstruction from *spherical Radon data* in *photoacoustic tomography*. It was pointed out by the author Wiatkowski that the computation of the Riesz representers and the inner products in the dual space is very tricky for those types of operators (cf. [148, page 23]). As a consequence, only the unsymmetric approach was numerically investigated.

The above-mentioned works regarding tomography have in common that none of them incorporated further tools known from standard interpolation such as *data-dependent bases*, *greedy data selection algorithms* or *regularization tools*, although most of these concepts can be transferred to the generalized case straightforward. A generalization of the well-known *P-greedy algorithm* (cf. [37], [41] for the standard case) was proposed in the preprint [118] by Schaback, which was recently followed by the preprint [146] of Wenzel, Winkle, Santin and Haasdonk on a larger class of greedy algorithms, the so-called *β-greedy algorithms*. Similar to the ideas of [118], the paper [146] yielded convergence results in terms of the L^∞-norm for *second-order elliptic boundary value problems* for *Sobolev kernels*. Moreover, both works shortly described the generalization of the *Newton basis* from the standard interpolation case (cf. [99], [105]).

Based on the previous research on this topic, the main objective of this part is to provide an extensive framework for generalized interpolation in reproducing kernel Hilbert spaces that brings together relevant theoretical results and useful tools to improve the reconstruction method. To this end, we want to keep our discussion and results as general as possible, so that the framework can be applied directly to many different problems. In particular, this is important in the second part of this thesis when we apply this framework in the context of computerized tomography, which is the main application in this work. Although our main focus lies on computerized tomography, we provide some examples that demonstrate how our derived theory links to the other applications mentioned before.

2. Positive (Semi-)Definite Kernel Functions

We begin this part with the introduction of *positive (semi-)definite kernel functions*, or in short, *kernel functions*. Since these functions build the core of our approximation algorithms later on, we give a summary of their construction in Section 2.1 including several examples. Further basic properties are stated in Section 2.2. For a more detailed treatment of this topic, we refer to the books [32], [70] and [143], where most of the presented results can be found. Moreover, the book [50] provides a great overview of kernel functions and their applications.

In order to motivate the use of kernel functions, we consider the standard problem of *scattered data interpolation*. Therefore, let $\Omega \subset \mathbb{R}^d$ for $d \in \mathbb{N}$. For a given point set $X = \{x_1, ..., x_n\} \subset \Omega$ of length $n \in \mathbb{N}$ and corresponding data values $\{f_1, ..., f_n\} \subset \mathbb{R}$, we aim to find a function $s : \Omega \to \mathbb{R}$ that satisfies the interpolation condition

$$s(x_i) = f_i \qquad \text{for all } i \in \{1, ..., n\}. \tag{2.1}$$

A common approach for this problem is to fix a set of n basis functions $B = \{b_1, ..., b_n\}$, where $b_i : \Omega \to \mathbb{R}$ for all $i \in \{1, ..., n\}$, and restrict the interpolant to

$$s = \sum_{i=1}^{n} c_i \cdot b_i \in \operatorname{span}_{\mathbb{R}}\{b_1, ..., b_n\}. \tag{2.2}$$

If we insert the restriction (2.2) into the interpolation condition (2.1), we get the equivalent linear system

$$\begin{pmatrix} b_1(x_1) & \dots & b_n(x_1) \\ \vdots & \ddots & \vdots \\ b_1(x_n) & \dots & b_n(x_n) \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}. \tag{2.3}$$

It is well-known from linear algebra that the system (2.3) has a unique solution for all sets of values $\{f_1, ..., f_n\} \subset \mathbb{R}$ if and only if the *Vandermonde matrix* $V_{B,X} = (b_i(x_j))_{1 \le i,j \le n}$ is regular. The regularity of this matrix is guaranteed if B is a *Haar system* (cf. [58], [143, Definition 2.1]).

Definition 2.1. Let $S \subset \mathscr{C}(\Omega)$ be a linear space of continuous real-valued functions on a domain $\Omega \subset \mathbb{R}^d$ with dimension $\dim(S) = n \in \mathbb{N}$. Then S is called a `Haar space` on Ω if any $s \in S \setminus \{0\}$ has at most $n - 1$ zeros in the domain Ω. A basis B of a Haar space on Ω is called `Haar system` on Ω.

Example 2.2. We want to provide some examples of Haar systems for the case $\Omega = \mathbb{R}$.

(i) For any $n \in \mathbb{N}$, the set

$$\mathscr{P}_n = \left\{ p : \mathbb{R} \to \mathbb{R} \;\middle|\; p(x) = \sum_{i=0}^{k} a_i \cdot x^i \text{ for all } x \in \mathbb{R}, \text{ where } 0 \le k \le n \text{ and } a_0, ..., a_k \in \mathbb{R} \right\}$$

of (algebraic) polynomial functions with degree less than $n + 1$ form a Haar space of dimension $n + 1$, see [70, Example 5.26].

(ii) For $\nu \in \mathbb{R}$, consider the exponential function

$$f_\nu(x) := e^{\nu \cdot x} \qquad \text{for all } x \in \mathbb{R}.$$

If $\nu_1, ..., \nu_n \in \mathbb{R}$ are pairwise distinct parameters, then $B = \{f_{\nu_1}, ..., f_{\nu_n}\}$ is a Haar system on $\mathbb{R}$, see [70, Example 5.28].

(iii) Given a fixed parameter $\nu > 0$ and nodes $Z = \{z_1, ..., z_n\} \subset \mathbb{R}$, the respective scaled / shifted *Gaussian functions* are given by

$$G_{\nu,z_i}(x) := e^{-\nu \cdot (x - z_i)^2} \qquad \text{for all } x \in \mathbb{R},\ i \in \{1, ..., n\}.$$

If Z contains pairwise distinct points, $B_{\nu,Z} = \{G_{\nu,z_1}, ..., G_{\nu,z_n}\}$ is a Haar system. To see this, we rewrite a linear combination s of the functions $B_{\nu,Z}$ as

$$s(x) = \sum_{i=1}^{n} c_i \cdot G_{\nu,z_i}(x) = e^{-\nu \cdot x^2} \cdot \sum_{i=1}^{n} c_i e^{-\nu z_i^2} \cdot e^{2 z_i \cdot x} \qquad \text{for all } x \in \mathbb{R},$$

with coefficients $c_1, ..., c_n \in \mathbb{R}$. Now assume that s has at least n zeros in $\mathbb{R}$. Since the exponential function is strictly positive on $\mathbb{R}$, this implies that the exponential sum

$$\tilde{s}(x) = \sum_{i=1}^{n} c_i e^{-\nu z_i^2} \cdot e^{2 z_i \cdot x} \qquad \text{for all } x \in \mathbb{R}$$

has at least n zeros in $\mathbb{R}$. But the parameters of the exponential functions are pairwise distinct, so that the coefficients of $\tilde{s}$, and therefore all c_i, $i = 1, ..., n$, have to vanish according to part (ii). The result can also be found in [3, Theorem 3.5]. We come back to this example when we talk about kernel functions in Section 2.1.

If we assume that B is a Haar system, the kernel of the Vandermonde matrix $V_{B,X}$ is trivial for any $X \subset \Omega$ with $|X| = n$, which directly implies the regularity of $V_{B,X}$. Hence, the interpolation problem (2.1) has a unique solution.

Unfortunately, Haar systems suffer from severe limitations. Note that the domain of the functions from Example 2.2 is only one-dimensional. In fact, there are no Haar systems on relevant domains in the case $d \geq 2$. This is the result of the famous *Mairhuber-Curtis Theorem* [35], [91]. Here, we state the version from [143, Theorem 2.3].

Theorem 2.3 (Mairhuber-Curtis)**.** *Let $\Omega \subset \mathbb{R}^d$ with $d \geq 2$. If Ω contains an interior point, then there is no Haar space of dimension $n \geq 2$ on Ω.*

Remark 2.4. Although there are no Haar spaces in the multivariate case, unique interpolation might still be possible, e.g. in the case of multivariate polynomial interpolation on a grid (cf. [143, Lemma 2.8]). However, we want to focus on scattered data interpolation, where we don't make any assumptions on the structure of X.

Due to the Mairhuber-Curtis Theorem, a workaround is required in the case of multivariate interpolation on scattered data. One possible approach for this problem is to make the basis functions depend on the data sites, i.e. $b_i \equiv b_{x_i}$ for all $i \in \{1, ..., n\}$. To this end, we consider a (kernel) function

$$K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

and define the basis functions

$$b_i(x) := K(x, x_i) \qquad \text{for all } x \in \mathbb{R}^d,\ i \in \{1, ..., n\}.$$

If we insert this definition into (2.3), we get the following linear system:

$$\begin{pmatrix} K(x_1, x_1) & \dots & K(x_1, x_n) \\ \vdots & \ddots & \vdots \\ K(x_n, x_1) & \dots & K(x_n, x_n) \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$$

Again, the regularity of the system matrix determines whether the system is well-posed or not. This leads us to the definition of *positive (semi-)definite kernel functions* (cf. [143, Definition 6.1]).

Definition 2.5. Let $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. We call K a `positive (semi-)definite kernel function`, if K is symmetric, i.e. $K(x, y) = K(y, x)$ for all $x, y \in \mathbb{R}^d$, and for any finite set $X = \{x_1, ..., x_n\} \subset \mathbb{R}^d$ of pairwise distinct points, the corresponding `interpolation matrix`

$$A_{K,X} := \begin{pmatrix} K(x_1, x_1) & \dots & K(x_1, x_n) \\ \vdots & \ddots & \vdots \\ K(x_n, x_1) & \dots & K(x_n, x_n) \end{pmatrix}$$

is positive (semi-)definite.

Our previous discussion shows that positive definite kernels, which are most relevant to our work, lead to well-posed interpolation problems. So far, it has not been important that the system matrices $A_{K,X}$ are not only regular but also symmetric and positive definite. In the next chapter, we make use of that when we dive deeper into the theory of kernel-based approximation.

Remark 2.6. Besides positive (semi-)definite kernels, there is also the concept of *conditionally positive (semi-)definite functions*, where the interpolant consists of a kernel part and a polynomial part. We do not cover these functions in this thesis but refer to [48, Chapter 7 ff.] and [143, Chapter 8 ff.] for detailed information.

2.1. Construction of positive definite kernel functions

The task of finding positive definite kernels is not trivial, as the requirement of $A_{K,X}$ being positive definite for all finite sets $X \subset \mathbb{R}^d$ seems to be very strong. Hence, we give a short overview of the construction of positive definite kernels. These include translation-invariant kernels, radial basis functions and more recent kernels that have been derived from the previous two classes.

We omit most of the proofs in this subsection since this is not the focus of this thesis and the respective proofs are very technical. More details and proofs can be found in [32, Chapter 12-16] and [143, Chapter 6-9].

2.1.1. Translation-invariant kernels

The first class of kernels that we want to present are *translation-invariant kernels*. Here, translation-invariant means that there is an even function $\Phi : \mathbb{R}^d \to \mathbb{R}$ such that

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

Note that for every translation-invariant kernel K and any fixed $z \in \mathbb{R}^d$, we have

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

which should explain the term 'translation-invariant' in this context. The analysis of multivariate translation-invariant kernels dates back to Bochner in the 1930s (cf. [23], [24]), who linked these kernels to the non-negativity of their Fourier transforms. Details on the Fourier transform are provided in Section A.3. We state the L^1-version of Bochner's characterization here (cf. [143, Theorem 6.11]).

Theorem 2.7 (Bochner's theorem)**.** *Let $\Phi \in L^1(\mathbb{R}^d) \cap \mathscr{C}(\mathbb{R}^d)$. Then $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by*

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

is a positive definite kernel function if and only if Φ is bounded and its Fourier transform $\mathcal{F}\Phi$ is non-negative and non-vanishing.

Example 2.8. Due to Bochner's theorem, the following kernels are positive definite:

(i) For every shape parameter $\nu > 0$, the respective *Gaussian kernel* (cf. Figure 2.1)

$$K(x, y) := e^{-\nu \cdot \|x-y\|_2^2} \qquad \text{for all } x, y \in \mathbb{R}^d,$$

where $\|x\|_2$ denotes the standard euclidean norm, is positive definite on $\mathbb{R}^d$ for all dimensions $d \in \mathbb{N}$. This is due to the fact that the Fourier transform of the Gaussian function

$$\Phi(x) := e^{-\nu \cdot \|x\|_2^2} \qquad \text{for all } x \in \mathbb{R}^d$$

is again a Gaussian function (cf. Example A.12), and therefore positive on the whole domain. Note that we have already seen this function in Example 2.2 for $d = 1$, but this time, the nodes coincide with the interpolation points.

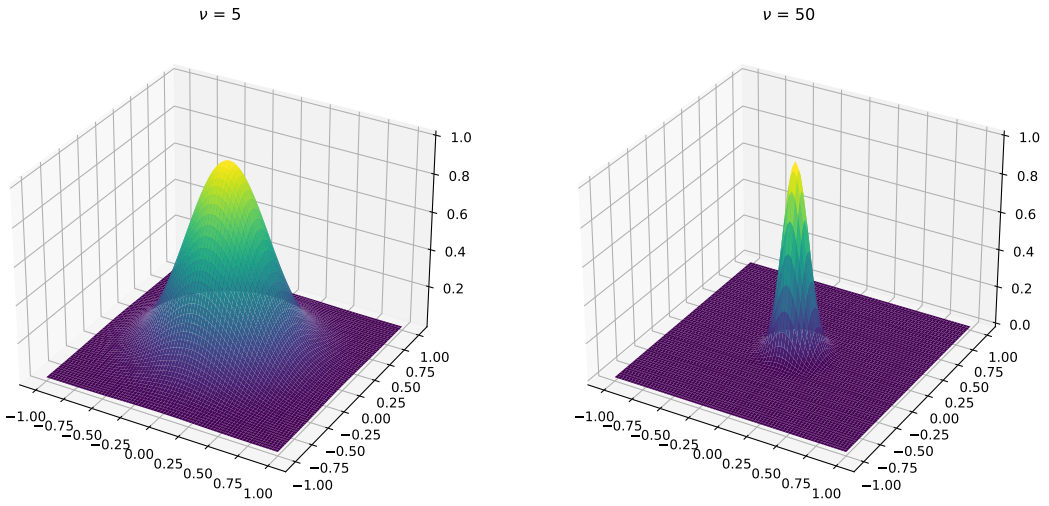

Figure 2.1.: Plot of two-dimensional Gaussian with shape parameter $\nu = 5$ (left) and $\nu = 50$ (right)

(ii) The *radial characteristic kernel* of Askey (cf. [13]), defined by

$$K(x, y) := (1 - \|x - y\|_2)_+^l := \begin{cases} (1 - \|x - y\|_2)^l, & \text{for } \|x - y\|_2 \leq 1 \\ 0, & \text{otherwise,} \end{cases}$$

is positive definite on $\mathbb{R}^d$ for $l \geq \lfloor d/2 \rfloor + 1$ (cf. [143, Theorem 6.20]).

2.1.2. Radial basis functions

Another popular class of kernels is formed by rotation-invariant kernels. We call the function K a radial kernel, if there is a univariate function $\phi : [0, \infty) \to \mathbb{R}$, such that

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

The function ϕ is often called *radial basis function*. Note that all functions from Example 2.8 belong to this class, and every radial kernel is also translation-invariant. Furthermore, if $Q \in \mathbb{R}^{d \times d}$ is an orthogonal matrix, we have the equation

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

A characterization of functions ϕ that generate kernels in every dimension $d \in \mathbb{N}$ was given by Schoenberg in 1938 (cf. [121]). This characterization associates radial basis functions with *completely monotone functions* (cf. [143, Definition 7.1 & Theorem 7.14]).

Definition 2.9. Let $\phi : [0, \infty) \to \mathbb{R}$ be a univariate function. We call ϕ `completely monotone` on $[0, \infty)$, if $\phi \in \mathscr{C}([0, \infty)) \cap \mathscr{C}^\infty((0, \infty))$ and

$$(-1)^l \cdot \phi^{(l)}(r) \geq 0 \qquad \text{for all } r \in (0, \infty)\,,\ l \in \mathbb{N}_0,$$

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

Theorem 2.10 (Schoenberg)**.** *For $\phi : [0, \infty) \to \mathbb{R}$, the following two properties are equivalent:*

(i) For every dimension $d \in \mathbb{N}$, the kernel function

$$K(x, y) := \phi(\|x - y\|_2) \qquad \text{for all } x, y \in \mathbb{R}^d$$

is positive definite on $\mathbb{R}^d$.

(ii) The function

$$\tilde{\phi} : [0, \infty) \to \mathbb{R},\ r \mapsto \phi\left(\sqrt{r}\right)$$

is completely monotone and not constant on $[0, \infty)$.

Example 2.11. For fixed $\nu > 0$, consider the *inverse multiquadrics*

$$\phi(r) := (1 + r^2)^{-\nu} \qquad \text{for all } r \in [0, \infty)\,.$$

Due to the estimate

$$(-1)^l \cdot \tilde{\phi}^{(l)}(r) = (-1)^{2l} \cdot \prod_{k=1}^{l-1} (\nu + k) \cdot (1 + r)^{-\nu - l} \geq 0$$

for any $r \in [0, \infty)$ (cf. [143, Theorem 7.15]), the kernel

$$K(x, y) := (1 + \|x - y\|_2^2)^{-\nu} \qquad \text{for all } x, y \in \mathbb{R}^d$$

is positive definite on $\mathbb{R}^d$ for any dimension $d \in \mathbb{N}$ according to Theorem 2.10.

2.1.3. Other kernels

Of course, we are not able to cover all types of kernels in this thesis. A broad list of positive definite kernel functions is provided in the book [50]. Instead, we want to give two more examples to briefly demonstrate the wide range and versatility of kernels.

Wendland's compactly supported functions

From the computational point of view, it is desirable to use radial basis functions with compact support, as these lead to sparse interpolation matrices if there is sufficient separation between the interpolation points. Some examples are given by the radial characteristic kernels from Example 2.8 (ii), which are generated by the functions

$$\phi_l(r) := (1 - |r|)_+^l := \begin{cases} (1 - |r|)^l & \text{for } |r| \leq 1 \\ 0, & \text{otherwise} \end{cases} \qquad \text{for } l \in \mathbb{N}, \tag{2.4}$$

where we evenly extended the original functions to the whole real line. Note that we had to make the restriction $l \geq \lfloor d/2 \rfloor + 1$ to guarantee that the resulting kernel is positive definite on $\mathbb{R}^d$. In contrast to functions like the Gaussians or the inverse multiquadrics, there are no continuous radial basis functions that have compact support and generate a positive definite kernel on $\mathbb{R}^d$ for every dimension $d \in \mathbb{N}$ (cf. [143, Theorem 9.2]). Hence, the construction of these functions is always dependent on the considered dimension.

As desired, the functions ϕ_l for $l \in \mathbb{N}$ have compact support, but they are not differentiable at the origin $r = 0$. This means that the basis functions $K(\cdot, x)$ for $x \in \mathbb{R}^d$ of the resulting kernel, and therefore all resulting interpolants are not differentiable as well, see Figure 2.2. However, regularity assumptions like differentiability are usually required in approximation theory to achieve high-order convergence. As a possible solution, Wendland constructed compactly supported functions with arbitrary smoothness in [140], often called *Wendland's functions*.

These functions are even piecewise polynomial functions of the form

$$\phi(r) := \begin{cases} p(|r|), & \text{for } |r| \leq 1 \\ 0, & \text{otherwise,} \end{cases} \tag{2.5}$$

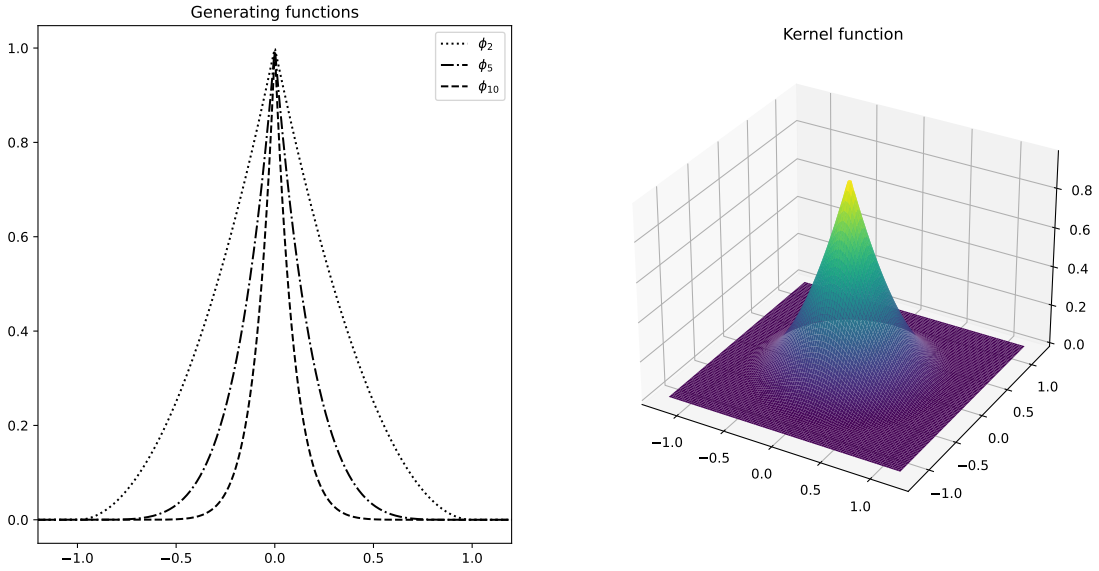

Figure 2.2.: Plot of generating functions ϕ_l for $l = 2, 5, 10$ (left) and resulting basis function $\phi_2(\|\cdot\|_2)$ on the two-dimensional domain $\mathbb{R}^2$ (right)

where $p : \mathbb{R} \to \mathbb{R}$ is a univariate polynomial. To optimize the computational costs, the degree of the polynomial p should be reduced to a minimum, such that the desired smoothness is still attained. In the following, we summarize the construction of Wendland's functions. Further explanations and proofs can be found in [143, Chapter 9].

The construction is based on the integral operator $\mathcal{I}$ defined as

$$(\mathcal{I}\phi)(r) = \int_{|r|}^{\infty} t \cdot \phi(t)\, dt \qquad \text{for } r \in \mathbb{R}$$

for functions $\phi : \mathbb{R} \to \mathbb{R}$ such that the mapping $t \mapsto t \cdot \phi(t)$ is in $L^1([0, \infty))$ (cf. [143, Definition 9.4]). If ϕ has the form (2.5), then $\mathcal{I}\phi$ has the same form, but the degree of the polynomial and the number of continuous derivatives is increased by 2 (cf. [143, Lemma 9.8]). With this, we define the functions

$$\phi_{d,k} = \mathcal{I}^k \phi_{\lfloor d/2 \rfloor + k + 1} \qquad \text{for } d, k \in \mathbb{N}, \tag{2.6}$$

where $\mathcal{I}$ is applied iteratively k times to the function $\phi_{\lfloor d/2 \rfloor + k + 1}$ defined in (2.4). Hence, the function $\phi_{d,k}$ is piecewise polynomial with degree $\lfloor d/2 \rfloor + 3k + 1$ and $2k$ continuous derivatives. Due to the Fourier properties of $\mathcal{I}$, $\phi_{d,k}$ generates a positive definite kernel on $\mathbb{R}^d$, and it turns out that $\lfloor d/2 \rfloor + 3k + 1$ is the minimal degree for a generating function of the form (2.5) to possess at least $2k$ continuous derivatives (cf. [143, Theorem 9.13]).

Theorem 2.12. *Let $d, k \in \mathbb{N}$ and $\phi_{d,k}$ be defined as in (2.6). Then, the induced function*

$$K(x, y) := \phi_{d,k}(\|x - y\|_2) \qquad \text{for } x, y \in \mathbb{R}^d$$

is a positive definite kernel on $\mathbb{R}^d$. Moreover, we have $\phi_{d,k} \in \mathscr{C}^{2k}(\mathbb{R})$, and $\phi_{d,k}$ has the form

$$\phi_{d,k}(r) = \begin{cases} p_{d,k}(|r|), & \text{for } |r| \leq 1 \\ 0, & \text{otherwise,} \end{cases}$$

where $p_{d,k}$ is a univariate polynomial of degree $\lfloor d/2 \rfloor + 3k + 1$. If $\phi \in \mathscr{C}^{2k}(\mathbb{R})$ is another function of the form (2.5) that generates a positive definite kernel on $\mathbb{R}^d$, then we have $\phi = c \cdot \phi_{d,k}$ for a constant $c > 0$.

Example 2.13. For $d = 2$, the first three Wendland functions are given by

$$\begin{aligned}\phi_{2,0}(r) &= (1-|r|)_+^2 \\ \phi_{2,1}(r) &= (1-|r|)_+^4 \cdot (4\cdot|r|+1) \\ \phi_{2,2}(r) &= (1-|r|)_+^6 \cdot \left(35\cdot|r|^2+18\cdot|r|+3\right)\end{aligned}$$

for $r \in \mathbb{R}$, see [143, Corollary 9.14]. A visualization of the functions and the resulting smooth kernel is given in Figure 2.3.

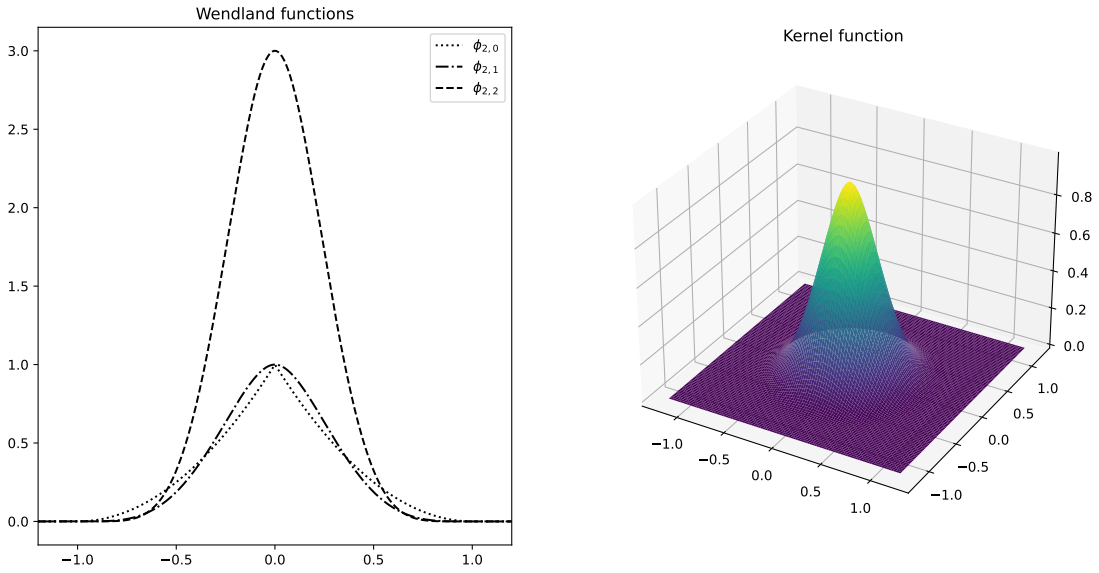

Figure 2.3.: Plot of Wendland functions (left) and plot of basis function $\phi_{2,1}(\|\cdot\|_2)$ (right)

Anisotropic product kernels

We can even use common kernel functions to construct new kernels via elementary operations. Here, we want to discuss the concept of *product kernels*, which was already introduced in the works [12, Section I.8], [103, Chapter VI]. In particular, we present the results of the preprint [6], which further elaborates on the properties of product kernels and their computational advantages.

The main idea of anisotropic product kernels is to separate the domain $\mathbb{R}^d$ into several axes and equip each axis with an individual kernel. In mathematical terms, we write

$$\mathbb{R}^d \simeq \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_M}$$

to split $\mathbb{R}^d$ into $M \in \mathbb{N}$ axes $\mathbb{R}^{d_i}$, where $d_i \in \mathbb{N}$ for $i = 1, ..., M$ and $d = \sum_{i=1}^M d_i$. In this context, we denote the projection onto the i-th axis as

$$p_i : \mathbb{R}^d \simeq \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_M} \to \mathbb{R}^{d_i},\ x = \left(x^{(1)}, ..., x^{(M)}\right) \mapsto x^{(i)}$$

for every $i \in \{1, ..., M\}$. If the functions $K_i : \mathbb{R}^{d_i} \times \mathbb{R}^{d_i} \to \mathbb{R}$ are positive semi-definite kernels on their respective axis, the product is a positive semi-definite kernel function on $\mathbb{R}^d$. This is a simple consequence of the *Schur product theorem* (see, e.g., [64, Theorem 5.2.1]).

Theorem 2.14. *Let $K_i : \mathbb{R}^{d_i} \times \mathbb{R}^{d_i} \to \mathbb{R}$ be positive semi-definite on $\mathbb{R}^{d_i}$ for $i = 1, ..., M$, where $d_i \in \mathbb{N}$. Set $d = \sum_{i=1}^M d_i$ and define the function*

$$K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R},\ (x, y) \mapsto \prod_{i=1}^M K_i(p_i(x), p_i(y)). \tag{2.7}$$

Then K is a positive semi-definite kernel function on $\mathbb{R}^d$.

The previous theorem motivates the definition of product kernels.

Definition 2.15. In the setting of Theorem 2.14, we call the function K defined in (2.7) the `product kernel` of the K_i, $i = 1, ..., M$, and use the notation $K = \prod_{i=1}^{M} K_i$. The kernel functions K_i are called the `components` of K.

Example 2.16. Consider the one-dimensional Wendland kernels

$$K_1(x, y) := (1 - |x - y|)_+$$
$$K_2(x, y) := (1 - |x - y|)_+^3 \cdot (3 \cdot |x - y| + 1)$$

for $x, y \in \mathbb{R}$ with different smoothness properties (cf. [143, Corollary 9.14]). The respective product kernel on $\mathbb{R}^2$ is given by

$$K(x, y) := (1 - |p_1(x) - p_1(y)|)_+ \cdot (1 - |p_2(x) - p_2(y)|)_+^3 \cdot (3 \cdot |p_2(x) - p_2(y)| + 1) \qquad \text{for } x, y \in \mathbb{R}^2.$$

As already pointed out, the main motivation of product kernels is the possibility to equip each axis with an individual kernel, depending on the application and its underlying data point set. In this example, the smoothness of the kernel varies for each axis of the domain. The one-dimensional kernels K_1, K_2 and the resulting product kernel K are visualized in Figure 2.4.

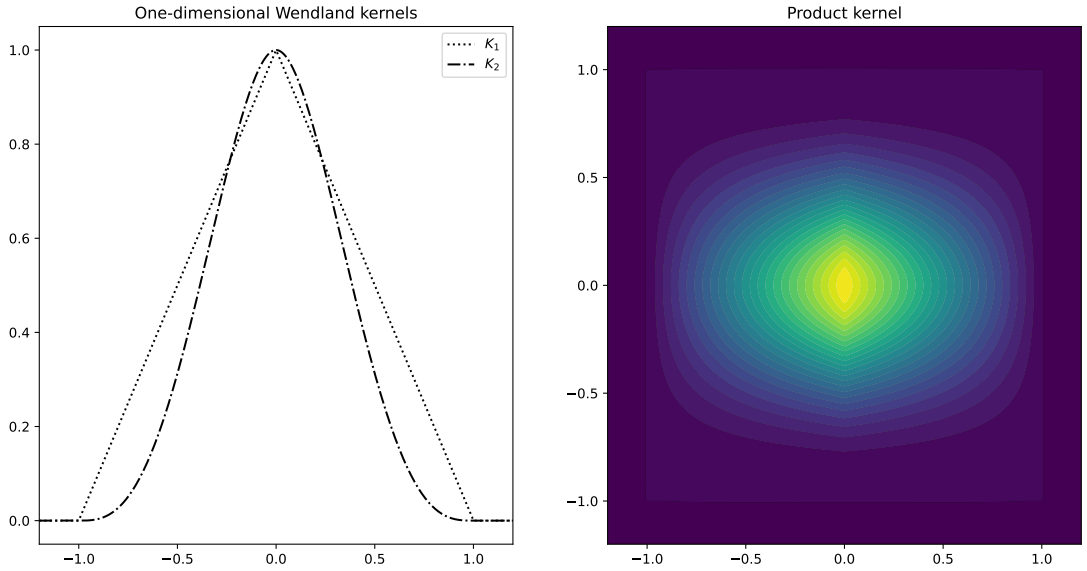

Figure 2.4.: Plot of basis functions $K_1(\cdot, x)$, $K_2(\cdot, x)$ for one-dimensional Wendland kernels with different smoothness at $x = 0$ (left) and contour plot of the respective product (right)

If $K = \prod_{i=1}^{M} K_i$ is a product kernel on $\mathbb{R}^d$ and $X = \{x_1, ..., x_n\} \subset \mathbb{R}^d$, the interpolation matrix $A_{K,X}$ is given by the *Hadamard product* (see, e.g., [64, Chapter 5]) of the components' interpolation matrices $A_{K_i, X^{(i)}}$, where

$$X^{(i)} = \{p_i(x_1), ..., p_i(x_n)\} \subset \mathbb{R}^{d_i} \qquad \text{for } i = 1, ..., M.$$

Even if the set X consists of pairwise distinct points, it is not guaranteed that the projected sets $X^{(i)}$ consist of pairwise distinct points as well. A simple counterexample is given by a finite grid in $\mathbb{R}^d$. Hence, the matrices do not have to be positive definite for positive definite K_i, which means that the Hadamard product is not suitable to show that the product kernel of positive definite components is again positive definite. In order to prove this assertion, we provide a workaround involving *grid-like point sets* and the *Kronecker product* for matrices (see, e.g., [64, Chapter 4.2]).

Definition 2.17. Let $X \subset \mathbb{R}^d \simeq \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_M}$ be finite. We call X a `grid-like point set`, if the set equality

$$X = p_1(X) \times ... \times p_M(X)$$

holds, i.e. X can be written as a Cartesian product of lower-dimensional subsets.

In the case of a grid-like point set, the matrix $A_{K,X}$ can be written as the Kronecker product of the component interpolation matrices. This result was already stated in [112, Section 5.2], but without precise proof.

Theorem 2.18. *Let $K = \prod_{i=1}^{M} K_i$ be a product kernel on $\mathbb{R}^d \simeq \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_M}$ and $X \subset \mathbb{R}^d$ be a grid-like point set of pairwise distinct points. Then we have*

$$A_{K,X} = \bigotimes_{i=1}^{M} A_{K_i,p_i(X)}, \tag{2.8}$$

where $\bigotimes$ denotes the Kronecker product and $p_i(X)$ only contains pairwise distinct points for $i = 1, ..., M$.

Proof. Let $N_i := |p_i(X)|$ for $i = 1, ..., M$ and $N := \prod_{i=1}^{M} N_i$. We prove the statement via induction on the number of components M.

<u>$M = 2$:</u> Let $p_1(X) = \left\{x_1^{(1)}, ..., x_{N_1}^{(1)}\right\}$ and $p_2(X) = \left\{x_1^{(2)}, ..., x_{N_2}^{(2)}\right\}$. For $k \in \{1, ..., N\}$, we set

$$x_k = \left(x_{\lceil k/N_2 \rceil}^{(1)},\ x_{(k-1) \bmod N_2+1}^{(2)}\right) \in X \subset \mathbb{R}^d,$$

where $\lceil \cdot \rceil$ denotes the ceiling function and *mod* denotes the remainder after division. This leads to an ordering $X = p_1(X) \times p_2(X) = \{x_k \mid k = 1, ..., N\}$ that results in

$$\begin{aligned}
(A_{K,X})_{j,k} &= K\left(\left(x_{\lceil j/N_2 \rceil}^{(1)},\ x_{(j-1) \bmod N_2+1}^{(2)}\right),\ \left(x_{\lceil k/N_2 \rceil}^{(1)},\ x_{(k-1) \bmod N_2+1}^{(2)}\right)\right) \\
&= K_1\left(x_{\lceil j/N_2 \rceil}^{(1)},\ x_{\lceil k/N_2 \rceil}^{(1)}\right) \cdot\ K_2\left(x_{(j-1) \bmod N_2+1}^{(2)},\ x_{(k-1) \bmod N_2+1}^{(2)}\right) \\
&= \left(A_{K_1,p_1(X)}\right)_{\lceil j/N_2 \rceil,\ \lceil k/N_2 \rceil} \cdot \left(A_{K_2,p_2(X)}\right)_{(j-1) \bmod N_2+1,\ (k-1) \bmod N_2+1} \\
&= \left(A_{K,p_1(X)} \otimes A_{K,p_2(X)}\right)_{j,k}
\end{aligned}$$

for any $j, k \in \{1, \ldots, N\}$. Hence, the equation $A_{K,X} = A_{K_1,p_1(X)} \otimes A_{K_2,p_2(X)}$ holds.

<u>$M \to M+1$:</u> Set $\tilde{X} = p_1(X) \times ... \times p_M(X)$, $\tilde{K} = \prod_{i=1}^{M} K_i$ and $\tilde{N} = \prod_{i=1}^{M} N_i$. Due to the induction hypothesis, there is an ordering $\tilde{X} = \{\tilde{x}_l \mid l = 1, ..., \tilde{N}\}$ such that the equation

$$A_{\tilde{K},\tilde{X}} = \bigotimes_{i=1}^{M} A_{K_i,p_i(X)}$$

holds. As in the case $M = 2$, we can order $X = \tilde{X} \times p_{M+1}(X)$ to get

$$A_{K,X} = A_{\tilde{K},\tilde{X}} \otimes A_{K_{M+1},p_{M+1}(X)}.$$

In total, we conclude

$$A_{K,X} = \bigotimes_{i=1}^{M+1} A_{K_i,p_i(X)}.$$

□

Corollary 2.19. *In the setting of Theorem 2.18, let K_i be positive definite on $\mathbb{R}^{d_i}$ for $i = 1, ..., M$. Then $A_{K,X}$ is positive definite.*

Proof. This directly follows from Theorem 2.18, as the Kronecker product of positive definite matrices is again positive definite (see, e.g., [64, Theorem 4.2.12]). □

With the previous two results, we are able to show that a product kernel inherits the positive definiteness of its components. Note that this result was already proven in [143, Proposition 6.25] for translation-invariant kernels, but not for general kernel functions.

Theorem 2.20. *Let $K = \prod_{i=1}^{M} K_i$ be a product kernel on $\mathbb{R}^d \simeq \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_M}$ such that K_i is positive definite on $\mathbb{R}^{d_i}$ for $i = 1, ..., M$. Then K is positive definite on $\mathbb{R}^d$.*

Proof. Let $X \subset \mathbb{R}^d$ be a finite set of pairwise distinct points. We have

$$X \subset Y := p_1(X) \times ... \times p_M(X),$$

where we assume that $p_i(X) \subset \mathbb{R}^{d_i}$ consists of pairwise distinct points for $i = 1, ..., M$. According to Corollary 2.19, the matrix $A_{K,Y}$ is positive definite. Since $A_{K,X}$ is a submatrix of $A_{K,Y}$, it is also positive definite. □

Remark 2.21. Consider Gaussian components

$$K_i(p_i(x), p_i(y)) = e^{-\nu_i \cdot \|p_i(x) - p_i(y)\|_2^2} \qquad \text{for } x, y \in \mathbb{R}^d$$

with shape parameters $\nu_i > 0$ for $i = 1, ..., M$. The respective product kernel is given by

$$K(x,y) = \prod_{i=1}^{M} e^{-\nu_i \cdot \|p_i(x) - p_i(y)\|_2^2} = e^{-\sum_{i=1}^{M} \nu_i \cdot \|p_i(x) - p_i(y)\|_2^2} = e^{-(x-y)^T \cdot D \cdot (x-y)} \qquad \text{for } x, y \in \mathbb{R}^d,$$

where

$$D = \begin{pmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_M \end{pmatrix} \in \mathbb{R}^{d \times d} \qquad \text{with} \qquad D_i = \begin{pmatrix} \nu_i & & 0 \\ & \ddots & \\ 0 & & \nu_i \end{pmatrix} \in \mathbb{R}^{d_i \times d_i} \quad \text{for i = 1, ..., M}$$

is a diagonal matrix with the shape parameters on its diagonal (cf. [49, Example 2.5]). Hence, the resulting product kernel exchanges the standard Euclidean norm $\|\cdot\|_2$ with the anisotropic norm

$$\|\cdot\|_D : \mathbb{R}^d \to \mathbb{R},\ x \mapsto \sqrt{x^T \cdot D \cdot x}$$

inside the standard Gaussian kernel with shape parameter $\nu = 1$. We want to point out that the multiplication of lower-dimensional Gaussian kernels builds a bridge between product kernels and the insertion of anisotropic norms. The effects of exchanging the standard Euclidean norm with an anisotropic norm inside radial basis functions have been investigated in [5], [16], [30].

2.2. Basic properties

We close this chapter with some basic properties of kernel functions (cf. [70, page 279]) that we use throughout our analysis.

Proposition 2.22. *Let K be positive semi-definite on $\mathbb{R}^d$. The following statements hold:*

(1) For any $x \in \mathbb{R}^d$, we have $K(x,x) \geq 0$. If K is positive definite, then $K(x,x) > 0$.

(2) Given $x, y \in \mathbb{R}^d$, we can estimate $K(x,y)^2 \leq K(x,x) \cdot K(y,y)$.

(3) K is bounded on $\mathbb{R}^d \times \mathbb{R}^d$ if and only if K is bounded on the diagonal

$$D = \left\{ (x,x) \mid x \in \mathbb{R}^d \right\} \subset \mathbb{R}^d \times \mathbb{R}^d.$$

Proof. For (1), we set $X = \{x\}$ for a given $x \in \mathbb{R}^d$. Then we have $A_{K,X} = (K(x,x))$ so that the desired inequality follows from Definition 2.5. Similarly, we set $X = \{x, y\}$ for given $x, y \in \mathbb{R}^d$. Inequality (2) then follows from

$$0 \leq \det(A_{K,X}) = K(x,x) \cdot K(y,y) - K(x,y)^2.$$

Part (3) is a consequence of (2). □

Corollary 2.23. *If K is a translation-invariant kernel, then K is bounded on $\mathbb{R}^d \times \mathbb{R}^d$.*

Proof. In this case, we have $K(x,y) = \Phi(x-y)$ for $x, y \in \mathbb{R}^d$ and a function $\Phi : \mathbb{R}^d \to \mathbb{R}$. This leads to

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

which proves the claim according to part (3) of Proposition 2.22. □

Corollary 2.24. *If K is positive semi-definite on $\mathbb{R}^d$ and $K(x,x) = 0$ for all $x \in \mathbb{R}^d$, then we have $K(x,y) = 0$ for all $x, y \in \mathbb{R}^d$ as well.*

Proof. This immediately follows from part (2) of Proposition 2.22:

$$0 \leq K(x,y)^2 \leq K(x,x) \cdot K(y,y) = 0 \qquad \text{for all } x, y \in \mathbb{R}^d.$$

□

3. Reproducing Kernel Hilbert Spaces

In this chapter, we dive deeper into the theory of kernel-based approximation. So far, we have shown how to solve multivariate interpolation problems using positive definite functions, where the unique interpolants belonged to the spaces

$$S_{K,X} := \operatorname{span}_{\mathbb{R}} \Big\{ K(\cdot, x_1), ..., K(\cdot, x_n) \Big\} \tag{3.1}$$

for given point sets $X = \{x_1, ..., x_n\} \subset \mathbb{R}^d$. Our goal now is to construct a complete linear space $\mathcal{H}$ that is a superset of all spaces of type (3.1). This construction leads to special Hilbert spaces, so-called *reproducing kernel Hilbert spaces*, which have many useful properties. We explain these properties in detail throughout the next sections, which are based on [22, Chapter 1], [143, Chapter 10].

The first rigorous analysis of reproducing kernel Hilbert spaces in its modern terminology was published by Aronszajn in 1943/1950, see [11], [12]. Amongst many other results, the papers explain the one-to-one correspondence between positive (semi-)definite functions and reproducing kernels. Due to the earlier work of Moore (cf. [96], [97]), this result is called the *Moore-Aronszajn Theorem*, which is explained in Section 3.1. In addition, we discuss how the kernel inherits properties like differentiability or integrability to the functions of the associated Hilbert space of functions. To demonstrate the relevance of reproducing kernel Hilbert spaces, we list further characterization of the native space in Section 3.2 that build a connection to *Sobolev spaces*. In Section 3.3, we end this chapter with an excursion about the native spaces of product kernels, where we show that the native space even inherits the structure of the kernel.

Although our treatment of this topic might seem excessive to some experienced readers, we mostly provide detailed explanations throughout the chapter, even for well-known results. In order to analyze generalized interpolation problems, it is essential to have an in-depth understanding of reproducing kernel Hilbert spaces, in particular of its distinctive properties such as the simplified determination of Riesz representers of a linear bounded functional, which is explained in Theorem 3.12 and the following. We start this chapter with the basic terminology (cf. [143, Definition 10.1]).

Definition 3.1. Let $\mathcal{H} \,\widehat{=}\, (\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ be a Hilbert space of real functions $f : \mathbb{R}^d \to \mathbb{R}$. We call $\mathcal{H}$ a `reproducing kernel Hilbert space`, if there is a function $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ that satisfies the following two properties:

(i) For all $x \in \mathbb{R}^d$, we have $K(\cdot, x) \in \mathcal{H}$.

(ii) We have the reproduction property

$$f(x) = \langle f, K(\cdot, x) \rangle_{\mathcal{H}} \qquad \text{for all } f \in \mathcal{H},\ x \in \mathbb{R}^d. \tag{3.2}$$

The function K is called `reproducing kernel` of $\mathcal{H}$. If $\mathcal{H}$ is a reproducing kernel Hilbert space, we use the notation $\mathcal{H} \equiv \mathcal{H}_K$ to indicate that $\mathcal{H}$ has the reproducing kernel K.

At first glance, it might not be clear whether a Hilbert space $\mathcal{H}$ possesses a reproducing kernel or not. But there is a connection between the existence of a reproducing kernel and the point evaluation functionals in $\mathcal{H}$. For every $x \in \mathbb{R}^d$, the respective *point evaluation functional* or *Dirac functional* δ_x is given by

$$\delta_x : \mathcal{H} \to \mathbb{R},\ f \mapsto f(x).$$

A necessary and sufficient criterion for the existence of a reproducing kernel is the continuity of the point evaluation functionals (cf. [143, Theorem 10.2]). The result is based on an important theorem from functional analysis, the *Fréchet-Riesz representation theorem*, which plays a key role in generalized interpolation, see Chapter 4. For convenience, we state the representation theorem here. A proof can be found in [76, Theorem 2.3.1] for example.

Theorem 3.2 (Fréchet-Riesz representation theorem)**.** *Let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ be a real Hilbert space and $\mathcal{H}^*$ denote its dual space. Then, the mapping*

$$\varphi : \mathcal{H} \to \mathcal{H}^*,\ v \mapsto \langle\cdot, v\rangle_{\mathcal{H}}$$

is an isometric isomorphism.

Theorem 3.3. *Let $\mathcal{H}$ be a Hilbert space of real functions on $\mathbb{R}^d$. Then $\mathcal{H}$ is a reproducing kernel Hilbert space if and only if δ_x is continuous on $\mathcal{H}$ for all $x \in \mathbb{R}^d$.*

Proof. Let $\mathcal{H}$ have a reproducing kernel K. With the Cauchy-Schwarz inequality, we get the estimate

$$|\delta_x(f)| = |f(x)| = |\langle f, K(\cdot, x)\rangle_{\mathcal{H}}| \leq \|f\|_{\mathcal{H}} \cdot \|K(\cdot, x)\|_{\mathcal{H}} \qquad \text{for all } f \in \mathcal{H},\ x \in \mathbb{R}^d.$$

Hence, δ_x is continuous for all $x \in \mathbb{R}^d$. Conversely, let δ_x be continuous for every $x \in \mathbb{R}^d$. Due to Theorem 3.2 , there exist functions $k_x \in \mathcal{H}$, $x \in \mathbb{R}^d$, that satisfy

$$f(x) = \delta_x(f) = \langle f, k_x\rangle_{\mathcal{H}} \qquad \text{for all } f \in \mathcal{H}, x \in \mathbb{R}^d.$$

By construction, the function $K(x, y) := k_y(x)$ for $x, y \in \mathbb{R}^d$ is a reproducing kernel of $\mathcal{H}$. □

Remark 3.4. Note that the point evaluation functionals are not well-defined for any Hilbert space $\mathcal{H}$. One example is the space $L^2(\mathbb{R}^d)$, which consists of equivalence classes. Hence, we exclude these spaces from our following discussion or restrict ourselves to the continuous representatives of the equivalence classes.

Example 3.5. As a first example of a reproducing kernel Hilbert space, we consider the space

$$\mathcal{H}_L = \left\{ f \in \mathscr{C}(\mathbb{R}) \cap L^2(\mathbb{R}) \;\middle|\; \operatorname{supp}(\mathcal{F}f) \subset [-L, L] \right\}$$

of square integrable functions with bandwidth $L > 0$ (cf. [153]). Here, $\mathcal{F}$ denotes the Fourier transform on $L^2(\mathbb{R})$. We equip this space with the standard inner product on $L^2(\mathbb{R})$. For every $f \in \mathcal{H}_L$, we have $\mathcal{F}f \in L^2(\mathbb{R})$. Due to the compact support of $\mathcal{F}f$, we also have $\mathcal{F}f \in L^1(\mathbb{R})$, so that the Fourier inversion formula from Theorem A.15 in combination with isometry property from Theorem A.16 yields

$$f(x) = (2\pi)^{-1/2} \cdot \int_{-L}^{L} \mathcal{F}f(\omega) \cdot e^{i\omega x}\, d\omega = \left\langle f, (2\pi)^{-1/2} \cdot \mathcal{F}^{-1}\left[\chi_{[-L,L]}(\cdot) \cdot e^{-i(\cdot)x}\right] \right\rangle_{L^2(\mathbb{R})} \quad \text{for all } x \in \mathbb{R}^d,$$

where $\chi_{[-L,L]}$ denotes the characteristic function on $[-L, L]$. The reproducing kernel is then given by

$$K(x, y) = (2\pi)^{-1/2} \cdot \mathcal{F}^{-1}\left[\chi_{[-L,L]}(\cdot) \cdot e^{-i(\cdot)y}\right](x) = \frac{\sin(L \cdot (x - y))}{\pi \cdot (x - y)} \qquad \text{for } x, y \in \mathbb{R},$$

where the evaluation for the case $x = y$ is interpreted as a limit. Further examples of reproducing kernels and their respective Hilbert spaces can be found in [22, Section 6.4].

We want to collect some general properties of reproducing kernels and their respective Hilbert spaces, see also [22, Section 1.2].

Proposition 3.6. *Let $\mathcal{H}_K$ be a reproducing kernel Hilbert space. Then, the following statements hold:*

(1) The reproducing kernel K is positive semi-definite on $\mathbb{R}^d$.

(2) K is positive definite if and only if the functionals $\{\delta_x \mid x \in \mathbb{R}^d\}$ are linearly independent.

(3) The set

$$\mathcal{S}_K := \operatorname{span}_{\mathbb{R}} \left\{ K(\cdot, x) \;\middle|\; x \in \mathbb{R}^d \right\} \tag{3.3}$$

is dense in $\mathcal{H}_K$.

(4) The dual set

$$\mathcal{S}_K^{(*)} := span_{\mathbb{R}} \left\{ \delta_x \;\middle|\; x \in \mathbb{R}^d \right\}$$

is dense in the dual space $\mathcal{H}_K^*$.

(5) If $(f_n)_{n\in\mathbb{N}} \subset \mathcal{H}_K$ *converges normwise to* $f \in \mathcal{H}_K$*, then* $(f_n)_{n\in\mathbb{N}}$ *converges pointwise to* f.

Proof. The proof of the statements mainly relies on the reproduction property (ii) from Definition 3.1:

(1) The symmetry of K follows from

$$K(x,y) = \langle K(\cdot,y), K(\cdot,x)\rangle_{\mathcal{H}_K} = \langle K(\cdot,x), K(\cdot,y)\rangle_{\mathcal{H}_K} = K(y,x) \qquad \text{for all } x,y \in \mathbb{R}^d,$$

where we used the symmetry of the inner product. To prove the second part of the definition, let $X = \{x_1, ..., x_n\} \subset \mathbb{R}^d$ and $c = (c_1, ..., c_n)^T \in \mathbb{R}^n$. Then we have

$$\begin{aligned} c^T \cdot A_{K,X} \cdot c &= \sum_{i=1}^n \sum_{j=1}^n c_i \cdot c_j \cdot K(x_i, x_j) \\ &= \sum_{i=1}^n \sum_{j=1}^n c_i \cdot c_j \cdot \langle K(\cdot,x_i), K(\cdot,x_j)\rangle_{\mathcal{H}_K} \\ &= \left\| \sum_{i=1}^n c_i \cdot K(\cdot,x_i) \right\|_{\mathcal{H}_K}^2 \geq 0. \end{aligned}$$

Hence, K is positive semi-definite on $\mathbb{R}^d$.

(2) Due to our previous computations and Theorem 3.2, we have the identity

$$A_{K,X} = \left(\langle K(\cdot,x_i), K(\cdot,x_j)\rangle_{\mathcal{H}_K}\right)_{1\leq i,j\leq n} = \left(\langle \delta_{x_i}, \delta_{x_j}\rangle_{\mathcal{H}_K^*}\right)_{1\leq i,j\leq n},$$

since δ_{x_i} is the Riesz representer of $K(\cdot,x_i)$ for all $i \in \{1,...,n\}$. Hence, K is positive definite on $\mathbb{R}^d$ if and only if the Gram matrices $A_{K,X}$ are positive definite for all finite sets $X \subset \mathbb{R}^d$ of pairwise distinct points, which is equivalent to the linear independence of all point evaluation functionals.

(3) Due to Theorem A.9 part (2), we can write the space $\mathcal{H}_K$ as the orthogonal sum

$$\mathcal{H}_K = \overline{\mathcal{S}_K} \oplus \overline{\mathcal{S}_K}^{\perp},$$

where $\overline{\mathcal{S}_K}^{\perp}$ is the orthogonal complement of $\overline{\mathcal{S}_K}$. For every $f \in \overline{\mathcal{S}_K}^{\perp}$, we have

$$f(x) = \langle f, K(\cdot,x)\rangle_{\mathcal{H}_K} = 0 \qquad \text{for all } x \in \mathbb{R}^d,$$

as $K(\cdot,x) \in \overline{\mathcal{S}_K}$. Thus, we can conclude $\overline{\mathcal{S}_K}^{\perp} = \{0\}$ and $\overline{\mathcal{S}_K} = \mathcal{H}_K$.

(4) Let φ denote the mapping from Theorem 3.2, i.e.

$$\varphi : \mathcal{H}_K \to \mathcal{H}_K^*,\ f \mapsto \langle \cdot, f\rangle_{\mathcal{H}_K}.$$

Note that we have $\varphi(\mathcal{S}_K) = \mathcal{S}_K^{(*)}$ and $\varphi(\mathcal{H}_K) = \mathcal{H}_K^*$. Since φ is an isometric isomorphism, the density follows from part (3).

(5) Let $(f_n)_{n\in\mathbb{N}} \subset \mathcal{H}_K$ converge normwise to $f \in \mathcal{H}_K$, i.e.

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

For fixed $x \in \mathbb{R}^d$, this implies

$$|f_n(x) - f(x)| = |\langle f_n - f, K(\cdot,x)\rangle_{\mathcal{H}_K}| \leq \|f_n - f\|_{\mathcal{H}_K} \cdot \|K(\cdot,x)\|_{\mathcal{H}_K} \longrightarrow 0 \qquad \text{for } n \to \infty.$$

Hence, $(f_n)_{n\in\mathbb{N}}$ converges pointwise to f. □

Remark 3.7. If K is bounded, i.e. $|K(x,y)| \leq C$ for a constant $C > 0$ and all $x, y \in \mathbb{R}^d$, we can make the estimate

$$|f(x)| = |\langle f, K(\cdot,x)\rangle|_{\mathcal{H}_K} \leq \|K(\cdot,x)\|_{\mathcal{H}_K} \cdot \|f\|_{\mathcal{H}_K} = \sqrt{K(x,x)} \cdot \|f\|_{\mathcal{H}_K} \leq \sqrt{C} \cdot \|f\|_{\mathcal{H}_K} \qquad \text{for } x \in \mathbb{R}^d.$$

In this case, every function $f \in \mathcal{H}_K$ is bounded with the estimate

$$\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)| \leq \sqrt{C} \cdot \|f\|_{\mathcal{H}_K},$$

so that normwise convergence implies uniform convergence (cf. [12, Section 1.2(5)]).

Before we continue with the construction of reproducing kernel Hilbert spaces for given positive semi-definite kernels, we want to answer the question of whether reproducing kernels and their respective Hilbert spaces are unique, see also [70, Remark 8.23] and [143, Theorem 10.11].

Theorem 3.8. *For reproducing kernel Hilbert spaces, we have the following uniqueness properties:*

(1) Let K be a reproducing kernel of two Hilbert spaces $\mathcal{G}$ and $\mathcal{H}$. Then we have $\mathcal{G} = \mathcal{H}$ and the inner products coincide.

(2) Let K_1 and K_2 be reproducing kernels of the Hilbert space $\mathcal{H}$. Then $K_1 = K_2$ holds.

Proof. Throughout this proof, we make use of the results from Theorem 3.6.

(1) We notice that the set $\mathcal{S}_K$ from (3.3) must be included in $\mathcal{G}$ and $\mathcal{H}$. Additionally, we have

$$\langle K(\cdot,x), K(\cdot,y)\rangle_{\mathcal{G}} = K(x,y) = \langle K(\cdot,x), K(\cdot,y)\rangle_{\mathcal{H}} \qquad \text{for all } x, y \in \mathbb{R}^d,$$

so that the inner products $\langle\cdot,\cdot\rangle_{\mathcal{G}}$ and $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ coincide on $\mathcal{S}_K$. Given an arbitrary $f \in \mathcal{G}$, there is a sequence

$$(s_n)_{n\in\mathbb{N}} \subset \mathcal{S}_K \qquad \text{such that} \qquad \|f - s_n\|_{\mathcal{G}} \longrightarrow 0 \quad \text{for } n \to \infty.$$

The sequence is a Cauchy sequence with respect to $\|\cdot\|_{\mathcal{G}}$, and due to

$$\|s_m - s_n\|_{\mathcal{G}} = \|s_m - s_n\|_{\mathcal{H}} \qquad \text{for all } m, n \in \mathbb{N},$$

it is also a Cauchy sequence with respect to $\|\cdot\|_{\mathcal{H}}$. This means that there is $g \in \mathcal{H}$ with

$$\|g - s_n\|_{\mathcal{H}} \longrightarrow 0 \qquad \text{for } n \to \infty.$$

Since normwise convergence implies pointwise convergence, we get

$$f(x) = \lim_{n\to\infty} s_n(x) = g(x) \qquad \text{for all } x \in \mathbb{R}^d,$$

so that $f = g$ holds. Moreover, we get

$$\|f\|_{\mathcal{G}} = \lim_{n\to\infty} \|s_n\|_{\mathcal{G}} = \lim_{n\to\infty} \|s_n\|_{\mathcal{H}} = \|f\|_{\mathcal{H}}.$$

In total, we have $\mathcal{G} \subset \mathcal{H}$ and the two norms coincide on $\mathcal{G}$. Analogously, we can show that $\mathcal{H} \subset \mathcal{G}$ and $\|f\|_{\mathcal{H}} = \|f\|_{\mathcal{G}}$ for $f \in \mathcal{H}$. The polarization identity for inner product spaces (see, e.g., [70, Theorem 3.9]) then implies that the inner products coincide on $\mathcal{G} = \mathcal{H}$ as well.

(2) In order to prove part (2) of this theorem, we use the reproduction property to compute

$$\|K_1(\cdot,x) - K_2(\cdot,x)\|_{\mathcal{H}}^2 = K_1(x,x) - K_1(x,x) - K_2(x,x) + K_2(x,x) = 0 \qquad \text{for } x \in \mathbb{R}^d.$$

This implies $K_1(\cdot,x) = K_2(\cdot,x)$ for all $x \in \mathbb{R}^d$ and therefore $K_1 = K_2$.

□

3.1. Construction of the native space

As we have seen in Proposition 3.6, every reproducing kernel K of a Hilbert space is positive semi-definite. But in practical cases, one usually starts with a positive (semi-)definite kernel. Hence, we go the other way round in this section and construct a reproducing kernel Hilbert space $\mathcal{H}_K$ for a given positive semi-definite kernel function K, called the *native space* of K. This means that we can base the resulting algorithms on positive (semi-)definite kernel functions and still take advantage of the theory of reproducing kernel Hilbert spaces. For the construction of the native space, we follow along [22, Section 1.3].

Let $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be positive definite. The construction starts with the set

$$\mathcal{S}_K = \operatorname{span}_{\mathbb{R}} \{K(\cdot, x) \mid x \in \mathbb{R}^d\}$$

from part (3) of Theorem 3.6. For elements

$$f = \sum_{i=1}^{M} c_i \cdot K(\cdot, x_i) \qquad \text{and} \qquad g = \sum_{j=1}^{N} d_j \cdot K(\cdot, y_j) \tag{3.4}$$

in $\mathcal{S}_K$, we define

$$\langle f, g \rangle_K := \sum_{i=1}^{M} \sum_{j=1}^{N} c_i \cdot d_j \cdot K(x_i, y_j). \tag{3.5}$$

Initially, it is not clear that (3.5) induces a well-defined map

$$\langle \cdot, \cdot \rangle_K : \mathcal{S}_K \times \mathcal{S}_K \to \mathbb{R}.$$

But if f and g from (3.4) have alternative representations

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

we get the equality

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

It is easy to see that $\langle \cdot, \cdot \rangle_K$ is bilinear, symmetric and positive semi-definite. For the positive definiteness, let $f \in \mathcal{S}_K$ satisfy $\langle f, f \rangle_K = 0$. The Cauchy-Schwarz inequality yields

$$0 \le |f(x)|^2 = |\langle f, K(\cdot, x) \rangle_K|^2 \le \langle f, f \rangle_K \cdot K(x, x) = 0 \qquad \text{for all } x \in \mathbb{R}^d,$$

which implies $f = 0$. We summarize our results in the next proposition (cf. [22, Theorem 3]).

Proposition 3.9. *For a given positive semi-definite kernel K, define $(\mathcal{S}_K, \langle \cdot, \cdot \rangle_K)$ as done in (3.3) and (3.5). Then $\mathcal{S}_K$ is a pre-Hilbert space.*

By definition of the inner product, we directly get the reproduction property

$$s(x) = \langle s, K(\cdot, x) \rangle_K \qquad \text{for all } s \in \mathcal{S}_K, x \in \mathbb{R}^d. \tag{3.6}$$

In general, $\mathcal{S}_K$ is not complete, so we need to complete the space with respect to the inner product. From functional analysis, it is well-known that any pre-Hilbert space can be densely embedded into a Hilbert space, see e.g. [76, Theorem 1.5.1 & Proposition 2.1.6]. But the elements of this space might not be functions again, which means that another step is required to convert these elements into functions, see [143, Section 10.2]. We follow a more direct approach here, which merges the two steps (cf. [22, Theorem 2 ff.]). This approach uses the following lemma:

Lemma 3.10. *Let $(s_n)_{n\in\mathbb{N}} \subset \mathcal{S}_K$ be a Cauchy sequence that converges pointwise to the zero function, i.e.*

$$\lim_{n\to\infty} s_n(x) = 0 \qquad \text{for all } x \in \mathbb{R}^d.$$

Then we have the normwise convergence $\|s_n\|_K \xrightarrow{n\to\infty} 0$.

Proof. Given $m, n \in \mathbb{N}$, we can rewrite the term $\|s_n\|_K^2$ as

$$\|s_n\|_K^2 = \langle s_n, s_n\rangle_K = \langle s_n - s_m, s_n\rangle_K + \langle s_m, s_n\rangle_K. \tag{3.7}$$

Now let $\varepsilon > 0$. Since Cauchy sequences are automatically bounded, we find $C > 0$ such that $\|s_n\|_K \le C$ for all $n \in \mathbb{N}$ and choose $N_1 \in \mathbb{N}$ such that

$$\|s_m - s_n\|_K < \frac{\varepsilon}{2C} \qquad \text{for } m, n \ge N_1.$$

Moreover, we can write s_{N_1} as

$$s_{N_1} = \sum_{i=1}^{M} c_i \cdot K(\cdot, x_i).$$

Using the reproduction property (3.6) on $\mathcal{S}_K$, we derive

$$\langle s_n, s_{N_1}\rangle_K = \sum_{i=1}^{M} c_i \cdot s_n(x_i) \xrightarrow{n\to\infty} 0$$

due to our assumptions. Hence, we can find $N_2 \in \mathbb{N}$ such that

$$\langle s_n, s_{N_1}\rangle_K \le \frac{\varepsilon}{2} \qquad \text{for all } n \ge N_2.$$

Setting $N = \max\{N_1, N_2\}$, we can use the identity (3.7) to get

$$\|s_n\|_K^2 \le \|s_{N_1} - s_n\|_K \cdot \|s_n\|_K + \langle s_{N_1}, s_n\rangle_K < \varepsilon \qquad \text{for } n \ge N,$$

which proves the assertion. □

Theorem 3.11 (Moore-Aronszajn)**.** *Under the assumptions of Proposition 3.9, let*

$$\mathcal{H}_K := \Big\{ f : \mathbb{R}^d \to \mathbb{R} \;\Big|\; \text{there is a Cauchy sequence } (s_n)_{n\in\mathbb{N}} \subset \mathcal{S}_K \text{ that converges pointwise to } f \Big\}.$$

Moreover, we equip $\mathcal{H}_K$ with the bilinear form

$$\langle f, g\rangle_K := \lim_{n\to\infty} \langle s_n, t_n\rangle_K \qquad \text{for } f, g \in \mathcal{H}_K, \tag{3.8}$$

where $(s_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ are Cauchy sequences in $\mathcal{S}_K$ that converge pointwise to f and g. Then $(\mathcal{H}_K, \langle\cdot,\cdot\rangle_K)$ is a Hilbert space with reproducing kernel K.

Proof. $\mathcal{H}_K$ is a vector space, as the set of Cauchy sequences in $\mathcal{S}_K$ is closed under componentwise addition and scalar multiplication. The proof is divided into six steps:

(i) **Well-definedness I:** Let $(s_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ be Cauchy sequences in $\mathcal{S}_K$. Due to the inequality

$$|\langle s_m, t_m\rangle_K - \langle s_n, t_n\rangle_K| \le \|s_m - s_n\|_K \cdot \|t_m\|_K + \|s_n\|_K \cdot \|t_m - t_n\|_K \qquad \text{for all } m, n \in \mathbb{N},$$

the sequence $(\langle s_n, t_n\rangle_K)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$ and therefore convergent. Hence, the term

$$\lim_{n\to\infty} \langle s_n, t_n\rangle_K$$

is well-defined.

(ii) **Well-definedness II:** Let $f, g \in \mathcal{H}_K$ and $(s_n)_{n\in\mathbb{N}}, (\tilde{s}_n)_{n\in\mathbb{N}}, (t_n)_{n\in\mathbb{N}}, (\tilde{t}_n)_{n\in\mathbb{N}}$ be Cauchy sequences in $\mathcal{S}_K$ such that $(s_n)_{n\in\mathbb{N}}$, $(\tilde{s}_n)_{n\in\mathbb{N}}$ converge pointwise to f and $(t_n)_{n\in\mathbb{N}}, (\tilde{t}_n)_{n\in\mathbb{N}}$ converge pointwise to g. Due to Lemma 3.10, we have the convergence

$$\|s_n - \tilde{s}_n\|_K,\ \|t_n - \tilde{t}_n\|_K \xrightarrow{n\to\infty} 0.$$

Similar to part (i), we get

$$|\langle s_n, t_n\rangle_K - \langle \tilde{s}_n, \tilde{t}_n\rangle_K| \le \|s_n - \tilde{s}_n\|_K \cdot \|t_n\|_K + \|\tilde{s}_n\|_K \cdot \|t_n - \tilde{t}_n\|_K \xrightarrow{n\to\infty} 0,$$

which implies

$$\lim_{n\to\infty} \langle s_n, t_n\rangle_K = \lim_{n\to\infty} \langle \tilde{s}_n, \tilde{t}_n\rangle_K.$$

This means that (3.8) induces a well-defined mapping. Moreover, we have $\mathcal{S}_K \subset \mathcal{H}_K$, so that $\langle\cdot,\cdot\rangle_K : \mathcal{H}_K \times \mathcal{H}_K \to \mathbb{R}$ is an extension of the previously defined inner product (3.5) on $\mathcal{S}_K$.

(iii) **Inner product:** It is clear that $\langle\cdot,\cdot\rangle_K$ is bilinear and symmetric on $\mathcal{H}_K \times \mathcal{H}_K$. For $f \in \mathcal{H}_K$, consider a Cauchy sequence $(s_n)_{n\in\mathbb{N}}$ that converges pointwise to f. By definition (3.5), we get

$$\langle f, f\rangle_K = \lim_{n\to\infty} \langle s_n, s_n\rangle_K = \lim_{n\to\infty} \|s_n\|_K^2 \ge 0.$$

Suppose that $\langle f, f\rangle_K = 0$. With the reproduction property on $\mathcal{S}_K$, we can make the estimate

$$|s_n(x)|^2 = |\langle s_n, K(\cdot, x)\rangle_K|^2 \le \langle s_n, s_n\rangle_K \cdot K(x,x) \xrightarrow{n\to\infty} \langle f, f\rangle_K \cdot K(x,x) = 0 \qquad \text{for all } x \in \mathbb{R}^d.$$

This yields

$$f(x) = \lim_{n\to\infty} s_n(x) = 0 \qquad \text{for all } x \in \mathbb{R}^d,$$

which is equivalent to $f = 0$. Hence, $\langle\cdot,\cdot\rangle_K$ is an inner product on $\mathcal{H}_K$.

(iv) **Completeness I:** Let $(s_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{S}_K$. Due to the estimate

$$|s_m(x) - s_n(x)|^2 = |\langle s_m - s_n, K(\cdot, x)\rangle_K|^2 \le \|s_m - s_n\|_K^2 \cdot K(x,x) \qquad \text{for all } x \in \mathbb{R}^d,$$

the sequence $(s_n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$ for every point $x \in \mathbb{R}^d$. Define the function $f : \mathbb{R}^d \to \mathbb{R}$ via

$$f(x) := \lim_{n\to\infty} s_n(x) \qquad \text{for all } x \in \mathbb{R}^d.$$

Then we have $f \in \mathcal{H}_K$ and

$$\|f - s_n\|_K^2 = \langle f - s_n, f - s_n\rangle_K = \lim_{m\to\infty} \langle s_m - s_n, s_m - s_n\rangle_K = \lim_{m\to\infty} \|s_m - s_n\|_K^2 \qquad \text{for all } n \in \mathbb{N}.$$

Now let $\varepsilon > 0$. We choose $N \in \mathbb{N}$ such that $\|s_m - s_n\|_K^2 < \varepsilon$ for all $m, n \ge N$. This results in

$$\|f - s_n\|_K^2 = \lim_{m\to\infty} \|s_m - s_n\|_K^2 \le \varepsilon \qquad \text{for } n \ge N$$

which proves that $(s_n)_{n\in\mathbb{N}}$ converges normwise to f. Moreover, this argument shows that $\mathcal{S}_K$ is dense in $\mathcal{H}_K$.

(v) **Completeness II:** Let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{H}_K$. Since $\mathcal{S}_K$ is dense in $\mathcal{H}_K$ (see part (iv)), we can find a sequence $(s_n)_{n\in\mathbb{N}}$ in $\mathcal{S}_K$ with

$$\|f_n - s_n\|_K < \frac{1}{n} \qquad \text{for all } n \in \mathbb{N}.$$

Due to the estimate

$$\|s_m - s_n\|_K \le \|f_m - s_m\|_K + \|f_m - f_n\|_K + \|f_n - s_n\|_K < \frac{1}{m} + \frac{1}{n} + \|f_m - f_n\|_K,$$

the sequence $(s_n)_{n\in\mathbb{N}}$ is a Cauchy sequence as well. With part (iv), we can conclude that

$$f = \lim_{n\to\infty} s_n \in \mathcal{H}_K$$

exists. But f is also the limit of the initial sequence $(f_n)_{n\in\mathbb{N}}$, as we have

$$\|f - f_n\|_K \leq \|f - s_n\|_K + \|s_n - f_n\| < \|f - s_n\|_K + \frac{1}{n} \xrightarrow{n\to\infty} 0.$$

(vi) **Reproduction property:** Let $f \in \mathcal{H}_K$ and choose a Cauchy sequence $(s_n)_{n\in\mathbb{N}}$ that converges pointwise to f. For every $x \in \mathbb{R}^d$, we get

$$f(x) = \lim_{n\to\infty} s_n(x) = \lim_{n\to\infty} \langle s_n, K(\cdot, x)\rangle_K = \langle f, K(\cdot, x)\rangle_K$$

by definition of the inner product.

In total, $(\mathcal{H}_K, \langle\cdot,\cdot\rangle_K)$ is a Hilbert space with reproducing kernel K. □

From now on, $\mathcal{H}_K$ always denotes the native reproducing kernel Hilbert space of the kernel K which we constructed in Theorem 3.11.

Since the evaluation of linear functionals from the dual space $\mathcal{H}_K^*$ plays a key role in the next chapters, we want to derive a generalized reproduction property that holds for all elements in the dual space, not only for point evaluation functionals. The result below can be found in [143, Theorem 16.7] and in [70, Theorem 8.24], where it is called *Madych-Nelson Theorem*, referring to the work of Madych and Nelson (cf. [87], [88], [89]).

Theorem 3.12 (Madych-Nelson)**.** *Let K be positive semi-definite on $\mathbb{R}^d$ and $\mathcal{H}_K$ be the respective native space. Then, the following two statements hold:*

(1) For every $\lambda \in \mathcal{H}_K^*$, the function $\lambda^y K(\cdot, y) : \mathbb{R}^d \to \mathbb{R}$, defined by

$$\lambda^y K(x, y) := \lambda(K(\cdot, x)) = \lambda(K(x, \cdot)) \qquad \text{for all } x \in \mathbb{R}^d,$$

is an element of the native space, i.e. $\lambda^y K(\cdot, y) \in \mathcal{H}_K$.

(2) We have the generalized reproduction property

$$\lambda(f) = \langle f, \lambda^y K(\cdot, y)\rangle_K \qquad \text{for all } f \in \mathcal{H}_K,\ \lambda \in \mathcal{H}_K^*.$$

Proof. Let $g_\lambda \in \mathcal{H}_K$ be the Riesz representer of $\lambda \in \mathcal{H}_K^*$, i.e.

$$\lambda(f) = \langle f, g_\lambda\rangle_K \qquad \text{for all } f \in \mathcal{H}_K.$$

Then we have $\lambda^y K(\cdot, y) = g_\lambda$, as the equation

$$\lambda^y K(x, y) = \lambda(K(\cdot, x)) = \langle K(\cdot, x), g_\lambda\rangle_K = g_\lambda(x)$$

holds for all $x \in \mathbb{R}$. This proves (1) and (2). □

Remark 3.13. Note that Theorem 3.12 presents an important advantage of reproducing kernel Hilbert spaces, as we can explicitly compute the Riesz representers of functionals via the mapping

$$\varphi^{-1} : \mathcal{H}_K^* \to \mathcal{H}_K,\ \lambda \mapsto \lambda^y K(\cdot, y).$$

This advantage is highly relevant for the implementation of kernel-based algorithms for generalized interpolation problems, see Chapter 4 and 10.

In order to determine whether well-known functionals (e.g. distributions) are elements of the dual native space, we have a look at the regularity of functions from the native space. As the construction of the native space is solely based on the given kernel, the functions of the native space inherit the regularity of the kernel function. We state some basic inclusions (cf. [70, Corollary 8.27], [132, Theorem 4.26], [143, Theorem 10.45]).

Proposition 3.14. *Let K be a positive semi-definite kernel function and $\mathcal{H}_K$ be the respective native space.*

(1) Let K be bounded. Then any function $f \in \mathcal{H}_K$ is bounded by

$$\|f\|_\infty \leq \sup_{x \in \mathbb{R}^d} \sqrt{K(x,x)} \cdot \|f\|_K < \infty.$$

(2) If $K \in \mathscr{C}(\mathbb{R}^d \times \mathbb{R}^d)$, then $\mathcal{H}_K \subset \mathscr{C}(\mathbb{R}^d)$ holds.

(3) Let $p \in [1, \infty)$ and $K \in \mathscr{C}(\mathbb{R}^d \times \mathbb{R}^d)$ satisfy

$$\int_{\mathbb{R}^d} K(x,x)^{p/2} \, dx < \infty.$$

Then we have $\mathcal{H}_K \subset L^p(\mathbb{R}^d)$ and the norm inequality

$$\|f\|_{L^p(\mathbb{R}^d)} \leq \left(\int_{\mathbb{R}^d} K(x,x)^{p/2} \, dx \right)^{1/p} \cdot \|f\|_K < \infty \qquad \text{for all } f \in \mathcal{H}_K.$$

(4) Suppose that $K \in \mathscr{C}^{2k}(\mathbb{R}^d \times \mathbb{R}^d)$ for $k \in \mathbb{N}_0$. Then we have $\mathcal{H}_K \subset C^k(\mathbb{R}^d)$. Moreover, we have

$$\lambda_{x,\alpha} := \delta_x \circ D^\alpha \in \mathcal{H}_K^* \qquad \text{for } x \in \mathbb{R}^d$$

and multi-indices $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq k$. Here, D^α denotes the partial derivative operator

$$D^\alpha(f) := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} \qquad \text{for all } f \in C^k(\mathbb{R}^d).$$

Proof. The proof of the stated inclusions mainly relies on the reproduction property (3.2):

(1) This is already proven in Remark 3.7.

(2) Let $f \in \mathcal{H}_K$ and $x \in \mathbb{R}^d$. Choose a sequence $(x_n)_{n \in \mathbb{N}} \subset \mathbb{R}^d$ that converges to x. We can make the estimate

$$|f(x) - f(x_n)| = |\langle f, K(\cdot, x) - K(\cdot, x_n) \rangle_K| \leq \|f\|_K \cdot \|K(\cdot, x) - K(\cdot, x_n)\|_K. \tag{3.9}$$

Since K is continuous, we have

$$\|K(\cdot, x) - K(\cdot, x_n)\|_K^2 = K(x,x) - 2 \cdot K(x, x_n) + K(x_n, x_n) \xrightarrow{n \to \infty} 2 \cdot (K(x,x) - K(x,x)) = 0,$$

and therefore $|f(x) - f(x_n)| \xrightarrow{n \to \infty} 0$ according to (3.9). Hence, f is continuous on $\mathbb{R}^d$.

(3) From part (1), we know that $\mathcal{H}_K \subset \mathscr{C}(\mathbb{R}^d)$. This implies that every function $f \in \mathcal{H}_K$ is measurable. For the norm estimate, we compute

$$\int_{\mathbb{R}^d} |f(x)|^p \, dx = \int_{\mathbb{R}^d} |\langle f, K(\cdot, x) \rangle_K|^p \, dx \leq \|f\|_K^p \cdot \int_{\mathbb{R}^d} \|K(\cdot, x)\|_K^p \, dx = \|f\|_K^p \cdot \int_{\mathbb{R}^d} K(x,x)^{p/2} \, dx < \infty$$

for $f \in \mathcal{H}_K$. This proves the norm inequality and the desired inclusion.

(4) We prove the assertion via induction on $k \in \mathbb{N}_0$. For $k = 0$, the statements follow from part (2) of this proposition and Theorem 3.3. Assuming that the assertion holds for $k \in \mathbb{N}_0$, let $K \in \mathscr{C}^{2k+2}(\mathbb{R}^d \times \mathbb{R}^d)$ and

$$\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d \qquad \text{with} \qquad |\alpha| \leq k+1.$$

If $|\alpha| < k+1$ holds, the statements follow from the induction hypothesis due to the inclusion $\mathscr{C}^{2k+2}(\mathbb{R}^d \times \mathbb{R}^d) \subset \mathscr{C}^{2k}(\mathbb{R}^d \times \mathbb{R}^d)$. Hence, we focus on the case $|\alpha| = k+1$. Divided into several steps, we show that $\lambda_{x,\alpha}$ is well-defined and bounded on $\mathcal{H}_K$ for all $x \in \mathbb{R}^d$:

(i) Select the minimal $i \in \{1, ..., d\}$ with $\alpha_i \neq 0$ and set

$$\tilde{\alpha} = (\alpha_1, ..., \alpha_i - 1, ..., \alpha_d)\,.$$

Then we have $\tilde{\alpha} \in \mathbb{N}_0^d$ with $|\tilde{\alpha}| = k$, and the induction hypothesis in combination with Theorem 3.12 implies

$$\lambda_{x,\tilde{\alpha}}(f) = \langle f, \lambda_{x,\tilde{\alpha}}^y K(\cdot, y)\rangle_{\mathcal{H}_K} \qquad \text{for } f \in \mathcal{H}_K, x \in \mathbb{R}^d. \tag{3.10}$$

Here, $\lambda_{x,\tilde{\alpha}}^y K(\cdot, y)$ is given by

$$\lambda_{x,\tilde{\alpha}}^y K(z, y) = D_2^{\tilde{\alpha}} K(z, x) \qquad \text{for } z \in \mathbb{R}^d,$$

where $D_2^{\tilde{\alpha}}$ denotes that the differential operator $D^{\tilde{\alpha}}$ for differentiable functions on $\mathbb{R}^d$ acts with respect to the second d-dimensional argument of K. The operator $D_1^{\tilde{\alpha}}$ is defined analogously. With (3.10), we get

$$\frac{D^{\tilde{\alpha}} f(x + h \cdot e^{(i)}) - D^{\tilde{\alpha}} f(x)}{h} = \left\langle f,\ \frac{\lambda_{x+h\cdot e^{(i)},\tilde{\alpha}}^y K(\cdot, y) - \lambda_{x,\tilde{\alpha}}^y K(\cdot, y)}{h} \right\rangle_K \tag{3.11}$$

for all $f \in \mathcal{H}_K, x \in \mathbb{R}^d$ and $h \neq 0$, where $e^{(i)} \in \mathbb{R}^d$ denotes the i-th standard basis vector.

(ii) Given $x \in \mathbb{R}^d$, we define the functions

$$g_{x,h} := \frac{\lambda_{x+h\cdot e^{(i)},\tilde{\alpha}}^y K(\cdot, y) - \lambda_{x,\tilde{\alpha}}^y K(\cdot, y)}{h} \in \mathcal{H}_K \qquad \text{for } h \neq 0.$$

Note that the corresponding functionals in the dual space are given by

$$\langle f, g_{x,h}\rangle_K = \left(\delta_x \circ \Delta_h^{e^{(i)}}\right)(D^{\tilde{\alpha}} f) \qquad \text{for all } f \in \mathcal{H}_K, x \in \mathbb{R}^d, h \neq 0,$$

where we interpret the vector $e^{(i)}$ as a multi-index and $\Delta_h^{e^{(i)}}$ is the forward difference operator from Definition A.7. Again, we denote the application of $\Delta_h^{e^{(i)}}$ with respect to the first and second d-dimensional argument of K as $\Delta_{h,1}^{e^{(i)}}$ and $\Delta_{h,2}^{e^{(i)}}$. In order to justify the differentiability, we show that

$$\lim_{h\to 0} g_{x,h}$$

exists in $\mathcal{H}_K$.

(iii) Let $(h_n)_{n\in\mathbb{N}}$ be a null sequence in $\mathbb{R} \setminus \{0\}$. For $m, n \in \mathbb{N}$, we have

$$\begin{aligned}\|g_{x,h_m} - g_{x,h_n}\|_K^2 = {} & \Delta_{h_m,1}^{e^{(i)}} \Delta_{h_m,2}^{e^{(i)}} \left[D_1^{\tilde{\alpha}} D_2^{\tilde{\alpha}} K\right](x, x) - \Delta_{h_m,1}^{e^{(i)}} \Delta_{h_n,2}^{e^{(i)}} \left[D_1^{\tilde{\alpha}} D_2^{\tilde{\alpha}} K\right](x, x) \\ & - \Delta_{h_n,1}^{e^{(i)}} \Delta_{h_m,2}^{e^{(i)}} \left[D_1^{\tilde{\alpha}} D_2^{\tilde{\alpha}} K\right](x, x) + \Delta_{h_n,1}^{e^{(i)}} \Delta_{h_n,2}^{e^{(i)}} \left[D_1^{\tilde{\alpha}} D_2^{\tilde{\alpha}} K\right](x, x).\end{aligned}$$

Due to Theorem A.8 part (2), we can find points

$$\zeta_{1,1}^{(m,n)}, \zeta_{1,2}^{(m,n)}, \zeta_{2,1}^{(m,n)}, \zeta_{3,2}^{(m,n)} \in \left[x, x + h_m \cdot e^{(i)}\right]$$

and

$$\zeta_{2,2}^{(m,n)}, \zeta_{3,1}^{(m,n)}, \zeta_{4,1}^{(m,n)}, \zeta_{4,2}^{(m,n)} \in \left[x, x + h_n \cdot e^{(i)}\right]$$

for $m, n \in \mathbb{N}$, where

$$\left[x, x + h \cdot e^{(i)}\right] := \left\{x + t \cdot e^{(i)} \;\middle|\; \min(0, h) \leq t \leq \max(0, h)\right\} \qquad \text{for } x \in \mathbb{R}^d,\ h \neq 0,$$

such that

$$\begin{aligned}\|g_{x,h_m} - g_{x,h_n}\|_K^2 = {} & D_1^\alpha D_2^\alpha K(\zeta_{1,1}^{(m,n)}, \zeta_{1,2}^{(m,n)}) - D_1^\alpha D_2^\alpha K(\zeta_{2,1}^{(m,n)}, \zeta_{2,2}^{(m,n)}) \\ & - D_1^\alpha D_2^\alpha K(\zeta_{3,1}^{(m,n)}, \zeta_{3,2}^{(m,n)}) + D_1^\alpha D_2^\alpha K(\zeta_{4,1}^{(m,n)}, \zeta_{4,2}^{(m,n)}).\end{aligned}$$

Given $\varepsilon > 0$, we can find $\delta > 0$ such that

$$\left| D_1^\alpha D_2^\alpha K(y,z) - D_1^\alpha D_2^\alpha K(x,x) \right| < \frac{\varepsilon}{4} \qquad \text{for} \qquad \| (y,z) - (x,x) \|_2 < \delta$$

since $D_1^\alpha D_2^\alpha K$ is continuous. Now choose $N \in \mathbb{N}$ such that $|h_n| < \delta/\sqrt{2}$ for $n \geq N$. Then we have

$$\left\| (\zeta_{j,1}^{(m,n)}, \zeta_{j,2}^{(m,n)}) - (x,x) \right\|_2 < \delta \qquad \text{for } m,n \geq N,\ j = 1,...,4,$$

and therefore

$$\begin{aligned}\|g_{x,h_m} - g_{x,h_n}\|_K^2 \leq {} & |D_1^\alpha D_2^\alpha K(\zeta_{1,1}^{(m,n)}, \zeta_{1,2}^{(m,n)}) - D_1^\alpha D_2^\alpha K(x,x)| \\ & + |D_1^\alpha D_2^\alpha K(\zeta_{2,1}^{(m,n)}, \zeta_{2,2}^{(m,n)}) - D_1^\alpha D_2^\alpha K(x,x)| \\ & + |D_1^\alpha D_2^\alpha K(\zeta_{3,1}^{(m,n)}, \zeta_{3,2}^{(m,n)}) - D_1^\alpha D_2^\alpha K(x,x)| \\ & + |D_1^\alpha D_2^\alpha K(\zeta_{4,1}^{(m,n)}, \zeta_{4,2}^{(m,n)}) - D_1^\alpha D_2^\alpha K(x,x)| \\ < {} & \varepsilon.\end{aligned}$$

Hence, the sequence $(g_{x,h_n})_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{H}_K$, so that

$$g_x := \lim_{n\to\infty} g_{x,h_n} \in \mathcal{H}_K \tag{3.12}$$

exists.

(iv) Since normwise convergence implies pointwise convergence in the native space, the limit function in (3.12) is given by

$$g_x(z) = \lim_{n\to\infty} g_{x,h_n}(z) = D^\alpha K(z,x) = \lambda_{x,\alpha}^y K(z,y) \qquad \text{for all } z \in \mathbb{R}^d.$$

Note that the limit function $\lambda_{x,\alpha}^y K(\cdot, y)$ does not depend on the chosen sequence $(h_n)_{n\in\mathbb{N}}$. In total, we get

$$\lim_{\substack{h\to 0 \\ h\neq 0}} g_{x,h} = \lambda_{x,\alpha}^y K(\cdot, y) \in \mathcal{H}_K.$$

(v) As a consequence of (iv), we can let $h \to 0$ in (3.11) to get

$$\lambda_{x,\alpha}(f) = D^\alpha f(x) = \lim_{h\to 0} \frac{D^{\tilde{\alpha}} f(x + h\cdot e^{(i)}) - D^{\tilde{\alpha}} f(x)}{h} = \left\langle f,\ \lambda_{x,\alpha}^y K(\cdot, y) \right\rangle_K,$$

so that $D^\alpha f(x)$ exists for all $f \in \mathcal{H}_K, x \in \mathbb{R}^d$. Moreover, this proves $\lambda_{x,\alpha} \in \mathcal{H}_K^*$. Similar to the proof of part (2), we estimate

$$\begin{aligned}|D^\alpha f(x) - D^\alpha f(z)| &\leq \|f\|_K \cdot \|\lambda_{x,\alpha}^y K(\cdot,y) - \lambda_{z,\alpha}^y K(\cdot,y)\|_K \\ &= \|f\|_K \cdot (D_1^\alpha D_2^\alpha K(x,x) - D_1^\alpha D_2^\alpha K(z,x) - D_1^\alpha D_2^\alpha K(x,z) + D_1^\alpha D_2^\alpha K(z,z))^{1/2}\end{aligned}$$

for $f \in \mathcal{H}_K$ and $x, z \in \mathbb{R}^d$. With the assumption $K \in \mathscr{C}^{2k+2}(\mathbb{R}^d \times \mathbb{R}^d)$, we can conclude that the derivative $D^\alpha f$ is continuous on $\mathbb{R}^d$ for each $f \in \mathcal{H}_K$. For $k > 0$, an application of *Schwarz's theorem* for partial derivatives in combination with the induction hypothesis shows that the order of differentiation does not matter.

Since α was chosen arbitrarily, we have $\mathcal{H}_K \subset \mathscr{C}^{k+1}(\mathbb{R}^d)$, and the respective partial differential operators are elements of the dual space. □

Corollary 3.15. *Let $K \in \mathscr{C}(\mathbb{R}^d \times \mathbb{R}^d)$ be a bounded positive semi-definite function and satisfy the condition $\int_{\mathbb{R}^d} K(x,x)^{1/2}\, dx < \infty$. Then we have*

$$\int_{\mathbb{R}^d} K(x,x)^{p/2}\, dx < \infty \qquad \text{for all } p \in [1,\infty)\,,$$

so that the statements of Proposition (3.14) part (3) hold for every $p \in [1,\infty)$.

Proof. We can estimate

$$\int_{\mathbb{R}^d} K(x,x)^{p/2}\, dx \leq \left(\max_{x\in\mathbb{R}^d} K(x,x)\right)^{(p-1)/2} \cdot \int_{\mathbb{R}^d} K(x,x)^{1/2}\, dx < \infty \qquad \text{for } p \in [1,\infty)\,.$$

□

Remark 3.16. The results from Proposition 3.14 and Corollary 3.15 are helpful when restricting operators to a native space. For example, consider a bounded linear operator $A : L^p(\mathbb{R}^d) \to Y$, where Y is a normed space. Part (3) of Proposition 3.14 then implies that A induces a bounded linear operator from $(\mathcal{H}_K, \|\cdot\|_K)$ to Y via the restriction

$$A|_{\mathcal{H}_K} : (\mathcal{H}_K, \|\cdot\|_K) \to Y,\ f \mapsto A(f).$$

We use these ideas later in the context of computerized tomography, see Chapter 10.

3.2. Alternative characterizations of native spaces

Besides the construction in Theorem 3.11, native spaces can be expressed in other terms, which can be useful for theoretical considerations. We state two alternative characterizations here without proof. For more details and proofs, we refer to [143, Section 10.2, 10.4].

The first alternative characterization is based on the dual set $\mathcal{S}_K^{(*)}$ from Proposition 3.6, given by

$$\mathcal{S}_K^{(*)} = \operatorname{span}_{\mathbb{R}} \{\delta_x \mid x \in \mathbb{R}^d\} \subset \mathcal{H}_K^*.$$

In this notation, the native space $\mathcal{H}_K$ consists of all functions $f : \mathbb{R}^d \to \mathbb{R}$, such that the evaluations of all $\lambda \in \mathcal{S}_K^{(*)}$ in f are uniformly bounded (cf. [143, Theorem 10.22]).

Theorem 3.17. *Let K be positive semi-definite. We set*

$$\mathcal{G} := \left\{ f : \mathbb{R}^d \to \mathbb{R} \;\middle|\; \text{there is } C_f > 0 \text{ such that } |\lambda(f)| \leq C_f \cdot \|\lambda\|_K \text{ for all } \lambda \in \mathcal{S}_K^{(*)} \right\}$$

and equip $\mathcal{G}$ with the functional

$$\|f\|_{\mathcal{G}} := \sup_{\lambda \in \mathcal{S}_K^{(*)} \setminus \{0\}} \frac{|\lambda(f)|}{\|\lambda\|_K} \qquad \text{for all } f \in \mathcal{G}.$$

Then $\mathcal{G} = \mathcal{H}_K$ and $\|\cdot\|_{\mathcal{G}} = \|\cdot\|_K$ hold.

The second characterization of native spaces is restricted to translation-invariant kernels, which we introduced in Subsection 2.1.1. Under suitable conditions, the inner product $\langle\cdot,\cdot\rangle_K$ can then be expressed with an integral formula (cf. [143, Theorem 10.12]).

Theorem 3.18. *Let $K(\cdot,\cdot) = \Phi(\cdot - \cdot)$ be positive definite, where $\Phi \in \mathscr{C}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$. Moreover, let the Fourier transform $\mathcal{F}\Phi$ of Φ be strictly positive on $\mathbb{R}^d$. We define*

$$\mathcal{G} := \left\{ f \in \mathscr{C}(\mathbb{R}^d) \cap L^2(\mathbb{R}^d) \;\middle|\; \mathcal{F}f \cdot \mathcal{F}\Phi^{-1/2} \in L^2(\mathbb{R}) \right\},$$

where $\mathcal{F}f$ denotes the L^2-Fourier transform of f, and equip $\mathcal{G}$ with the bilinear form

$$\langle f, g\rangle_{\mathcal{G}} := (2\pi)^{-d/2} \cdot \int_{\mathbb{R}^d} \frac{\mathcal{F}f(\omega) \cdot \overline{\mathcal{F}g(\omega)}}{\mathcal{F}\Phi(\omega)}\, d\omega \qquad \text{for all } f, g \in \mathcal{G}.$$

Then $(\mathcal{G}, \langle\cdot,\cdot\rangle_{\mathcal{G}})$ coincides with the native space $(\mathcal{H}_K, \langle\cdot,\cdot\rangle_K)$ of K.

This characterization is of importance as it builds a connection to *Sobolev spaces* (see Section A.5) and their underlying properties (cf. [143, Corollary 10.13]).

Corollary 3.19. *Under the assumptions of Theorem 3.18, let $\mathcal{F}\Phi$ satisfy the estimate*

$$C_1 \cdot \left(1 + \|\omega\|_2^2\right)^{-a} \leq \mathcal{F}\Phi(\omega) \leq C_2 \cdot \left(1 + \|\omega\|_2^2\right)^{-a} \qquad \text{for all } \omega \in \mathbb{R}^d$$

for constants $C_1, C_2 > 0$ and $a > d/2$. Then $\mathcal{H}_K$ coincides with the Sobolev space $\mathcal{H}^a(\mathbb{R}^d)$ and the norm $\|\cdot\|_K$ is equivalent to the respective Sobolev norm.

It was shown in [141] that Wendland's kernels from Subsection 2.1.3 satisfy the conditions of Corollary 3.19 for suitable constants.

3.3. Native spaces of product kernels

In addition to the results of Proposition 3.14, the native space usually inherits the structure of the kernel in special cases. One example of structure preservation is given by product kernels, which were introduced in Subsection 2.1.3 of this thesis. Given $x \in \mathbb{R}^d \simeq \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_M}$ and a product kernel $K = \prod_{i=1}^M K_i$, the respective standard basis function

$$K(\cdot, x) = K_1(p_1(\cdot), p_1(x)) \cdot ... \cdot K_M(p_M(\cdot), p_M(x)) \in \mathcal{H}_K \tag{3.13}$$

is the product of the standard basis functions $K_i(p_i(\cdot), p_i(x))$, $i = 1, ..., M$. Regarding the inner product, we get the identity

$$\langle K(\cdot, x), K(\cdot, y)\rangle_K = \prod_{i=1}^M K_i(p_i(x), p_i(y)) = \prod_{i=1}^M \langle K_i(\cdot, p_i(x)), K_i(\cdot, p_i(y))\rangle_{K_i} \qquad \text{for all } x, y \in \mathbb{R}^d. \tag{3.14}$$

By linear continuation of (3.13) and (3.14), we conclude that

$$\Pi_K : \mathcal{S}_{K_1} \times ... \times \mathcal{S}_{K_M} \to \mathcal{H}_K,\ (s_1, ..., s_M) \mapsto \prod_{i=1}^M s_i = s_1 \cdot ... \cdot s_M \tag{3.15}$$

is a well-defined multilinear mapping that satisfies

$$\Big\langle \Pi_K(s_1, ..., s_M), \Pi_K(\tilde{s}_1, ..., \tilde{s}_M)\Big\rangle_K = \prod_{i=1}^M \langle s_i, \tilde{s}_i\rangle_{K_i} \qquad \text{for all } s_i, \tilde{s_i} \in \mathcal{S}_{K_i}, i = 1, ..., M, \tag{3.16}$$

where $\prod_{i=1}^M s_i$ denotes the mapping

$$\prod_{i=1}^M s_i : \mathbb{R}^d \simeq \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_M} \to \mathbb{R},\ x \mapsto \prod_{i=1}^M s_i(p_i(x))$$

for $s_i \in \mathcal{S}_{K_i}$, $i = 1, ..., M$. In [12, Chapter I.8], [103, Chapter VI], it was already shown that the mapping Π_K from (3.15) and the property (3.16) can be extended to $\mathcal{H}_{K_1} \times ... \times \mathcal{H}_{K_M}$ for the case $M = 2$. We provide an extended proof for all $M \in \mathbb{N}$.

Theorem 3.20. *Let $K = \prod_{i=1}^M K_i$ be a product kernel on $\mathbb{R}^d \simeq \mathbb{R}^d \simeq \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_M}$. Then*

$$\Pi_K : \mathcal{H}_{K_1} \times ... \times \mathcal{H}_{K_M} \to \mathcal{H}_K,\ (f_1, ..., f_M) \mapsto \prod_{i=1}^M f_i \tag{3.17}$$

is a well-defined multilinear mapping. Moreover, the linear space $\operatorname{span}_{\mathbb{R}}(\Pi_K(\mathcal{H}_{K_1} \times ... \times \mathcal{H}_{K_M}))$ is dense in $\mathcal{H}_K$ and we have

$$\Big\langle \Pi_K(f_1, ..., f_M), \Pi_K(g_1, ..., g_M)\Big\rangle_K = \prod_{i=1}^M \langle f_i, g_i\rangle_{K_i} \qquad \text{for all } f_i, g_i \in \mathcal{H}_{K_i}, i = 1, ..., M. \tag{3.18}$$

Proof. If $M = 1$, there is nothing to show. For $M \geq 2$, we prove the proposed statements via induction on M.

$\underline{M = 2:}$ We start with the well-definedness of Π_K. Therefore, let $f_1 \in \mathcal{H}_{K_1}$ and $f_2 \in \mathcal{H}_{K_2}$. We can find convergent sequences

$$\left(s_n^{(1)}\right)_{n\in\mathbb{N}} \subset \mathcal{S}_{K_1},\ \left(s_n^{(2)}\right)_{n\in\mathbb{N}} \subset \mathcal{S}_{K_2} \quad \text{with} \quad \lim_{n\to\infty} s_n^{(1)} = f_1,\ \lim_{n\to\infty} s_n^{(2)} = f_2.$$

Due to (3.16),the estimate

$$\begin{aligned}\|s_m^{(1)} \cdot s_m^{(2)} - s_n^{(1)} \cdot s_n^{(2)}\|_K &= \left\| \left(s_m^{(1)} - s_n^{(1)}\right) \cdot s_m^{(2)} + s_n^{(1)} \cdot \left(s_m^{(2)} - s_n^{(2)}\right) \right\|_K \\ &\leq \|s_m^{(1)} - s_n^{(1)}\|_{K_1} \cdot \|s_m^{(2)}\|_{K_2} + \|s_n^{(1)}\|_{K_1} \cdot \|s_m^{(2)} - s_n^{(2)}\|_{K_2}\end{aligned}$$

holds, so that $\left(s_n^{(1)} \cdot s_n^{(2)}\right)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{H}_{K,\Omega}$ and therefore attains a limit

$$\mathcal{H}_K \ni f = \lim_{n\to\infty} s_n^{(1)} \cdot s_n^{(2)}.$$

As norm convergence implies pointwise convergence (cf. Proposition 3.6 part (5)), we have

$$f(x) = \lim_{n\to\infty} s_n^{(1)}(p_1(x)) \cdot s_n^{(2)}(p_2(x))$$

for any $x \in \mathbb{R}^d$. But we also have

$$\lim_{n\to\infty} s_n^{(1)}(p_1(x)) \cdot s_n^{(2)}(p_2(x)) = \lim_{n\to\infty} s_n^{(1)}(p_1(x)) \cdot \lim_{n\to\infty} s_n^{(2)}(p_2(x)) = f_1(p_1(x)) \cdot f_2(p_2(x))$$

by applying the same argument with respect to the norm convergence in the spaces $\mathcal{H}_{K_1}$ and $\mathcal{H}_{K_2}$. This implies

$$f_1 \cdot f_2 = f \in \mathcal{H}_K,$$

which means that Π_K is indeed a well-defined, bilinear mapping on $\mathcal{H}_{K_1} \times \mathcal{H}_{K_2}$. Given additional elements $g_1 \in \mathcal{H}_{K_1}$ and $g_2 \in \mathcal{H}_{K_2}$, we can approximate these with convergent sequences

$$\left(\tilde{s}_n^{(1)}\right)_{n\in\mathbb{N}} \subset \mathcal{S}_{K_1},\ \left(\tilde{s}_n^{(2)}\right)_{n\in\mathbb{N}} \subset \mathcal{S}_{K_2}$$

as well. The continuity of inner products and (3.16) then lead to the desired identity

$$\begin{aligned}\langle f_1 \cdot f_2, g_1 \cdot g_2\rangle_K &= \lim_{n\to\infty} \left\langle s_n^{(1)} \cdot s_n^{(2)}, \tilde{s}_n^{(1)} \cdot \tilde{s}_n^{(2)} \right\rangle_K \\ &= \lim_{n\to\infty} \left\langle s_n^{(1)}, \tilde{s}_n^{(1)} \right\rangle_{K_1} \cdot \left\langle s_n^{(2)}, \tilde{s}_n^{(2)} \right\rangle_{K_2} \\ &= \langle f_1, g_1\rangle_{K_1} \cdot \langle f_2, g_2\rangle_{K_2}.\end{aligned}$$

As a consequence of (3.13), we get

$$\mathcal{S}_K \subset \operatorname{span}_{\mathbb{R}} \left(\Pi_K \left(\mathcal{H}_{K_1} \times \mathcal{H}_{K_2}\right)\right),$$

which yields the density in $\mathcal{H}_K$.

$\underline{M \to M+1:}$ Set $\tilde{K} = \prod_{i=1}^{M} K_i$. Due to the induction basis and hypothesis, the mappings

$$\Pi_{\tilde{K}} : \mathcal{H}_{K_1} \times ... \times \mathcal{H}_{K_M} \to \mathcal{H}_{\tilde{K}}, \quad (f_1, ..., f_M) \mapsto \prod_{i=1}^{M} f_i$$

and

$$\Pi_2 : \mathcal{H}_{\tilde{K}} \times \mathcal{H}_{K_{M+1}} \to \mathcal{H}_K, \quad \left(\tilde{f}, f_{M+1}\right) \mapsto \tilde{f} \cdot f_{M+1}$$

are well-defined multilinear mappings that satisfy the property (3.18) on their respective domains. Since

$$\prod_{i=1}^{M+1} f_i = \Pi_2(\Pi_{\tilde{K}}(f_1, ..., f_M), f_{M+1})$$

holds for $(f_1, ..., f_{M+1}) \in \mathcal{H}_{K_1} \times ... \times \mathcal{H}_{K_{M+1}}$, the mapping

$$\Pi_K : \mathcal{H}_{K_1} \times ... \times \mathcal{H}_{K_{M+1}} \to \mathcal{H}_{K,\Omega}, \quad (f_1, ..., f_{M+1}) \mapsto \prod_{i=1}^{M+1} f_i$$

is well-defined and multilinear. For elements

$$f = (f_1, ..., f_{M+1}),\ g = (g_1, ..., g_{M+1}) \in \mathcal{H}_{K_1} \times ... \times \mathcal{H}_{K_{M+1}},$$

we have

$$\begin{aligned}
\langle \Pi_K(f), \Pi_K(g) \rangle_K &= \langle \Pi_2(\Pi_{\tilde{K}}(f_1, ..., f_M), f_{M+1}), \Pi_2(\Pi_{\tilde{K}}(g_1, ..., g_M), g_{M+1}) \rangle_K \\
&= \langle \Pi_{\tilde{K}}(f_1, ..., f_M), \Pi_{\tilde{K}}(g_1, ..., g_M) \rangle_{\tilde{K}} \cdot \langle f_{M+1}, g_{M+1} \rangle_{K_{M+1}} \\
&= \prod_{i=1}^{M} \langle f_i, g_i \rangle_{K_i} \cdot \langle f_{M+1}, g_{M+1} \rangle_{K_{M+1}} \\
&= \prod_{i=1}^{M+1} \langle f_i, g_i \rangle_{K_i},
\end{aligned}$$

which proves (3.18). A similar argument as in the case $M = 2$ shows that $\operatorname{span}_{\mathbb{R}}(\Pi_K(\mathcal{H}_{K_1} \times ... \times \mathcal{H}_{K_{M+1}}))$ is dense in $\mathcal{H}_K$. □

In general, the mapping Π_K from (3.17) is neither injective nor surjective, which means that the Cartesian product $\mathcal{H}_{K_1} \times ... \times \mathcal{H}_{K_M}$ is not sufficient to describe the native space $\mathcal{H}_K$ of the product kernel K. Instead, we need to use the tensor product to derive a representation of the native space $\mathcal{H}_K$ in terms of the native spaces $\mathcal{H}_{K_i}$, $i = 1, ..., M$. Since the standard tensor product of Hilbert spaces is not necessarily a Hilbert space again, we make use of a special version of the tensor product, the *Hilbert tensor product*.

Remark 3.21. In [12, Chapter I.8], the native space of the product kernel is constructed via the completion of the space of functions

$$\operatorname{span}_{\mathbb{R}}(\Pi_K(\mathcal{H}_{K_1} \times ... \times \mathcal{H}_{K_M}))$$

with respect to the inner product defined by the linear continuation of (3.18). However, we can skip this procedure in our approach, as we already know that the product kernel generates a Hilbert space of functions and by now, the existence of the Hilbert tensor product as well as its properties have been extensively studied for arbitrary Hilbert spaces. We just have to connect the dots here, similar to the approach in [103, Chapter VI]. But, in contrast to [103, Proposition 6.7], we do not have to convert the abstract elements of the constructed Hilbert tensor product into functions due to our preliminaries.

For the definition of the Hilbert tensor product, we introduce further terminology (cf. [76, Definition 2.6.3]).

Definition 3.22. Let $\mathcal{H}_1, ..., \mathcal{H}_M, Z$ be real Hilbert spaces and

$$\Pi : \mathcal{H}_1 \times ... \times \mathcal{H}_M \to Z$$

be a multilinear function on the standard Cartesian product of $\mathcal{H}_1, ..., \mathcal{H}_M$.

(i) The multilinear function Π is called `bounded` if there is a constant $c > 0$ such that

$$\|\Pi(x_1, ..., x_M)\|_Z \leq c \cdot \prod_{i=1}^{M} \|x_i\|_{\mathcal{H}_i}$$

holds for all $(x_1, ..., x_M) \in \mathcal{H}_1 \times ... \times \mathcal{H}_M$.

(ii) We call Π a `weak Hilbert-Schmidt mapping` if it is bounded and there is a constant $C > 0$ such that the estimate

$$\sum_{b_1 \in B_1} \dots \sum_{b_M \in B_M} |\langle \varphi(b_1, ..., b_n), z \rangle_Z|^2 \leq C^2 \cdot \|z\|_Z^2$$

holds for any orthonormal bases $B_1 \subset \mathcal{H}_1, ..., B_M \subset \mathcal{H}_M$ and $z \in Z$.

The Hilbert tensor product of Hilbert spaces $\mathcal{H}_1, ..., \mathcal{H}_M$ is a pair $(\mathcal{H}, \Pi)$ consisting of a Hilbert space $\mathcal{H}$ and a weak Hilbert-Schmidt mapping $\Pi : \mathcal{H}_1 \times ... \times \mathcal{H}_M \to \mathcal{H}$ that matches the properties of Π_K from Proposition 3.20. We guarantee the existence with the following theorem, which is a slight modification of [76, Theorem 2.6.4] and provides an alternative characterization in the form of a universal property.

Theorem 3.23. *Let $\mathcal{H}_1, ..., \mathcal{H}_M$ be real Hilbert spaces.*

(1) There exists a Hilbert space $\mathcal{H}$ and a multilinear mapping $\Pi : \mathcal{H}_1 \times ... \times \mathcal{H}_M \to \mathcal{H}$ such that

$$\langle \Pi(x_1, ..., x_M), \Pi(y_1, ..., y_M) \rangle_{\mathcal{H}} = \prod_{i=1}^{M} \langle x_i, y_i \rangle_{\mathcal{H}_i}$$

holds for any $(x_1, ..., x_M), (y_1, ..., y_M) \in \mathcal{H}_1 \times ... \times \mathcal{H}_M$ and

$$\mathcal{H}_0 := span_{\mathbb{R}}\{\Pi(\mathcal{H}_1 \times ... \times \mathcal{H}_M)\}$$

is dense in $\mathcal{H}$.

(2) The mapping Π from part (1) is a weak Hilbert-Schmidt mapping and satisfies the following universal property: If Z is a Hilbert space and $\tilde{\Pi} : \mathcal{H}_1 \times ... \times \mathcal{H}_M \to Z$ is a weak Hilbert-Schmidt mapping, there is a unique bounded linear map $T : \mathcal{H} \to Z$ such that $\tilde{\Pi} = T \circ \Pi$.

(3) Let $\tilde{\mathcal{H}}$ be a Hilbert space and $\tilde{\Pi} : \mathcal{H}_1 \times ... \times \mathcal{H}_M \to \tilde{\mathcal{H}}$ be a weak Hilbert-Schmidt mapping that satisfies the universal property from part (2). Then, there exists an isometric isomorphism $U : \mathcal{H} \to \tilde{\mathcal{H}}$ with $U \circ \Pi = \tilde{\Pi}$. Hence, the pair $(\tilde{\mathcal{H}}, \tilde{\Pi})$ satisfies the properties of part (1).

(4) If $(\tilde{\mathcal{H}}, \tilde{\Pi})$ satisfies the properties from part (1), there exists an isometric isomorphism $U : \mathcal{H} \to \tilde{\mathcal{H}}$ with $U \circ \varphi = \tilde{\varphi}$.

Definition 3.24. Let $\mathcal{H}_1, ..., \mathcal{H}_M$ be Hilbert spaces and $(\mathcal{H}, \Pi)$ satisfy the properties from Theorem 3.23 part (1). Then $(\mathcal{H}, \Pi)$ is called the `Hilbert tensor product` of $\mathcal{H}_1, ..., \mathcal{H}_M$, denoted by

$$(\mathcal{H}, \Pi) \simeq \bigotimes_{i=1}^{M} \mathcal{H}_i, \quad \text{or in short,} \quad \mathcal{H} \simeq \bigotimes_{i=1}^{M} \mathcal{H}_i.$$

We remark that the sign $\simeq$ in Definition 3.24 shall indicate that the Hilbert tensor product is unique up to isometric isomorphy. For further reading on the Hilbert tensor product in general, we refer to [76, Section 2.6]. If we compare the results of Proposition 3.20 with Theorem 3.23 and Definition 3.24, we can deduce the relation between $\mathcal{H}_K$ and the native spaces $\mathcal{H}_{K_i}$ immediately. We recall from Remark 3.21 that this result was already stated and proven in [103, Proposition 6.7].

Theorem 3.25. *Let $K = \prod_{i=1}^{M} K_i$ be a product kernel on $\mathbb{R}^d \simeq \mathbb{R}^d \simeq \mathbb{R}^{d_1} \times ... \times \mathbb{R}^{d_M}$. Then $(\mathcal{H}_K, \Pi_K)$ is the Hilbert tensor product of $\mathcal{H}_{K_i}$, $i = 1, ..., M$, i.e.*

$$\mathcal{H}_K \simeq \bigotimes_{i=1}^{M} \mathcal{H}_{K_i}.$$

To end this section, we want to point out that the structure of the Hilbert tensor product can be used to improve the efficiency of standard interpolation with product kernels. For example, the structure of the inner product simplifies the construction of orthonormal systems. Details on this as well as numerical examples are given in [6, Section 4-7]. A numerical investigation of product kernels can also be found in [112, Chapter 5].

4. Generalized Interpolation

With the theoretical insights of the previous chapter, we can start our analysis on the main subject of this thesis, which is the generalized interpolation in reproducing kernel Hilbert spaces. For the introduction to generalized interpolation, we discuss the general Hilbert space setting first. Thereby, it becomes clear that especially those Hilbert spaces that possess a reproducing kernel are well-suited to treat this kind of interpolation problem. Note that a similar discussion can be found in [143, Section 16.1]. We show that the well-posedness of these problems is linked to the linear independence of the considered functionals by modeling the interpolant as a linear combination of the associated Riesz representers. However, determining the linear independence of functionals is quite complicated in the native dual space, since we do not know the native space of a kernel function in general. In Section 4.1, we build on [68] and [151] to prove that the discussion about linear independence for *compactly supported distributions* can be shifted to the space of distributions. This setting covers most of the applications stated in the introductory text of the first part of this thesis, and we provide two examples that demonstrate the relevance of the derived linear independence result. Throughout Section 4.2, we assume linear independence of the considered functionals and discuss the basic properties of the resulting, well-defined interpolation operator, which turn out to be the same as in the standard interpolation case.

As a starting point of our analysis, let $\mathcal{H}$ be a real Hilbert space. Furthermore, let $\Lambda = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}^*$ be a set of bounded linear functionals and $f = (f_1, ..., f_n)^T \in \mathbb{R}^n$ be an array of given data values. The generalized interpolation problem then consists of finding an element $s \in \mathcal{H}$ such that the generalized interpolation conditions

$$\lambda_i(s) = f_i \qquad \text{for all } i \in \{1, ..., n\} \tag{4.1}$$

are fulfilled. As in the standard interpolation case (cf. Chapter 2), we have to restrict ourselves to a suitable subset $S_\Lambda \subset \mathcal{H}$, if the problem (4.1) is supposed to have a unique solution. For every $i \in \{1, ..., n\}$, let $v_i \in \mathcal{H}$ denote the Riesz representer of λ_i, i.e.

$$\lambda_i(g) = \langle g, v_i \rangle_{\mathcal{H}} \qquad \text{for all } g \in \mathcal{H}, i \in \{1, ..., n\},$$

and define S_Λ as the span of all the Riesz representers:

$$S_\Lambda := \operatorname{span}_{\mathbb{R}} \{v_1, ..., v_n\} \tag{4.2}$$

If we assume $s = \sum_{i=1}^n c_i \cdot v_i \in S_\Lambda$, the generalized interpolation problem (4.1) leads to the linear system

$$A \cdot c = f, \tag{4.3}$$

where the system matrix A is given by

$$A = \begin{pmatrix} \lambda_1(v_1) & \dots & \lambda_1(v_n) \\ \vdots & \ddots & \vdots \\ \lambda_n(v_1) & \dots & \lambda_n(v_n) \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Due to our choice of the basis functions v_i, $i \in \{1, ..., n\}$, we can write A as

$$A = \begin{pmatrix} \langle v_1, v_1 \rangle_{\mathcal{H}} & \dots & \langle v_1, v_n \rangle_{\mathcal{H}} \\ \vdots & \ddots & \vdots \\ \langle v_n, v_1 \rangle_{\mathcal{H}} & \dots & \langle v_n, v_n \rangle_{\mathcal{H}} \end{pmatrix} = \begin{pmatrix} \langle \lambda_1, \lambda_1 \rangle_{\mathcal{H}^*} & \dots & \langle \lambda_1, \lambda_n \rangle_{\mathcal{H}^*} \\ \vdots & \ddots & \vdots \\ \langle \lambda_n, \lambda_1 \rangle_{\mathcal{H}^*} & \dots & \langle \lambda_n, \lambda_n \rangle_{\mathcal{H}^*} \end{pmatrix},$$

so that A is the Gram matrix of Λ with respect to $\langle \cdot, \cdot \rangle_{\mathcal{H}^*}$. Hence, A is regular if and only if Λ is linearly independent. In this case, the problem (4.1) restricted to S_Λ is well-posed, so that we can find a unique

element $s = \sum_{i=1}^{n} c_i \cdot v_i \in S_\Lambda$ that matches the desired evaluations on Λ in theory. But, even if the Gram matrix can be computed quite efficiently, we need to determine the Riesz representers v_i, $i \in \{1, ..., n\}$, in order to make further computations with the interpolant s. This is a major problem in the general Hilbert space setting since the determination of the Riesz representers is very difficult.

However, in reproducing kernel spaces $\mathcal{H}_K$, we have already shown that the correspondence between linear functionals and their Riesz representers is described by the mapping

$$\varphi^{-1} : \mathcal{H}_K^* \to \mathcal{H}_K,\ \lambda \mapsto \lambda^y K(\cdot, y), \tag{4.4}$$

see Theorem 3.12 and Remark 3.13. This means that the determination of the Riesz representers is relatively easy in this special setting, so we restrict our theoretical analysis to reproducing kernel Hilbert spaces in this thesis, although large parts of the results could be transferred to the general Hilbert space setting.

In the setting of a reproducing kernel Hilbert space $\mathcal{H}_K$, we denote the space of interpolants from (4.2) as

$$S_{K,\Lambda} := \operatorname{span}_{\mathbb{R}} \{\lambda_1^y K(\cdot, y), ..., \lambda_n^y K(\cdot, y)\} \subset \mathcal{H}_K$$

and the respective system matrix from (4.3) as

$$A_{K,\Lambda} := \begin{pmatrix} \lambda_1(\lambda_1^y K(\cdot, y)) & \dots & \lambda_1(\lambda_n^y K(\cdot, y)) \\ \vdots & \ddots & \vdots \\ \lambda_n(\lambda_1^y K(\cdot, y)) & \dots & \lambda_n(\lambda_n^y K(\cdot, y)) \end{pmatrix}$$

for a given set $\Lambda = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}_K^*$ of functionals. We summarize our previous discussion in the following theorem (cf. [143, Theorem 16.1]).

Theorem 4.1. *Let $\mathcal{H}_K$ be a reproducing kernel Hilbert space. Moreover, let $\Lambda = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}_K^*$ and $f_1, ..., f_n \in \mathbb{R}$. If Λ is linearly independent, there is a unique $s \in S_{K,\Lambda}$ that satisfies the interpolation conditions (4.1).*

Remark 4.2. The generalized interpolation problem has also been investigated for a more general Banach space setting in [86]. However, the numerical tests in this paper only consider the Hilbert space setting, where the function spaces are chosen as $L^2(\Omega)$ or the Sobolev space $H^1(\Omega)$ for suitable domains $\Omega \subset \mathbb{R}^d$.

4.1. Linear independence

In order to have a unique solution to the generalized interpolation problem (4.1), it is essential that the considered set of functionals $\Lambda \subset \mathcal{H}_K^*$ is linearly independent. For example, if the kernel K is positive definite, part (2) of Proposition 3.6 ensures that the set

$$\left\{\delta_{x_1}, ..., \delta_{x_n}\right\} \subset \mathcal{H}_K^*$$

of point evaluation functions is linearly independent for any finite set $X = \{x_1, ..., x_n\} \subset \mathbb{R}^d$ of pairwise distinct points. Given other functionals, it is not a trivial task to decide whether Λ is linearly independent or not. In general, we do not know the native space explicitly, which means that the analysis in the dual space $\mathcal{H}_K^*$ is limited. Instead, one tries to shift the discussion about linear independence to better-known spaces.

Here, we follow the ideas of [68], [151] and discuss the case that Λ consists of *distributions with compact support*. The space $\mathscr{E}'(\mathbb{R}^d)$ of compactly supported distributions is a large class of functionals and, to our advantage, there is a large number of theoretical results and analytical tools available, see Section A.4. For the following discussion, the intersection

$$\mathcal{H}_K^* \cap \mathscr{E}'(\mathbb{R}^d)$$

denotes the set of all functionals $\lambda \in \mathcal{H}_K^*$ that are also well-defined and continuous on the space $\mathscr{D}(\mathbb{R}^d)$ of *test functions* (cf. Definition A.17) and compactly supported. Note that we have

$$\mathcal{S}_K^{(*)} = \operatorname{span}_{\mathbb{R}} \{\delta_x \mid x \in \mathbb{R}^d\} \ \subset\ \mathcal{H}_K^* \cap \mathscr{E}'(\mathbb{R}^d). \tag{4.5}$$

Our first goal is to find an alternative representation of the inner product $\langle \cdot, \cdot \rangle_K$ on $\mathcal{H}_K^* \cap \mathscr{E}'(\mathbb{R}^d)$ that builds a bridge between the two spaces of functionals. To this end, we restrict to translation-invariant kernels, as introduced in Subsection 2.1.1, and borrow the ideas from Bochner's theorem (cf. [143, Section 6.2]) to derive a suitable integral representation. Throughout the next steps, we use that the Fourier transform of the generating function is integrable under the assumptions of Theorem 2.7 (cf. [143, Corollary 6.12]).

Proposition 4.3. *Let $\Phi \in \mathscr{C}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ such that $K = \Phi(\cdot - \cdot)$ is positive definite on $\mathbb{R}^d$. Then $\mathcal{F}\Phi \in L^1(\mathbb{R}^d)$ holds.*

In the setting of Proposition 4.3, we can use the Fourier inversion formula (cf. Theorem A.15) to derive

$$\begin{aligned}\langle \delta_x, \delta_y \rangle_K = \Phi(x-y) &= (2\pi)^{-d/2} \cdot \int_{\mathbb{R}^d} e^{i \cdot \langle x-y, \omega \rangle_2} \cdot \mathcal{F}\Phi(\omega)\, d\omega \\ &= (2\pi)^{-d/2} \cdot \int_{\mathbb{R}^d} e^{i \cdot \langle x, \omega \rangle_2} \cdot e^{-i \cdot \langle y, \omega \rangle_2} \cdot \mathcal{F}\Phi(\omega)\, d\omega \\ &= (2\pi)^{d/2} \cdot \int_{\mathbb{R}^d} \overline{\mathcal{F}_{\mathscr{E}'} \delta_x(\omega)} \cdot \mathcal{F}_{\mathscr{E}'} \delta_y(\omega) \cdot \mathcal{F}\Phi(\omega)\, d\omega\end{aligned}$$

for $x, y \in \mathbb{R}^d$, where $\mathcal{F}_{\mathscr{E}'}$ is the Fourier-Laplace transform on $\mathscr{E}'(\mathbb{R}^d)$, see Definition A.22 and Example A.23. By multilinear continuation, we extend this representation to $\mathcal{S}_K^{(\star)}$ from (4.5), which results in

$$\langle \lambda, \mu \rangle_K = (2\pi)^{d/2} \cdot \int_{\mathbb{R}^d} \overline{\mathcal{F}_{\mathscr{E}'} \lambda(\omega)} \cdot \mathcal{F}_{\mathscr{E}'} \mu(\omega) \cdot \mathcal{F}\Phi(\omega)\, d\omega \qquad \text{for all } \lambda, \mu \in \mathcal{S}_K^{(\star)}. \tag{4.6}$$

To further extend the domain of (4.6), we assume that the considered functionals can be approximated by elements from $\mathcal{S}_K^{(*)}$ in $\mathcal{H}_K^*$ and $\mathscr{E}'(\mathbb{R}^d)$ simultaneously.

Lemma 4.4. *Let $\lambda, \mu \in \mathcal{H}_K^* \cap \mathscr{E}'(\mathbb{R}^d)$. Assume that there are sequences $(\lambda_n)_{n \in \mathbb{N}}, (\mu_n)_{\mathbb{N}}$ in $\mathcal{S}_K^{(*)}$ that satisfy the following conditions:*

(i) We have $\|\lambda_n - \lambda\|_K \xrightarrow{n \to \infty} 0$ and $\|\mu_n - \mu\|_K \xrightarrow{n \to \infty} 0$.

(ii) The convergence $\lambda_n \xrightarrow{n \to \infty} \lambda$ and $\mu_n \xrightarrow{n \to \infty} \mu$ holds in the sense of distributions.

(iii) There are constants $C_\lambda, C_\mu > 0$ such that

$$|\mathcal{F}_{\mathscr{E}'} \lambda_n(\omega)| \leq C_\lambda, \quad |\mathcal{F}_{\mathscr{E}'} \mu_n(\omega)| \leq C_\mu \qquad \text{for all } n \in \mathbb{N}, \omega \in \mathbb{R}^d.$$

Then we have

$$\langle \lambda, \mu \rangle_K = (2\pi)^{d/2} \cdot \int_{\mathbb{R}^d} \overline{\mathcal{F}_{\mathscr{E}'} \lambda(\omega)} \cdot \mathcal{F}_{\mathscr{E}'} \mu(\omega) \cdot \mathcal{F}\Phi(\omega)\, d\omega.$$

Proof. Due to condition (ii) and Lemma A.25 part (2), we have

$$\mathcal{F}_{\mathscr{E}'} \lambda_n(\omega) \xrightarrow{n \to \infty} \mathcal{F}_{\mathscr{E}'} \lambda(\omega), \quad \mathcal{F}_{\mathscr{E}'} \mu_n(\omega) \xrightarrow{n \to \infty} \mathcal{F}_{\mathscr{E}'} \mu(\omega) \qquad \text{for all } \omega \in \mathbb{R}^d.$$

Moreover, we use (iii) to estimate

$$|\overline{\mathcal{F}_{\mathscr{E}'} \lambda_n(\omega)} \cdot \mathcal{F}_{\mathscr{E}'} \mu_n(\omega) \cdot \mathcal{F}\Phi(\omega)| \leq C_\lambda \cdot C_\mu \cdot \mathcal{F}\Phi(\omega) \qquad \text{for all } n \in \mathbb{N}, \omega \in \mathbb{R}^d,$$

where Proposition 4.3 ensures that $C_\lambda \cdot C_\mu \cdot \Phi \in L^1(\mathbb{R}^d)$. The representation (4.6) in combination with *Lebesgue's dominated convergence theorem* then yields

$$\begin{aligned}\langle \lambda, \mu \rangle_K = \lim_{n \to \infty} \langle \lambda_n, \mu_n \rangle_K &= \lim_{n \to \infty} (2\pi)^{d/2} \cdot \int_{\mathbb{R}^d} \overline{\mathcal{F}_{\mathscr{E}'} \lambda_n(\omega)} \cdot \mathcal{F}_{\mathscr{E}'} \mu_n(\omega) \cdot \mathcal{F}\Phi(\omega)\, d\omega \\ &= (2\pi)^{d/2} \cdot \int_{\mathbb{R}^d} \overline{\mathcal{F}_{\mathscr{E}'} \lambda(\omega)} \cdot \mathcal{F}_{\mathscr{E}'} \mu(\omega) \cdot \mathcal{F}\Phi(\omega)\, d\omega.\end{aligned}$$

□

Under the previous assumptions, we can shift the discussion about linear independence to the space of distributions.

Theorem 4.5. *Let $K = \Phi(\cdot - \cdot)$ be a translation-invariant positive definite kernel function with $\Phi \in \mathscr{C}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ and let $\Lambda = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}_K^* \cap \mathscr{E}'(\mathbb{R}^d)$, such that every functional in Λ satisfies the conditions of Lemma 4.4. Then Λ is linearly independent in $\mathcal{H}_K^*$ if and only if Λ is linearly independent in $\mathscr{E}'(\mathbb{R}^d)$.*

Proof. For $c = (c_1, ..., c_n)^T \in \mathbb{R}^n$, we use the inner product representation of Lemma 4.4 to get

$$c^T \cdot A_{K,\Lambda} \cdot c = (2\pi)^{d/2} \cdot \int_{\mathbb{R}^d} \Big| \sum_{i=1}^n c_i \cdot \mathcal{F}_{\mathscr{E}'} \lambda_i(\omega) \Big|^2 \cdot \mathcal{F}\Phi(\omega)\, d\omega.$$

According to Theorem 2.7, $\mathcal{F}\Phi$ is non-negative and non-vanishing. Hence, there is an open set $U \subset \mathbb{R}^d$ such that $\mathcal{F}\Phi(\omega) > 0$ for $\omega \in U$. If $c^T \cdot A_{K,\Lambda} \cdot c = 0$, which is equivalent to

$$\sum_{i=1}^n c_i \cdot \lambda_i = 0 \tag{4.7}$$

in $\mathcal{H}_K^*$, this implies that the function

$$\sum_{i=1}^n c_i \cdot \mathcal{F}_{\mathscr{E}'} \lambda_i,$$

which is an entire function according to Theorem A.24, vanishes on U. By applying the *identity theorem* for holomorphic functions, we see that it must vanish on the whole space $\mathbb{C}^d$, so that (4.7) also holds in the sense of distributions due to Lemma A.25 part (1). If Λ is linearly independent as a set of distributions, we must have $c = 0$, so that Λ is linearly independent in $\mathcal{H}_K^*$ as well. For the reverse implication, we can simply backtrack the previous steps. □

The main advantage of this approach is that we can now make use of the well-known space of test functions in our analysis. The next application demonstrates the utility of Theorem 4.5.

Cell average functionals

Given a compact set $E \subset \mathbb{R}^d$ with Lebesgue measure $\mathrm{vol}(E) > 0$, we define the respective *cell average functional* as

$$\lambda_E(f) := \frac{1}{\mathrm{vol}(E)} \cdot \int_E f(x)\, dx \qquad \text{for } f \in \mathscr{D}(\mathbb{R}^d), \tag{4.8}$$

which is a distribution with compact support. If $K \in \mathscr{C}(\mathbb{R}^d \times \mathbb{R}^d)$, we have the inclusion $\mathcal{H}_K \subset \mathscr{C}(\mathbb{R}^d)$ due to Proposition 3.14 part (2), so that λ_E is well-defined on $\mathcal{H}_K$. Moreover, the estimate

$$|\lambda_E(f)| \le \frac{1}{\mathrm{vol}(E)} \cdot \int_E |f(x)|\, dx = \frac{1}{\mathrm{vol}(E)} \cdot \int_E |\langle f, K(\cdot, x)\rangle_K|\, dx \le \max_{x \in E} \sqrt{K(x,x)} \cdot \|f\|_K \quad \text{for all } f \in \mathcal{H}_K$$

shows that $\lambda_E \in \mathcal{H}_K^*$ holds as well. We construct an approximating sequence in $\mathcal{S}_K^{(*)}$ as follows: For $n \in \mathbb{N}$, there exist points

$$x_1^{(n)}, ..., x_{M_n}^{(n)} \in E \qquad \text{such that} \qquad E \subset \bigcup_{j=1}^{M_n} B_{1/n}\left(x_j^{(n)}\right), \tag{4.9}$$

i.e. E is covered by open balls with radius $1/n$. Then, the modified sets

$$E_1^{(n)} := B_{1/n}\left(x_1^{(n)}\right) \cap E, \qquad E_j^{(n)} := \left(B_{1/n}\left(x_j^{(n)}\right) \cap E\right) \setminus \bigcup_{j=1}^{j-1} E_j^{(n)} \quad \text{for } j = 2, ..., M_n$$

decompose E into measurable subsets. To make sure that each subset is non-empty, we assume that M_n is the minimal natural number such that (4.9) can hold. The resulting sequence of functionals is then given by

$$\lambda_n := \sum_{j=1}^{M_n} \frac{\mathrm{vol}(E_j^{(n)})}{\mathrm{vol}(E)} \cdot \delta_{x_j^{(n)}} \in \mathcal{S}_K^{(*)} \qquad \text{for } n \in \mathbb{N}. \tag{4.10}$$

We show that $(\lambda_n)_{n\in\mathbb{N}}$ has the desired properties.

Lemma 4.6. *Let $K \in \mathscr{C}(\mathbb{R}^d \times \mathbb{R}^d)$ and $E \subset \mathbb{R}^d$ be compact with $\mathrm{vol}(E) > 0$. Then λ_E, as defined in (4.8), satisfies the conditions of Lemma 4.4.*

Proof. Define λ_n as in (4.10) for $n \in \mathbb{N}$.

(i) Since $K \in \mathscr{C}(\mathbb{R}^d \times \mathbb{R}^d)$, the mapping $x \mapsto K(\cdot, x)$ is uniformly continuous on E. Given $\varepsilon > 0$, we can find $\delta > 0$ such that

$$\|K(\cdot, x) - K(\cdot, y)\|_K < \varepsilon \qquad \text{for} \qquad x, y \in E \text{ with } \|x - y\|_2 < \delta.$$

Now choose $N \in \mathbb{N}$ such that $1/N < \delta$. Then we have

$$\begin{aligned}
|\lambda_n(f) - \lambda_E(f)| &= \frac{1}{\mathrm{vol}(E)} \cdot \left| \sum_{j=1}^{M_n} \mathrm{vol}(E_j^{(n)}) \cdot f(x_j^{(n)}) - \int_E f(x)\, dx \right| \\
&\le \frac{1}{\mathrm{vol}(E)} \cdot \sum_{j=1}^{M_n} \int_{E_j^{(n)}} |f(x_j^{(n)}) - f(x)|\, dx \\
&\le \frac{1}{\mathrm{vol}(E)} \cdot \sum_{j=1}^{M_n} \int_{E_j^{(n)}} \|K(\cdot, x_j^{(n)}) - K(\cdot, x)\|_K\, dx \cdot \|f\|_K \\
&< \varepsilon \cdot \|f\|_K
\end{aligned}$$

for all $f \in \mathcal{H}_K$, $n \ge N$.

(ii) We can use the same arguments as in part (i) to show that

$$\lambda_n(f) \xrightarrow{n\to\infty} \lambda_E(f) \qquad \text{for all } f \in \mathscr{D}(\mathbb{R}^d).$$

(iii) For $n \in \mathbb{N}$ and $\omega \in \mathbb{R}^d$, we have

$$|\mathcal{F}_{\mathscr{E}'}\lambda_n(\omega)| = \left| \sum_{j=1}^{M_n} \frac{\mathrm{vol}(E_j^{(n)})}{\mathrm{vol}(E)} \cdot e^{-i\cdot\langle x_j^{(n)},\, \omega\rangle_2} \right| \le \sum_{j=1}^{M_n} \frac{\mathrm{vol}(E_j^{(n)})}{\mathrm{vol}(E)} = 1.$$

□

With this, we can formulate a feasible criterion that ensures linear independence of the cell average functionals based on the interior of the considered sets.

Theorem 4.7. *Let $K = \Phi(\cdot - \cdot)$ with $\Phi \in \mathscr{C}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ and let $E_1, ..., E_n \subset \mathbb{R}^d$ be compact subsets with positive Lebesgue measure, such that*

$$\mathring{E}_i \setminus \bigcup_{i\neq k} E_k \neq \emptyset \qquad \textit{for } i = 1, ..., n.$$

Then $\Lambda = \{\lambda_{E_1}, ..., \lambda_{E_n}\}$ is linearly independent in $\mathcal{H}_K^*$.

Proof. Due to Lemma 4.6 and Theorem 4.5, it suffices to show that Λ is linearly independent in the sense of distributions. For every $i \in \{1, ..., n\}$, we can find $x_i \in E_i$ and $r_i > 0$, such that

$$B_{r_i}(x_i) \subset E_i \setminus \bigcup_{i \neq k} E_k.$$

Choose $\varepsilon < \min_{i=1,...,n} r_i$ and define the functions

$$f_i(x) := \mathrm{vol}(E_i) \cdot g_\varepsilon(x - x_i) \qquad \text{for } x \in \mathbb{R}^d, i = 1, ..., n,$$

where g_ε is the L^1-normalized function defined in Example A.18. By construction, we then have

$$\lambda_{E_i}(f_j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases}$$

If $\sum_{i=1}^n c_i \cdot \lambda_{E_i} = 0$ with $c_i \in \mathbb{R}$ holds in $\mathscr{E}'(\mathbb{R}^d)$, we get

$$0 = \sum_{i=1}^n c_i \cdot \lambda_{E_i}(f_j) = c_j \qquad \text{for } j = 1, ..., n,$$

so that Λ is linearly independent in $\mathscr{E}'(\mathbb{R}^d)$. □

For example, cell average functionals occur in *finite volume methods*, where the solution is reconstructed from its cell average values that are estimated for certain time steps (cf. [71], [128]). In [1], the computational domain $\Omega \subset \mathbb{R}^2$ is decomposed into conforming triangulations. Given a finite point set $X \subset \Omega$, this can be done via the *Delauney triangulation* (cf. [69, Section 2.2]), see Figure 4.1. Note that the corresponding triangles satisfy the conditions of Theorem 4.7. A convergence analysis for this interpolation setting can be found in [142].

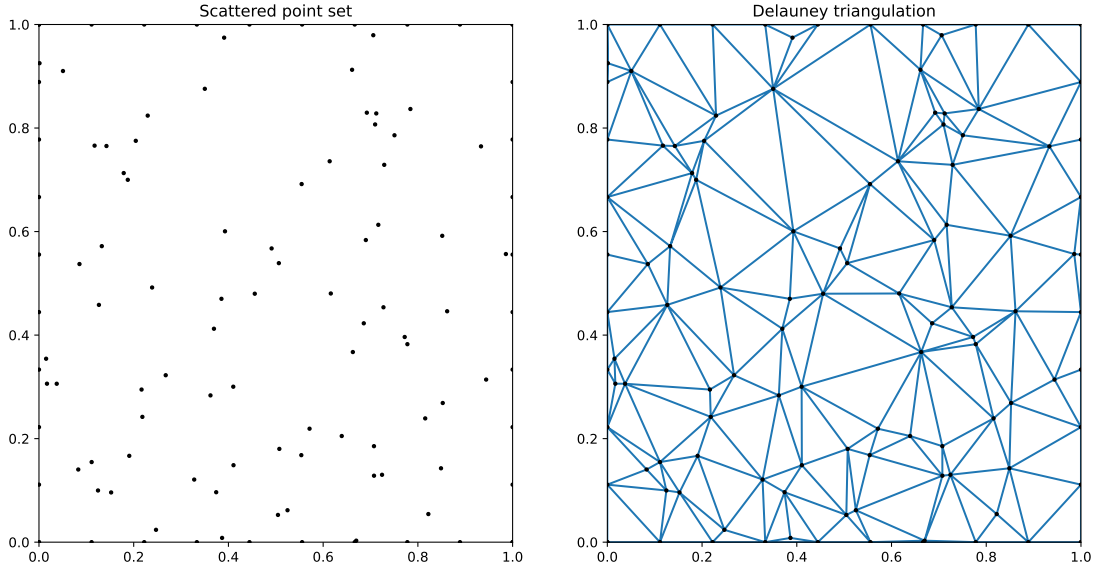

Figure 4.1.: Scattered point set in the domain $[0, 1]^2$ (left) and its Delauney triangulation (right)

4.2. Interpolation operator

For our further analysis, we assume that $\Lambda = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}_K^*$ is linearly independent. Then, Theorem 4.1 states that for given data values $f_1, ..., f_n \in \mathbb{R}$, there is a unique interpolant $s \in S_{K,\Lambda}$ that satisfies the interpolation conditions (4.1). In applications, the given data values are usually not some arbitrary

numbers but rather come from a function we want to interpolate. More precisely, we assume that there is a function $f \in \mathcal{H}_K$ such that

$$f_i = \lambda_i(f) \qquad \text{for all } i \in \{1, ..., n\}.$$

Under these assumptions, s is called the *(unique) interpolant to f on Λ* and denoted by $s \equiv s_{f,\Lambda}$. Thus, we can define a (generalized) interpolation operator on the whole space $\mathcal{H}_K$.

Definition 4.8. Let $\Lambda = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}_K^*$ be a linearly independent set of functionals. We define the respective `interpolation operator` on $\mathcal{H}_K$ as

$$I_{K,\Lambda} : \mathcal{H}_K \to S_{K,\Lambda},\ f \mapsto s_{f,\Lambda}$$

that maps every function $f \in \mathcal{H}_K$ to its unique interpolant $s_{f,\Lambda}$ on Λ.

It is clear that $I_{K,\Lambda}$ is a linear operator. The next lemma allows the identification as a projection operator (cf. [70, Corollary 8.28 & 8.29]).

Lemma 4.9. *Given $\Lambda = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}_K^*$, where Λ is not necessarily linearly independent, we can write the orthogonal complement $S_{K,\Lambda}^\perp$ of $S_{K,\Lambda}$ as*

$$S_{K,\Lambda}^\perp = \{f \in \mathcal{H}_K \mid \lambda_i(f) = 0 \text{ for all } i \in \{1, ..., n\}\ \}.$$

Proof. We have $f \in S_{K,\Lambda}^\perp$ if and only if f is orthogonal to the elements $\lambda_i^y K(\cdot, y)$ for all $i \in \{1, ..., n\}$. Due to the generalized reproduction property from Theorem 3.12, this can be rewritten as

$$0 = \langle f, \lambda_i^y K(\cdot, y)\rangle_{\mathcal{H}_K} = \lambda_i(f) \qquad \text{for all } i \in \{1, ..., n\}.$$

□

Theorem 4.10. *Let $\Lambda = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}_K^*$ be linearly independent. Moreover, let $\mathcal{P}_{S_{K,\Lambda}}$ denote the orthogonal projection operator onto $S_{K,\Lambda}$, defined in Definition A.10. Then we have the operator equality $\mathcal{P}_{S_{K,\Lambda}} = I_{K,\Lambda}$. In particular, $I_{K,\Lambda}(f)$ is the unique best approximation to f from $S_{K,\Lambda}$, i.e.*

$$\|f - I_{K,\Lambda}(f)\|_K = \inf_{s \in S_{K,\Lambda}} \|f - s\|_K \qquad \text{for all } f \in \mathcal{H}_K,$$

and $I_{K,\Lambda}$ is a bounded operator with operator norm

$$\|I_{K,\Lambda}\|_K = \sup_{f \in \mathcal{H}_K \setminus \{0\}} \frac{\|I_{K,\Lambda}(f)\|_K}{\|f\|_K} = 1.$$

Proof. Let $f \in \mathcal{H}_K$. By Definition 4.8, we have

$$\lambda_i(f - I_{K,\Lambda}(f)) = \lambda_i(f) - \lambda_i(I_{K,\Lambda}(f)) = 0 \qquad \text{for all } i \in \{1, ..., n\}.$$

Due to Lemma 4.9, this implies $f - I_{K,\Lambda}(f) \in S_{K,\Lambda}^\perp$. The assertion now follows from Theorem A.9 part (1). □

This characterization of the interpolation operator simplifies our further analysis, as projection operators in Hilbert spaces have already been well-investigated. For example, if $B = \{b_1, ..., b_n\}$ is an orthonormal basis of $S_{K,\Lambda}$, the interpolant can be written as

$$I_{K,\Lambda}(f) = \sum_{i=1}^{n} \langle f, b_i\rangle_K \cdot b_i \qquad \text{for all } f \in \mathcal{H}_K, \tag{4.11}$$

see Theorem A.9 part (3). Hence, the computation of orthonormal bases also plays an important role in our approximation algorithms. We discuss the construction of orthonormal systems in Chapter 6.

Another implication of Theorem 4.10 is that, given $f \in \mathcal{H}_K$, the element $I_{K,\Lambda}(f)$ minimizes the native space norm amongst all possible interpolants to f from $\mathcal{H}_K$ (cf. [70, Corollary 8.30]). This property is often referred to as *optimal recovery* (cf. [143, Theorem 16.1]).

Corollary 4.11. *Let $\Lambda = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}_K^*$ be linearly independent. For $f \in \mathcal{H}_K$, define the set of all possible interpolants on Λ as*

$$\mathcal{I}_{f,\Lambda} := \{g \in \mathcal{H}_K \mid \lambda_i(f) = \lambda_i(g) \text{ for all } i \in \{1, ..., n\} \} \subset \mathcal{H}_K.$$

Then $I_{K,\Lambda}(f)$ minimizes the native space norm on $\mathcal{I}_{f,\Lambda}$, i.e.

$$\|I_{K,\Lambda}(f)\|_K < \|g\|_K \qquad \text{for all } g \in \mathcal{I}_{f,\Lambda} \setminus \{I_{K,\Lambda}(f)\}.$$

Proof. If $g \in \mathcal{I}_{f,\Lambda} \setminus \{I_{K,\Lambda}(f)\}$, we must have

$$I_{K,\Lambda}(f) = I_{K,\Lambda}(g),$$

and therefore $g - I_{K,\Lambda}(f) \in S_{K,\Lambda}^{\perp} \setminus \{0\}$. Then Pythagoras' theorem gives

$$\|g\|_K^2 = \|g - I_{K,\Lambda}(f)\|_K^2 + \|I_{K,\Lambda}(f)\|_K^2 > \|I_{K,\Lambda}(f)\|_K^2.$$

□

Note that the proof of Corollary 4.11 also shows that

$$\mathcal{I}_{f,\Lambda} = I_{K,\Lambda}(f) + S_{K,\Lambda}^{\perp} := \left\{ I_{K,\Lambda}(f) + s \;\middle|\; s \in S_{K,\Lambda}^{\perp} \right\} \qquad \text{for all } f \in \mathcal{H}_K.$$

Hence, the previous corollary can be interpreted in the following way: For our interpolation problem, which just consists of matching the evaluation of the functionals in Λ with respect to a certain function $f \in \mathcal{H}_K$, it is sufficient to restrict to the set $S_{K,\Lambda}$. Every other interpolant $g \in \mathcal{I}_{f,\Lambda} \setminus \{I_{K,\Lambda}(f)\}$ contains some additional information, which is unnecessary for the interpolation process and increases the norm in comparison to $I_{K,\Lambda}(f)$.

5. Convergence of the Interpolation Method

In the previous chapter, we have discussed the generalized interpolation problem with respect to a given fixed set $\Lambda \subset \mathcal{H}_K^*$ of linearly independent functions. However, the interpolation method often is a dynamic process, where we successively increase the number of interpolation points to improve the approximation quality. Motivated by this setting, we want to derive a class of functions and respective convergence criteria, such that the contained functions can be approximated arbitrarily well by the interpolation method. In kernel-based approximation theory, it is common to derive error estimates in terms of the *power function* or *fill distance*. For the standard interpolation problem, such error estimates have been derived for popular kernels in [90],[152], an overview is given in the book chapter [143, Chapter 11]. Here, we provide generalizations of these two error indicators and derive corresponding convergence criteria for the generalized interpolation method in Section 5.1 & 5.2. Note that the generalized version of the power function was already introduced in [53], [54]. In order to build a stronger connection to the standard interpolation theory, we introduce the notion of *parametrizations* for subsets of functionals in Section 5.3, which simplifies the analysis and results in known convergence results for the standard interpolation case. Ultimately, the results of this chapter serve as a foundation for Chapter 7, where we discuss suitable point selection strategies that guarantee the convergence of the interpolation method.

For our analysis, let $\Gamma \subset \mathcal{H}_K^*$ be a fixed set of functionals, which serves as the domain for the selection of the interpolation functionals. We consider nested sequences $(\Lambda_n)_{n\in\mathbb{N}}$ of finite linearly independent subsets of Γ, i.e.

(i) $\Lambda_n \subset \Gamma$ is finite and linearly independent for all $n \in \mathbb{N}$

(ii) $\Lambda_n \subset \Lambda_{n+1}$ for all $n \in \mathbb{N}$.

In this setting, we analyze the corresponding sequence of interpolants

$$\left(I_{K,\Lambda_n}(f)\right)_{n\in\mathbb{N}}, \tag{5.1}$$

to $f \in \mathcal{H}_K$, which is a subset of the closed subspace

$$\mathcal{H}_{K,\Gamma} := \overline{\operatorname{span}_{\mathbb{R}} \{\lambda^y K(\cdot, y) \mid \lambda \in \Gamma\}} \subset \mathcal{H}_K.$$

Hence, if the sequence (5.1) converges, the limit is an element of $\mathcal{H}_{K,\Gamma}$ as well. Naturally, we wish to have the convergence

$$\|f - I_{K,\Lambda_n}(f)\|_K \xrightarrow{n\to\infty} 0,$$

which means that we have to restrict our convergence analysis to the case $f \in \mathcal{H}_{K,\Gamma}$. In the following, we derive conditions on Γ and $(\Lambda_n)_{n\in\mathbb{N}}$ that guarantee convergence in $\mathcal{H}_{K,\Gamma}$.

Remark 5.1. Before we continue, we remark that most of the results from this chapter do not rely on the assumptions (ii) that the subsets are nested. However, we want to introduce this notion right from the start of our convergence analysis, as we only deal with nested sequences in the later parts of this thesis. In practical cases, nested sequences are favorable, since we can use pre-computed information with the help of suitable update strategies. More details on update strategies for the generalized interpolation method are given in Section 6.3.

5.1. Power function

For the first convergence criterion, we initially consider a finite linearly independent set $\Lambda \subset \mathcal{H}_K^*$. For any $\lambda \in \mathcal{H}_K^*$, we can define the pointwise error functional $\varepsilon_{\Lambda,\lambda}$ (cf. [70, Subsection 8.3.2]) as

$$\varepsilon_{\Lambda,\lambda} : \mathcal{H}_K \to \mathbb{R},\ f \mapsto \lambda(f) - \lambda(I_{K,\Lambda}(f)).$$

Since $I_{K,\Lambda}$ is a bounded operator, we have $\varepsilon_{\Lambda,\lambda} \in \mathcal{H}_K^*$, and therefore

$$\|\varepsilon_{\Lambda,\lambda}\|_K = \sup_{f \in \mathcal{H}_K \setminus \{0\}} \frac{|\varepsilon_{\Lambda,\lambda}(f)|}{\|f\|_K} < \infty \qquad \text{for all } \lambda \in \mathcal{H}_K^*.$$

By definition, we have

$$|\lambda(f) - \lambda(I_{K,\Lambda})(f)| \leq \|\varepsilon_{\Lambda,\lambda}\|_K \cdot \|f\|_K \qquad \text{for all } f \in \mathcal{H}_K, \lambda \in \mathcal{H}_K^*, \tag{5.2}$$

so that these operator norms are useful to estimate the pointwise error of the interpolation method. We collect the norm values as a function of $\lambda \in \mathcal{H}_K^*$, the so-called *power function* (cf. [114, Section 1]).

Definition 5.2. Let $\Lambda \subset \mathcal{H}_K^*$ be finite and linearly independent. We define the `power function` with respect to Λ as

$$P_\Lambda(\lambda) := \|\varepsilon_{\Lambda,\lambda}\|_K \qquad \text{for all } \lambda \in \mathcal{H}_K^*.$$

In the format of Definition 5.2, the power function is quite unhandily. For our analysis and later implementations, we use a different representation (cf. [114, Lemma 2.3]).

Lemma 5.3. *The power function w.r.t $\Lambda \subset \mathcal{H}_K^*$ can be written as*

$$P_\Lambda(\lambda) = \|\lambda^y K(\cdot, y) - I_{K,\Lambda}(\lambda^y K(\cdot, y))\|_K \qquad \textit{for all } \lambda \in \mathcal{H}_K^*.$$

Proof. We set $n = |\Lambda|$. Let $\lambda \in \mathcal{H}_K^*$ and $B = \{b_1, ..., b_n\}$ be an orthonormal basis of $S_{K,\Lambda}$. According to formula (4.11), we can write

$$I_{K,\Lambda}(f) = \sum_{i=1}^{n} \langle f, b_i \rangle_K \cdot b_i \qquad \text{for all } f \in \mathcal{H}_K.$$

We can use this identity and the generalized reproduction property from Theorem 3.12 to get

$$\begin{aligned}
\varepsilon_{\Lambda,\lambda}(f) &= \lambda(f) - \lambda\left(\sum_{i=1}^{n} \langle f, b_i \rangle_K \cdot b_i\right) \\
&= \langle f, \lambda^y K(\cdot, y) \rangle_K - \sum_{i=1}^{n} \langle f, b_i \rangle_K \cdot \langle \lambda^y K(\cdot, y), b_i \rangle_K \\
&= \left\langle f, \lambda^y K(\cdot, y) - \sum_{i=1}^{n} \langle \lambda^y K(\cdot, y), b_i \rangle_K \cdot b_i \right\rangle_K \\
&= \langle f, \lambda^y K(\cdot, y) - I_{K,\Lambda}(\lambda^y K(\cdot, y)) \rangle_K
\end{aligned}$$

for all $f \in \mathcal{H}_K$. Hence, $\lambda^y K(\cdot, y) - I_{K,\Lambda}(\lambda^y K(\cdot, y))$ is the Riesz representer of $\varepsilon_{\Lambda,\lambda}$, which directly implies

$$\|\varepsilon_{\Lambda,\lambda}\|_K = \|\lambda^y K(\cdot, y) - I_{K,\Lambda}(\lambda^y K(\cdot, y))\|_K.$$

□

According to Lemma 5.3, the term $P_\Lambda(\lambda)$ measures the distance, and therefore the approximation quality between $\lambda^y K(\cdot, y)$ and the subspace $S_{K,\Lambda}$ for each $\lambda \in \mathcal{H}_K$. Due to the representation theorem, this is the same as the distance between λ and the span of Λ. With this representation of the power function, we can derive some basic properties, see also [104, Definition 2.1.3 ff.].

Corollary 5.4. *Let $\Lambda = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}_K^*$ be a linearly independent set of functionals. Then P_Λ satisfies the following properties:*

(1) The power function P_Λ is Lipschitz continuous on $\mathcal{H}_K^*$.

(2) For every $\lambda \in \mathcal{H}_K^*$, we have $P_\Lambda(\lambda) = 0$ if and only if $\lambda \in \text{span}_{\mathbb{R}} \{\lambda_1, ..., \lambda_n\}$.

(3) If $\tilde{\Lambda} \subset \mathcal{H}_K$ is another finite linearly independent set with $\Lambda \subset \tilde{\Lambda}$, then we can estimate

$$P_{\tilde{\Lambda}}(\lambda) \leq P_\Lambda(\lambda) \qquad \text{for all } \lambda \in \mathcal{H}_K^*.$$

(4) Given an orthonormal basis $B = \{b_1, ..., b_n\}$ of $S_{K,\Lambda}$, we can compute the power function via the formula

$$P_\Lambda(\lambda)^2 = \langle \lambda, \lambda \rangle_K - \sum_{i=1}^{n} \lambda(b_i)^2 \qquad \text{for all } \lambda \in \mathcal{H}_K^*.$$

Proof. We use the representation from Lemma 5.3:

(1) In the notation of Definition A.1, we have

$$P_\Lambda(\lambda) = \|\lambda^y K(\cdot, y) - I_{K,\Lambda}(\lambda^y K(\cdot, y))\|_K = \inf_{s \in S_{K,\Lambda}} \|\lambda^y K(\cdot, y) - s\|_K = \operatorname{dist}(\lambda^y K(\cdot, y), S_{K,\Lambda})$$

for $\lambda \in \mathcal{H}_K^*$. Since the mapping from (4.4) is an isometric isomorphism, we can further write

$$P_\Lambda(\lambda) = \operatorname{dist}(\lambda^y K(\cdot, y), S_{K,\Lambda}) = \operatorname{dist}(\lambda, \operatorname{span}_{\mathbb{R}} \{\lambda_1, ..., \lambda_n\}). \tag{5.3}$$

Now Lemma A.2 part (1) implies that P_Λ is Lipschitz continuous with Lipschitz constant 1.

(2) Let $\lambda \in \mathcal{H}_K^*$. With the representation (5.3) and Lemma A.2 part (2), we get

$$0 = P_\Lambda(\lambda) = \operatorname{dist}(\lambda, \operatorname{span}_{\mathbb{R}} \{\lambda_1, ..., \lambda_n\}) \quad \Longleftrightarrow \quad \lambda \in \operatorname{span}_{\mathbb{R}} \{\lambda_1, ..., \lambda_n\},$$

as $\operatorname{span}_{\mathbb{R}} \{\lambda_1, ..., \lambda_n\}$ is a closed subspace.

(3) If $\Lambda \subset \tilde{\Lambda}$, then $S_{K,\Lambda} \subset S_{K,\tilde{\Lambda}}$. This immediately leads to

$$P_\Lambda(\lambda) = \inf_{s \in S_{K,\Lambda}} \|\lambda^y K(\cdot, y) - s\|_K \geq \inf_{s \in S_{K,\tilde{\Lambda}}} \|\lambda^y K(\cdot, y) - s\|_K = P_{\tilde{\Lambda}}(\lambda)$$

for every $\lambda \in \mathcal{H}_K^*$.

(4) With part (3) of Theorem A.9 and the generalized reproduction property from Theorem 3.12, we get

$$P_\Lambda(\lambda)^2 = \|\lambda^y K(\cdot, y) - I_{K,\Lambda}(\lambda^y K(\cdot, y))\|_K^2 = \langle \lambda, \lambda \rangle_K - \sum_{i=1}^{n} \langle \lambda^y K(\cdot, y), b_i \rangle_K^2 = \langle \lambda, \lambda \rangle_K - \sum_{i=1}^{n} \lambda(b_i)^2$$

for all $\lambda \in \mathcal{H}_K^*$, provided that $B = \{b_1, ..., b_n\}$ is an orthonormal basis of $S_{K,\Lambda}$.

□

Remark 5.5. Note that the power function P_Λ provides a practical method for identifying linear dependencies between functionals. Due to part (4) of Corollary 5.4, we can evaluate the power function numerically. Then, part (2) of the same corollary states that the updated set $\Lambda \cup \{\lambda\}$ is (nearly) linearly dependent if $P_\Lambda(\lambda)$ is (nearly) equal to zero.

With this knowledge, we can derive the first convergence criterion. So let $(\Lambda_n)_{n \in \mathbb{N}}$ be a nested sequence of linearly independent subsets of $\Gamma \subset \mathcal{H}_K^*$. In order to have convergence on $\mathcal{H}_{K,\Gamma}$, we must at least have

$$P_{\Lambda_n}(\lambda) = \|\lambda^y K(\cdot, y) - I_{K,\Lambda_n}(\lambda^y K(\cdot, y))\|_K \xrightarrow{n \to \infty} 0 \qquad \text{for all } \lambda \in \Gamma.$$

Since the basis functions $\lambda^y K(\cdot, y)$ for $\lambda \in \Gamma$ span a dense subset by construction, the pointwise convergence of the power function towards zero is also sufficient for the convergence of the interpolation method on the whole space $\mathcal{H}_{K,\Lambda}$.

Theorem 5.6. *Let $\Gamma \subset \mathcal{H}_K^*$ and $(\Lambda_n)_{n\in\mathbb{N}}$ be a nested sequence of finite linearly independent subsets of Γ. Then, the generalized interpolation method converges on $\mathcal{H}_{K,\Gamma}$, i.e.*

$$\|f - I_{K,\Lambda_n}(f)\|_K \xrightarrow{n\to\infty} 0 \qquad \text{for all } f \in \mathcal{H}_{K,\Gamma},$$

if and only if the power function converges pointwise to zero on Γ, i.e.

$$P_{\Lambda_n}(\lambda) \xrightarrow{n\to\infty} 0 \qquad \text{for all } \lambda \in \Gamma.$$

Proof. Due to the previous discussion, we only have to prove that the convergence of the power function implies convergence of the interpolation method. We begin with an arbitrary element

$$s = \sum_{i=1}^{N} c_i \cdot \mu_i^y K(\cdot, y) \in \operatorname{span}_{\mathbb{R}} \{\lambda^y K(\cdot, y) \mid \lambda \in \Gamma\} =: S_{K,\Gamma}.$$

Since I_{K,Λ_n} is a linear operator for each $n \in \mathbb{N}$, we get the convergence

$$\|s - I_{K,\Lambda_n}(s)\|_K = \left\| \sum_{i=1}^{N} c_i \cdot \left(\mu_i^y K(\cdot, y) - I_{K,\Lambda_n}(\mu_i^y K(\cdot, y)) \right) \right\|_K \leq \sum_{i=1}^{N} |c_i| \cdot P_{\Lambda_n}(\mu_i) \xrightarrow{n\to\infty} 0.$$

Hence, the interpolation method converges on $S_{K,\Gamma}$. Now let $f \in \mathcal{H}_{K,\Gamma}$ be arbitrary and $\varepsilon > 0$. Since $S_{K,\Gamma}$ is dense in $\mathcal{H}_{K,\Gamma}$, we can find $s \in S_{K,\Lambda}$ with

$$\|f - s\|_K < \frac{\varepsilon}{3}.$$

We can choose $N \in \mathbb{N}$, such that

$$\|s - I_{K,\Lambda_n}(s)\|_K < \frac{\varepsilon}{3} \qquad \text{for all } n \geq N,$$

due to the previously shown convergence on $S_{K,\Gamma}$. From Theorem 4.10, we also know that $\|I_{K,\Lambda_n}\|_K = 1$ for all $n \in \mathbb{N}$. In total, we get

$$\|f - I_{K,\Lambda_n}(f)\|_K \leq \|f - s\|_K + \|s - I_{K,\Lambda_n}(s)\|_K + \|I_{K,\Lambda_n}\|_K \cdot \|f - s\|_K < \varepsilon \qquad \text{for all } n \geq N.$$

□

Remark 5.7. According to Theorem 5.6, the interpolation method converges normwise on $\mathcal{H}_{K,\Gamma}$ if the power function converges pointwise to zero on $\Gamma \subset \mathcal{H}_K^*$. By definition of the power function, this also implies the convergence

$$|\lambda(f) - \lambda(I_{K,\Lambda_n}(f))| \leq P_{\Lambda_n}(\lambda) \cdot \|f\|_K \xrightarrow{n\to\infty} 0 \qquad \text{for all } \lambda \in \Gamma, f \in \mathcal{H}_K.$$

If the superset Γ is compact and the power function converges pointwise to zero, then *Dini's theorem* implies that the power function also converges uniformly to zero on Γ. Note that Corollary 5.4 shows that the requirements of Dini's Theorem are fulfilled.

5.2. Fill distance

Another indicator for the approximation quality of the interpolation method is the *fill distance* (cf. [143, Definition 1.4]), which is based on the distance function defined in Definition A.1.

Definition 5.8. Let (Z, d_Z) be a metric space and $A \subset Z$. We define the `fill distance` $h_{A,Z}$ of A in Z as the supremum of all distances to the subset A, i.e.

$$h_{A,Z} = \sup_{z\in Z} \operatorname{dist}(z, A).$$

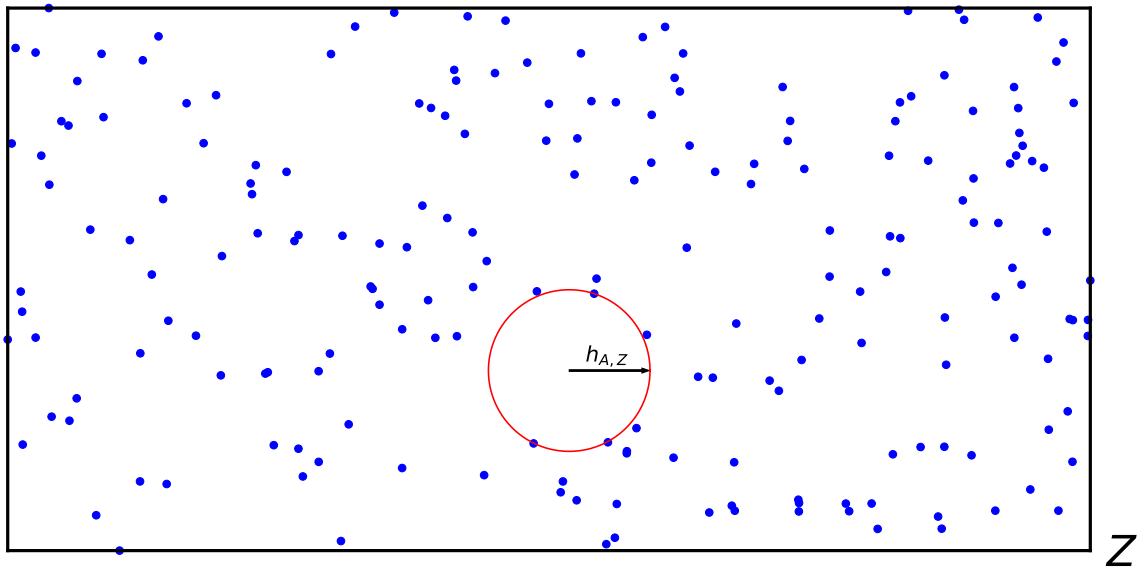

Figure 5.1.: Fill distance of a scattered data point set A (blue) inside a rectangular domain $Z \subset \mathbb{R}^2$

In other words, the fill distance is the radius of the largest ball in Z that does not include an element of A, see Figure 5.1. For our generalized interpolation problem, we are interested in the fill distances $h_{\Lambda_n,\Gamma}$, where $\Gamma \subset \mathcal{H}_K^*$ is the given superset of functionals and $(\Lambda_n)_{n\in\mathbb{N}}$ is again a nested sequence of linearly independent subsets of Γ.

Similar to the properties of the power function, the inequality $h_{\Lambda_m,X} \leq h_{\Lambda_n,X}$ holds for $m \geq n$. Note that the distance function $\operatorname{dist}(\cdot,\Lambda_n)$, and therefore the fill distance $h_{\Lambda_n,\Gamma}$ gives an upper bound for the power function P_{Λ_n} via

$$P_{\Lambda_n}(\lambda) = \operatorname{dist}(\lambda, \operatorname{span}_{\mathbb{R}}(\Lambda_n)) \leq \operatorname{dist}(\lambda,\Lambda_n) \leq h_{\Lambda_n,\Gamma} \qquad \text{for all } n \in \mathbb{N}, \tag{5.4}$$

as we have $\Lambda_n \subset \operatorname{span}_{\mathbb{R}}(\Lambda_n)$. Hence, if the sequence of fill distances $(h_{\Lambda_n,\Gamma})_{n\in\mathbb{N}}$ converges to zero, the interpolation method converges on $\mathcal{H}_{K,\Gamma}$.

Theorem 5.9. *Let $\Gamma \subset \mathcal{H}_K^*$ and $(\Lambda_n)_{n\in\mathbb{N}}$ be a nested sequence of finite linearly independent subsets of Γ that satisfies*

$$h_{\Lambda_n,\Gamma} \xrightarrow{n\to\infty} 0.$$

Then we have

$$\|f - I_{K,\Lambda_n}(f)\|_K \xrightarrow{n\to\infty} 0 \qquad \text{for all } f \in \mathcal{H}_{K,\Gamma}.$$

Proof. Due to the inequality (5.4), we immediately get

$$P_{\Lambda_n}(\lambda) \leq h_{\Lambda_n,\Gamma} \xrightarrow{n\to\infty} 0 \qquad \text{for all } \lambda \in \Gamma.$$

With Theorem 5.6, we conclude

$$\|f - I_{K,\Lambda_n}(f)\|_K \xrightarrow{n\to\infty} 0 \qquad \text{for all } f \in \mathcal{H}_{K,\Gamma}.$$

□

With the inequality (5.4) and Remark 5.7, we also obtain an estimate for the pointwise error on the superset Γ.

Corollary 5.10. *For every $n \in \mathbb{N}$, we have the pointwise error estimate*

$$|\lambda(f) - \lambda(I_{K,\Lambda_n})(f)| \leq h_{\Lambda_n,\Gamma} \cdot \|f\|_K \qquad \text{for all } \lambda \in \Gamma, f \in \mathcal{H}_K.$$

5.3. Parametrization

In many cases, the set $\Gamma \subset \mathcal{H}_K^*$ can be parametrized by another well-known set. For example, the set of all point evaluation functionals is parametrized by the (standard) mapping

$$\varrho : \mathbb{R}^d \to \{\delta_x \mid x \in \mathbb{R}^d\},\ x \mapsto \delta_x.$$

This parametrization allows us to shift our convergence analysis to the parameter space $\mathbb{R}^d$ under a certain regularity of the mapping ϱ. But, before we get into detail, we want to clarify what we exactly mean by the term *parametrization*.

Definition 5.11. Let $\Gamma \subset \mathcal{H}_K^*$. Moreover, let Ω be a set and $\varrho : \Omega \to \Gamma$. We call the tuple (Ω, ϱ) a `parametrization` of Γ, if the mapping ϱ is surjective. In short, we also say that Γ is parametrized by Ω or that Γ is parametrized by the mapping ϱ and Ω is called the `parameter space`.

Example 5.12. Consider a kernel function $K \in C^{2k}(\mathbb{R}^d \times \mathbb{R}^d)$ with $k \in \mathbb{N}$ and the set

$$\Gamma = \{\lambda_{x,\alpha} = \delta_x \circ D^\alpha \mid x \in \mathbb{R}^d, \alpha \in \mathbb{N}_0^d \text{ with } |\alpha| \leq k\} \subset \mathcal{H}_K^*,$$

see Proposition 3.14. Then, the mapping

$$\varrho : \mathbb{R}^d \times \{\alpha \in \mathbb{N}_0^d \mid |\alpha| \leq k\} \to \Gamma,\ (x, \alpha) \mapsto \lambda_{x,\alpha}$$

yields a parametrization of Γ.

The main advantage of parametrizations is that, in most cases, the analysis of the parameter space is much simpler and computations are less costly. Hence, it is desirable to transfer the previous convergence results to the parameter space Ω.

This idea is mainly relevant for the fill distance approach, since a parametrization (Ω, ϱ) of $\Gamma \subset \mathcal{H}_K^*$ only leads to a reformulated power function on Ω, i.e.

$$P_X(x) := P_{\varrho(X)}(\varrho(x)) \qquad \text{for all } x \in \Omega \tag{5.5}$$

and for finite $X \subset \Omega$, provided that $\varrho(X)$ is linearly independent. Usually, the reformulation (5.5) does not lead to a simplified version of the power function.

In contrast, the concept of the fill distance can be used to derive a convergence criterion in Ω similar to Theorem 5.9 under suitable assumptions on Ω and ϱ. As the metric space (Ω, d_Ω) is usually well-known, it is easier to find sequences of subsets $X_n \subset \Omega$, $n \in \mathbb{N}$, such that the sequence of fill distances in Ω converges to zero.

Theorem 5.13. *Let $\Gamma \subset \mathcal{H}_K^*$ and (Ω, ϱ) be a parametrization of Γ, such that Ω is a metric space and ϱ is uniformly continuous. Moreover, let $(X_n)_{n\in\mathbb{N}}$ be a nested sequence of finite subsets in Ω, such that $\varrho(X_n) \subset \mathcal{H}_K^*$ is linearly independent for all $n \in \mathbb{N}$ and*

$$h_{X_n,\Omega} \xrightarrow{n\to\infty} 0.$$

Then, the induced generalized interpolation method converges on $\mathcal{H}_{K,\Gamma}$, i.e.

$$\|f - I_{K,\varrho(X_n)}(f)\|_K \xrightarrow{n\to\infty} 0 \qquad \textit{for all } f \in \mathcal{H}_{K,\Gamma}.$$

Proof. For $n \in \mathbb{N}$, we set $\Lambda_n = \varrho(X_n)$. With Theorem 5.9, it is sufficient to show that

$$h_{\Lambda_n,\Gamma} \xrightarrow{n\to\infty} 0.$$

To this end, let $\varepsilon > 0$. Since ϱ is uniformly continuous on Ω, there is $\delta > 0$ such that

$$\|\varrho(x) - \varrho(y)\|_K < \varepsilon \quad \text{for} \quad d_\Omega(x, y) < \delta.$$

Now choose $N \in \mathbb{N}$, such that $h_{X_N,\Omega} < \delta$ and write $X_N = \{x_1, ..., x_m\}$, $m = |X_N|$. For any $\lambda \in \Gamma$, there is $x_\lambda \in \Omega$ with $\lambda = \varrho(x_\lambda)$, since (Ω, ϱ) is a parametrization of Γ. Moreover, there is an index $i_\lambda \in \{1, ..., m\}$ that satisfies

$$d_\Omega(x_\lambda, x_{i_\lambda}) = \inf_{j\in\{1,...,M\}} d_\Omega(x_\lambda, x_j) = \operatorname{dist}(x_\lambda, X_N).$$

Then we have

$$\operatorname{dist}(\lambda, \varrho(X_N)) \leq \|\lambda - \varrho(x_{i_\lambda})\|_K = \|\varrho(x_\lambda) - \varrho(x_{i_\lambda})\|_K < \varepsilon \qquad \text{for all } \lambda \in \Gamma,$$

as the inequality

$$d_\Omega(x_\lambda, x_{i_\lambda}) = \operatorname{dist}(x_\lambda, X_N) \leq h_{X_N,\Omega} < \delta$$

holds. Therefore, we get

$$h_{\Lambda_n,\Gamma} \leq h_{\Lambda_N,\Gamma} \leq \varepsilon \qquad \text{for all } n \geq N.$$

□

If the mapping ϱ is *Hölder continuous*, we can derive an estimate for the pointwise interpolation error in dependence of the fill distance in the parameter space, which is a generalization of [70, Exercise 8.66].

Corollary 5.14. *Under the assumptions of Theorem 5.13, let ϱ be Hölder continuous with exponent $a \in (0, 1]$, i.e. there is $C > 0$ such that*

$$\|\varrho(x) - \varrho(y)\|_K \leq C \cdot d_\Omega(x, y)^a \qquad \textit{for all } x, y \in \Omega.$$

Then, we have the pointwise error estimate

$$|\lambda(f) - \lambda(I_{K,\Lambda_n})(f)| \leq C \cdot h_{X_n,\Omega}^a \cdot \|f\|_K \qquad \textit{for all } \lambda \in \Gamma, f \in \mathcal{H}_K.$$

Proof. Again, for any $\lambda \in \Gamma$, there exists $x_\lambda \in \Omega$ such that $\varrho(x_\lambda) = \lambda$. Using the Hölder continuity of ϱ, we can estimate

$$\operatorname{dist}(\lambda, \varrho(X_n)) = \inf_{x \in X_n} \|\varrho(x_\lambda) - \varrho(x)\|_K \leq C \cdot \inf_{x \in X_n} d_\Omega(x_\lambda, x)^a \leq C \cdot h_{X_n,\Omega}^a \qquad \text{for all } \lambda \in \Gamma,$$

and therefore

$$h_{\varrho(X_n),\Gamma} \leq C \cdot h_{X_n,\Omega}^a.$$

By Corollary 5.10, we get

$$|\lambda(f) - \lambda(I_{K,\Lambda_n})(f)| \leq h_{\varrho(X_n),\Gamma} \cdot \|f\|_K \leq C \cdot h_{X_n,\Omega}^a \cdot \|f\|_K \qquad \text{for all } \lambda \in \Gamma, f \in \mathcal{H}_K.$$

□

Thus, we can shift our convergence analysis to the parameter space, if the requirements of Theorem 5.13 and Corollary 5.14 are fulfilled.

Example 5.15. If $\Omega \subset \mathbb{R}^d$ is compact and $K \in \mathscr{C}(\mathbb{R}^d \times \mathbb{R}^d)$ is positive definite, then K is automatically uniformly continuous on the compact set $\Omega \times \Omega$. Due to the equation

$$\|\delta_x - \delta_y\|_K^2 = \|K(\cdot, x) - K(\cdot, y)\|_K^2 = K(x, x) - 2 \cdot K(x, y) + K(y, y) \qquad \text{for all } x, y \in \Omega,$$

the mapping

$$\varrho : \Omega \to \{\delta_x \mid x \in \Omega\},\ x \mapsto \delta_x$$

is uniformly continuous, so that the resulting interpolation method converges according to Theorem 5.13 for suitable sequences of subsets of Ω. Similar convergence results for the standard interpolation were proven in [70, Section 8.4.2].

Remark 5.16. If $\Gamma \subset \mathcal{H}_K^*$ is linearly independent, we can use the parametrization (Ω, ϱ) of Γ to define a positive definite kernel function on Ω via

$$\tilde{K}(x, y) := \langle \varrho(x), \varrho(y) \rangle_K \qquad \text{for all } x, y \in \Omega,$$

see also [143, Theorem 16.8]. The resulting kernel function is positive definite on Ω, as we have

$$c^T \cdot A_{\tilde{K},X} \cdot c = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i \cdot c_j \cdot \tilde{K}(x_i, x_j) = \left\| \sum_{i=1}^{n} c_i \cdot \varrho(x_i) \right\|_K^2 > 0$$

for any finite set $X = \{x_1, ..., x_n\} \subset \Omega$ of pairwise distinct points and any vector $c = (c_1, ..., c_n) \in \mathbb{R}^n \setminus \{0\}$, as we assumed that Γ is linearly independent. In [53], [54] and [143, Section 16.3], the new kernel $\tilde{K}$ is used to interpret the generalized interpolation problem as a standard interpolation problem.

Throughout this chapter, we have shown that the decay of the power function or the fill distance leads to convergence of the interpolation method. But we can also show that the convergence can be arbitrarily slow on $\mathcal{H}_{K,\Gamma}$. We provide a generalized version of [70, Exercise 8.64] to close out this chapter.

Theorem 5.17. *Let $\Gamma \subset \mathcal{H}_K^*$ and $(\Lambda_n)_{n\in\mathbb{N}}$ be a nested sequence of finite linearly independent subsets from Γ with strictly increasing cardinality, i.e. $|\Lambda_n| < |\Lambda_{n+1}|$ for all $n \in \mathbb{N}$. For any monotonically decreasing null sequence $(\eta_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{\geq 0}$, there exists a function $f \in \mathcal{H}_{K,\Gamma}$ such that*

$$\|f - I_{K,\Lambda_n}(f)\|_K \geq \eta_n \qquad \text{for all } n \in \mathbb{N}.$$

Proof. Due to our assumptions, we can find an orthonormal sequence $(b_n)_{n\in\mathbb{N}}$ in $\mathcal{H}_{K,\Gamma}$, i.e.

$$\langle b_i, b_j \rangle_K = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j, \end{cases}$$

such that

$$b_{i+1} \in S_{K,\Lambda_{i+1}} \cap S_{K,\Lambda_i}^{\perp} \qquad \text{for all } n \in \mathbb{N}.$$

Moreover, we define the coefficients c_i via

$$c_i^2 := \begin{cases} 0 & \text{for } i = 1 \\ \eta_{i-1}^2 - \eta_i^2 & \text{for } i \geq 2 \end{cases}$$

and set

$$s_n = \sum_{i=1}^{n} c_i \cdot b_i \qquad \text{for all } n \in \mathbb{N}.$$

The resulting sequence $(s_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{H}_{K,\Gamma}$, as we have

$$\|s_m - s_n\|_K^2 = \sum_{i=n+1}^{m} c_i^2 = \eta_n^2 - \eta_m^2 \qquad \text{for } m \geq n.$$

Since $\mathcal{H}_{K,\Gamma}$ is complete, there exists

$$f = \lim_{n\to\infty} s_n = \sum_{i=1}^{\infty} c_i \cdot b_i \in \mathcal{H}_{K,\Gamma}.$$

By construction of f and the basis elements b_i, $i \in \mathbb{N}$, we have $I_{K,\Lambda_n}(f) = s_n$ and therefore

$$\|f - I_{K,\Lambda_n}(f)\|_K^2 = \lim_{m\to\infty} \|s_m - s_n\|_K^2 = \lim_{m\to\infty} \eta_n^2 - \eta_m^2 = \eta_n^2$$

for every $n \in \mathbb{N}$. □

6. Alternative Bases

In Chapter 4, we discussed the construction of a generalized interpolant in terms of the Riesz representers of Λ. More precisely, given a linearly independent set $\Lambda = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}_K^*$ and a function $f \in \mathcal{H}_K$, we can compute the interpolant $I_{K,\Lambda}(f)$ by solving the linear system

$$A_{K,\Lambda} \cdot c = f_\Lambda,$$

where $f_\Lambda = (\lambda_1(f), ..., \lambda_n(f))^T \in \mathbb{R}^n$. The solution $c \in \mathbb{R}^n$ of this linear system is then used to write the interpolant as

$$I_{K,\Lambda}(f) = \sum_{i=1}^{n} c_i \cdot \lambda_i^y K(\cdot, y).$$

In the case of standard interpolation, it is well-known that the standard basis

$$B = \left\{ \lambda_1^y K(\cdot, y), ..., \lambda_n^y K(\cdot, y) \right\} \tag{6.1}$$

can lead to severe numerical problems like high condition numbers of $A_{K,\Lambda}$, see e.g. [25]. Moreover, this basis is quite inefficient in a situation where we successively increase the number of interpolation points and need to update the interpolant. Hence, we are interested in alternative data-dependent bases, i.e. bases whose construction depends on the considered set of functionals Λ, that improve the stability of the interpolation method and can be updated easily when adding new functionals.

First, we introduce a generalized *Lagrangian basis* in Section 6.1, which leads to a simple representation of the interpolant and is therefore useful for theoretical concerns. However, this basis does not come with an update formula, so we only use it to derive stability estimates in Section 7.1. Second, we discuss the construction of orthonormal bases via matrix decompositions in Section 6.2, which generalizes [105, Theorem 6.1 ff.]. As in the standard interpolation case, this approach results in the *(generalized) Newton basis*, initially introduced in [99]. We translate relevant properties of the Newton basis (cf. [98, Section 4 & 5]) to the generalized case. In Section 6.3, we provide update formulas for the Newton basis that are derived from the standard case (cf. [105, Section 10]). The utility of the Newton basis is further demonstrated in Chapter 7, where it is supplemented with suitable point selection strategies.

Remark 6.1. Besides the construction of data-dependent (orthonormal) bases, there are numerous other stabilization methods that use a series expansion of the considered kernel function. A list of such methods can be found in [81, Section 1]. One example is given in [51], where *Mercer's theorem* is used to rewrite the kernel in terms of the eigenvalues and orthonormal eigenfunctions of the associated integral operator. An application of this approach to fractional derivative interpolation can be found in [46].

6.1. Lagrangian basis

The first basis we discuss here is the *Lagrangian basis*, which is a popular concept from polynomial interpolation (cf. [70, Section 2.3]). It is uniquely characterized by the equations

$$\lambda_i(l_j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j, \end{cases} \tag{6.2}$$

where $\mathcal{L} = \{l_1, ..., l_n\} \subset S_{K,\Lambda}$ denotes the Lagrangian basis with respect to $\Lambda = \{\lambda_1, ..., \lambda_n\}$. Note that solving for the j-th Lagrangian basis element l_j in the standard basis representation is equivalent to solving the linear system

$$A_{K,\Lambda} \cdot c^{(j)} = e^{(j)}, \tag{6.3}$$

with $c^{(j)} \in \mathbb{R}^n$ and $e^{(j)}$ being the j-th standard basis vector in $\mathbb{R}^n$. As $A_{K,\Lambda}$ is regular for the linearly independent set $\Lambda \subset \mathcal{H}_K^*$ (cf. Chapter 4), there is indeed a unique coefficient vector $c^{(j)} \in \mathbb{R}^d$ such that

$$l_j = \sum_{i=1}^{n} c_i^{(j)} \cdot \lambda_i^y K(\cdot, y) \in S_{K,\Lambda}$$

satisfies the equations (6.2) for each $j \in \{1, ..., n\}$. By construction of this basis, the interpolant to a function $f \in \mathcal{H}_K$ can simply be written as

$$I_{K,\Lambda}(f) = \sum_{j=1}^{n} \lambda_j(f) \cdot l_j, \tag{6.4}$$

since we have

$$\lambda_i \left(\sum_{j=1}^{n} \lambda_j(f) \cdot l_j \right) = \sum_{j=1}^{n} \lambda_j(f) \cdot \lambda_i(l_j) = \lambda_i(f) \qquad \text{for all } i \in \{1, ..., n\}.$$

In other words, the interpolation matrix with respect to the Lagrangian basis is the $n \times n$ identity matrix, which means that no computations are necessary to identify the coefficients of the interpolant in the Lagrangian basis representation. We summarize the previous discussion in the following theorem (cf. [1, Section 3.2], [70, Proposition 8.4]).

Theorem 6.2. *Let $\Lambda = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}_K^*$ be a linearly independent set of functionals. Then, there is a Lagrangian basis $\mathcal{L} = \{l_1, ..., l_n\} \subset S_{K,\Lambda}$ that satisfies the conditions (6.2). For each $f \in \mathcal{H}_K$, the unique interpolant to f on Λ can be written as*

$$I_{K,\Lambda}(f) = \sum_{j=1}^{n} \lambda_j(f) \cdot l_j.$$

Similar to the case of (univariate) polynomial interpolation, the Lagrangian basis is not very useful in practical cases, as its computation requires the solution of the n linear systems (6.3). But, due to the simplicity of the representation (6.4), it is useful in the theoretical analysis of the interpolation method. For example, the Lagrangian basis can be used to derive another representation of the power function from Section 5.1. To this end, we need to evaluate the inner products between the Lagrangian basis functions (cf. [70, Proposition 8.14 ff.]).

Lemma 6.3. *Let $\Lambda \subset \mathcal{H}_K^*$ be finite and linearly independent. We define the inner product $\langle \cdot, \cdot \rangle_{A_{K,\Lambda}}$ via*

$$\langle c, d \rangle_{A_{K,\Lambda}} := c^T \cdot A_{K,\Lambda} \cdot d \qquad \text{for } c, d \in \mathbb{R}^n.$$

Then, the mapping

$$G_\Lambda : \mathbb{R}^n \to S_{K,\Lambda},\ (c_1, ..., c_n)^T \mapsto \sum_{i=1}^{n} c_i \cdot \lambda_i^y K(\cdot, y)$$

is an isometric isomorphism between the spaces $\left(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{A_{K,\Lambda}}\right)$ and $(S_{K,\Lambda}, \langle \cdot, \cdot \rangle_K)$. Moreover, we have

$$\langle l_i, l_j \rangle_K = {e^{(i)}}^T \cdot A_{K,\Lambda}^{-1} \cdot e^{(j)} \qquad \text{for all } i, j \in \{1, ..., n\},$$

which means that $A_{K,\Lambda}^{-1}$ is the Gram matrix of $\mathcal{L}$ with respect to the inner product $\langle \cdot, \cdot \rangle_K$.

Proof. By definition of $S_{K,\Lambda}$, G_Λ is an isomorphism. For given $c = (c_1, ..., c_n)^T, d = (d_1, ..., d_n)^T \in \mathbb{R}^n$, we compute

$$\langle G_\Lambda(c), G_\Lambda(d) \rangle_K = \Big\langle \sum_{i=1}^{n} c_i \cdot \lambda_i^y K(\cdot, y), \sum_{j=1}^{n} d_j \cdot \lambda_j^y K(\cdot, y) \Big\rangle_K = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i \cdot d_j \cdot \langle \lambda_i, \lambda_j \rangle_K = c^T \cdot A_{K,\Lambda} \cdot d.$$

Hence, G_Λ is also isometric. From equation (6.3), we know that the coefficient vectors of the Lagrangian basis functions are given by $c^{(i)} = A_{K,\Lambda}^{-1} \cdot e^{(i)}$ for $i = 1, ..., n$. Using the first part, this leads to

$$\langle l_i, l_j \rangle_K = {e^{(i)}}^T \cdot A_{K,\Lambda}^{-1} \cdot A_{K,\Lambda} \cdot A_{K,\Lambda}^{-1} \cdot e^{(j)} = {e^{(i)}}^T \cdot A_{K,\Lambda}^{-1} \cdot e^{(j)} \qquad \text{for all } i, j \in \{1, ..., n\}.$$

□

Corollary 6.4. *The power function with respect to Λ can be expressed as*

$$P_\Lambda(\lambda)^2 = \langle \lambda, \lambda \rangle_K - R^T \cdot A_{K,\Lambda}^{-1} \cdot R = \langle \lambda, \lambda \rangle_K - l^T \cdot A_{K,\Lambda} \cdot l \qquad \textit{for all } \lambda \in \mathcal{H}_K^*,$$

where $R = (\langle \lambda, \lambda_1 \rangle_K, ..., \langle \lambda, \lambda_n \rangle_K)^T$ and $l = (\lambda(l_1), ..., \lambda(l_n))^T$.

Proof. According to Lemma 5.3, we can write

$$P_\Lambda(\lambda)^2 = \| \lambda^y K(\cdot, y) - I_{K,\Lambda}(\lambda^y K(\cdot, y)) \|_K^2 = \langle \lambda^y K(\cdot, y) - I_{K,\Lambda}(\lambda^y K(\cdot, y)), \lambda^y K(\cdot, y) - I_{K,\Lambda}(\lambda^y K(\cdot, y)) \rangle_K$$

for a given $\lambda \in \mathcal{H}_K^*$. Inserting the representation (6.4) for the interpolant $I_{K,\Lambda}(\lambda^y K(\cdot, y))$ leads to

$$\begin{aligned} P_\Lambda(\lambda)^2 &= \langle \lambda, \lambda \rangle_K - 2 \cdot \sum_{i=1}^n \langle \lambda, \lambda_i \rangle_K \cdot \lambda(l_j) + \sum_{i=1}^n \sum_{j=1}^n \langle \lambda, \lambda_i \rangle_K \cdot \langle \lambda, \lambda_j \rangle_K \cdot \langle l_i, l_j \rangle_K \\ &= \langle \lambda, \lambda \rangle_K - 2 \cdot R^T \cdot l + R^T \cdot A_{K,\Lambda}^{-1} \cdot R, \end{aligned}$$

where we used the second part of Lemma 6.3. Note that we have the relation

$$\lambda(l_j) = \lambda \left(\sum_{i=1}^n c_i^{(j)} \cdot \lambda_i^y K(\cdot, y) \right) = {e^{(i)}}^T \cdot A_{K,\Lambda}^{-1} \cdot R \qquad \text{for all } j \in \{1, ..., n\},$$

or in other words, $l = A_{K,\Lambda}^{-1} \cdot R$. In conlcusion, we get

$$P_\Lambda(\lambda)^2 = \langle \lambda, \lambda \rangle_K - 2 \cdot R^T \cdot l + R^T \cdot A_{K,\Lambda}^{-1} \cdot R = \begin{cases} \langle \lambda, \lambda \rangle_K - R^T \cdot A_{K,\Lambda}^{-1} \cdot R \\ \langle \lambda, \lambda \rangle_K - l^T \cdot A_{K,\Lambda} \cdot l. \end{cases}$$

□

6.2. Orthonormal bases

We observed in Section 4.2 that the interpolant to a function $f \in \mathcal{H}_K$ coincides with its orthogonal projection onto the space $S_{K,\Lambda}$ so that we can make use of the orthogonal projection formula (4.11). Note that the coefficients of this representation satisfy the (stability) estimate

$$|\langle f, b_i \rangle_K| \leq \|f\|_K \qquad \text{for } i \in \{1, ..., n\}.$$

As we stated in the introduction of this chapter, the construction of orthonormal bases of $S_{K,\Lambda}$ has already been discussed for the standard interpolation setting (cf. [98], [99], [104], [105]). In particular, an efficient *Newton basis* was introduced in [99], inspired by the properties of the Newton basis known from univariate polynomial interpolation (see, e.g., [70, Section 2.4]).

Analogously to the results of Chapter 4, the Newton basis can be transferred to the setting of generalized interpolation. But, before we get into detail, we give a characterization that identifies the construction of orthonormal bases of $S_{K,\Lambda}$ with symmetric matrix decompositions of the interpolation matrix $A_{K,\Lambda}$ (cf. [105, Theorem 6.1]).

Theorem 6.5. *Under the assumptions of Lemma 6.3, let $B = \{b_1, ..., b_n\}$ be a set of functions of $S_{K,\Lambda}$. We set $c^{(i)} = G_\Lambda^{-1}(b_i)$ for $i = 1, ..., n$ and define the coefficient matrix C_B as*

$$C_B := \begin{pmatrix} c^{(1)} & \dots & c^{(n)} \end{pmatrix} \in \mathbb{R}^{n \times n}$$

containing the coefficient vectors as columns. Then B is an orthonormal basis of $S_{K,\Lambda}$ if and only if

$$A_{K,\Lambda} = {C_B^{-1}}^T \cdot C_B^{-1}.$$

Proof. With the result from Lemma 6.3, we get that B is an orthonormal basis of $S_{K,\Lambda}$ if and only if C_B is invertible and

$$C_B^T \cdot A_{K,\Lambda} \cdot C_B = I_n$$

holds, where I_n is the $n \times n$ identity matrix. Due to the invertibility of C_B, this is equivalent to

$$A_{K,\Lambda} = {C_B^{-1}}^T \cdot C_B^{-1}.$$

□

The previous theorem now enables us to construct orthonormal bases with the help of suitable matrix decompositions. If $C \in \mathbb{R}^{n \times n}$ is a regular matrix satisfying $A_{K,\Lambda} = C^T \cdot C$, we can set $b_i = G_\Lambda(c^{(i)})$ for $i = 1, ..., n$, where $c^{(i)}$ denotes the i-th column of C^{-1}. Then, Theorem 6.5 guarantees that the set $B = \{b_1, ..., b_n\}$ is an orthonormal system in $S_{K,\Lambda}$. We discuss two different symmetric matrix decompositions here:

- **Eigenvalue decomposition:** Since $A_{K,\Lambda}$ is symmetric positive definite, there is an eigenvalue decomposition

  $$A_{K,\Lambda} = Q^T \cdot D \cdot Q,$$

  where $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix and

  $$D = \begin{pmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n \end{pmatrix}$$

  is a diagonal matrix containing the positive eigenvalues of $A_{K,\Lambda}$. If we set

  $$D^{1/2} := \begin{pmatrix} \sqrt{\sigma_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{\sigma_n} \end{pmatrix}$$

  and $C := D^{1/2} \cdot Q$, we obtain a symmetric decomposition of $A_{K,\Lambda}$. Hence, the inverse matrix $C^{-1} = Q^T \cdot {D^{1/2}}^{-1}$ contains the coefficients of an orthonormal basis of $S_{K,\Lambda}$ (cf. [70, Proposition 8.35]). However, it is not feasible to use the eigenvalue decomposition, since the computation of the eigenvalues and orthogonal eigenvectors is very costly, especially for large matrices. In general, the eigenvalue decomposition of the interpolation matrices cannot be updated efficiently if the number of interpolation points increases throughout the interpolation process.

- **Cholesky decomposition:** Again, due to the positive definiteness of $A_{K,\Lambda}$, there is a *Cholesky decomposition*

  $$A_{K,\Lambda} = L \cdot L^T,$$

  where

  $$L = \begin{pmatrix} l_{1,1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ * & \dots & l_{n,n} \end{pmatrix}$$

  is a lower triangular matrix containing positive entries on the diagonal. Hence, the matrix ${L^T}^{-1}$ also contains the coefficients of an orthonormal basis. In contrast to the eigenvalue decomposition, there is an efficient update formula for the Cholesky decomposition and its inverse factor. We explain the details of this update strategy in Section 6.3.

Due to the aforementioned reasons, we restrict ourselves to the Cholesky decomposition of the interpolation matrix for the computation of an orthonormal system. This particular basis coincides with a normalized version of the Newton basis from [99].

Theorem 6.6. *Let L be the Cholesky factor of the interpolation matrix $A_{K,\Lambda}$, i.e. $A_{K,\Lambda} = L \cdot L^T$. We define the functions*

$$\mathfrak{n}_i = G_\Lambda(L^{T^{-1}} \cdot e^{(i)}) \qquad \text{for } i = 1, ..., n,$$

where $e^{(i)} \in \mathbb{R}^n$ denotes the i-th standard basis vector. Then $\mathcal{N} = \{\mathfrak{n}_1, ..., \mathfrak{n}_n\}$ is an orthonormal basis of $S_{K,\Lambda}$ that satisfies the following properties:

(1) We have $\operatorname{span}_{\mathbb{R}} \{\mathfrak{n}_1, ..., \mathfrak{n}_i\} = S_{K,\Lambda_i}$ for each $i \in \{1, ..., n\}$, where $\Lambda_i = \{\lambda_1, ..., \lambda_i\}$.

(2) The basis functions satisfy the Newton basis condition

$$\lambda_i(\mathfrak{n}_j) = 0 \qquad \text{for } i < j.$$

Proof. According to Theorem 6.5 and our previous discussion, $\mathcal{N}$ is an orthonormal basis of $S_{K,\Lambda}$. Due to the upper triangular structure of $L^{T^{-1}}$, we have $\mathfrak{n}_j \in S_{K,\Lambda_i}$ for $j \leq i$. Part (1) follows from

$$\dim\left(\operatorname{span}_{\mathbb{R}} \{\mathfrak{n}_1, ..., \mathfrak{n}_i\}\right) = i = \dim\left(S_{K,\Lambda_i}\right).$$

If $j > i$, we have $\mathfrak{n}_j \perp \operatorname{span}_{\mathbb{R}} \{\mathfrak{n}_1, ..., \mathfrak{n}_i\} = S_{K,\Lambda_i}$ and therefore

$$0 = \langle \mathfrak{n}_j, \lambda_i^y K(\cdot, y) \rangle_K = \lambda_i(\mathfrak{n}_j).$$

□

From now on, we call the basis $\mathcal{N}$ in Theorem 6.6 the Newton basis of $S_{K,\Lambda}$. It should be mentioned here that, due to the structure of the Cholesky factor L, an application of the Gram-Schmidt orthogonalization process to the standard basis of $S_{K,\Lambda}$ results in the constructed Newton basis as well. Nevertheless, the characterization of orthonormal bases via matrix decompositions is still an important result, as it leaves room for further bases.

Remark 6.7. Since the coefficients of the Newton basis elements $\mathfrak{n}_1, ..., \mathfrak{n}_n$ with respect to the standard basis of $S_{K,\Lambda}$ are given by the columns of the transposed inverse Cholesky factor L^{-1^T} of $A_{K,\Lambda}$, the generalized Vandermonde matrix of the Newton basis and Λ is given by the Cholesky factor L of $A_{K,\Lambda}$, i.e.

$$\begin{pmatrix} \lambda_1(\mathfrak{n}_1) & \dots & \lambda_1(\mathfrak{n}_n) \\ \vdots & \ddots & \vdots \\ \lambda_n(\mathfrak{n}_1) & \dots & \lambda_n(\mathfrak{n}_n) \end{pmatrix} = A_{K,\Lambda} \cdot L^{-1^T} = L.$$

It is possible to link the Newton basis to the Lagrangian basis of the previous section (cf. [70, Theorem 8.41]).

Corollary 6.8. *Let $\mathcal{L}^{(i)} = \{l_1^{(i)}, ..., l_i^{(i)}\}$ denote the Lagrangian basis of S_{K,Λ_i} for each $i \in \{1, ..., n\}$, where $\Lambda_i = \{\lambda_1, ..., \lambda_i\}$. Then we have*

$$l_i^{(i)} = \lambda_i(\mathfrak{n}_i)^{-1} \cdot \mathfrak{n}_i \qquad \text{for } i = 1, ..., n.$$

Proof. First, we need to verify that $\lambda_i(\mathfrak{n}_i) \neq 0$. If this is not the case, we have

$$\lambda_j(\mathfrak{n}_i) = 0 \qquad \text{for each } j \in \{1, ..., i\}.$$

Since $\mathfrak{n}_i \in S_{K,\Lambda}$, this implies $\mathfrak{n}_i = 0$, as the interpolation with respect to Λ_i is uniquely solvable. But this cannot hold for an element of a vector space basis. With part (2) of Theorem 6.6, we get

$$\lambda_j(l_i^{(i)}) = \lambda_j(\lambda_i(\mathfrak{n}_i)^{-1} \cdot \mathfrak{n}_i) \qquad \text{for } j \leq i.$$

Again, due to unique interpolation on S_{K,Λ_i}, we must have $l_i^{(i)} = \lambda_i(\mathfrak{n}_i)^{-1} \cdot \mathfrak{n}_i$. □

6.3. Update strategies

One of the biggest advantages of the Newton basis is the possibility to update efficiently. More precisely, if we add a new functional $\lambda_{n+1} \notin \operatorname{span}_{\mathbb{R}}(\Lambda)$, we update the current set of Newton basis functions $\mathcal{N} = \{\mathfrak{n}_1, ..., \mathfrak{n}_n\}$ to the Newton basis $\tilde{\mathcal{N}} = \{\tilde{\mathfrak{n}}_1, ..., \tilde{\mathfrak{n}}_{n+1}\}$ of $S_{K,\tilde{\Lambda}}$, where $\tilde{\Lambda} = \Lambda \cup \{\lambda_{n+1}\}$. The update formula for the Newton basis $\tilde{N}$ can be derived by updating the inverse Cholesky factor of the interpolation matrix $A_{K,\Lambda}$ (cf. [70, Theorem 8.45]).

Lemma 6.9. *Let L be the Cholesky factor of $A_{K,\Lambda}$. Moreover, let $\lambda_{n+1} \notin \operatorname{span}_{\mathbb{R}}(\Lambda)$ and $\tilde{\Lambda} = \Lambda \cup \{\lambda_{n+1}\}$. Then, the Cholesky factor of $A_{K,\tilde{\Lambda}}$ is given by the block matrix*

$$\tilde{L} = \begin{pmatrix} L & 0 \\ (L^{-1} \cdot R)^T & P_\Lambda(\lambda_{n+1}) \end{pmatrix},$$

where $R = (\langle \lambda_1, \lambda_{n+1} \rangle_K, ..., \langle \lambda_n, \lambda_{n+1} \rangle_K)^T \in \mathbb{R}^n$. Consequently, the inverse Cholesky factor is given by

$$\tilde{L}^{-1} = \begin{pmatrix} L^{-1} & 0 \\ -P_\Lambda(\lambda_{n+1})^{-1} \cdot l^T & P_\Lambda(\lambda_{n+1})^{-1} \end{pmatrix},$$

where $l = (\lambda_{n+1}(l_1), ..., \lambda_{n+1}(l_n))^T = A_{K,\Lambda}^{-1} \cdot R = {L^{-1}}^T \cdot L^{-1} \cdot R$, see also Corollary 6.4.

Proof. Due to our assumptions and part (2) of Corollary 5.4, we must have $P_\Lambda(\lambda_{n+1}) > 0$. We can compute

$$\tilde{L} \cdot \tilde{L}^T = \begin{pmatrix} L \cdot L^T & L \cdot L^{-1} \cdot R \\ R^T \cdot {L^T}^{-1} \cdot L^T & \|L^{-1} \cdot R\|_2^2 + P_\Lambda(\lambda_{n+1})^2 \end{pmatrix} = \begin{pmatrix} A_{K,\Lambda} & R \\ R^T & \|L^{-1} \cdot R\|_2^2 + P_\Lambda(\lambda_{n+1})^2 \end{pmatrix}.$$

The desired equality follows from Corollary 6.4, as we have

$$\|L^{-1} \cdot R\|_2^2 + P_\Lambda(\lambda_{n+1})^2 = R^T \cdot A_{K,\Lambda}^{-1} \cdot R + P_\Lambda(\lambda_{n+1})^2 = \langle \lambda_{n+1}, \lambda_{n+1} \rangle_K.$$

The second part can be verified via basic matrix multiplication as well. □

Recall that we can evaluate the power function with the Newton basis and the formula from Corollary 5.4, part (4). Due to the structure of the updated (inverse) Cholesky factor from Lemma 6.9, we can make the following conclusions:

(i) To update the Newton basis, we can keep the current set of basis functions and compute an additional basis element. This leads to an efficient recursive computation of the Newton basis. As a consequence, we can write $\tilde{\mathcal{N}} = \{\mathfrak{n}_1, ..., \mathfrak{n}_{n+1}\}$ instead of $\tilde{\mathcal{N}} = \{\tilde{\mathfrak{n}}_1, ..., \tilde{\mathfrak{n}}_{n+1}\}$.

(ii) The generalized Vandermonde matrix of the updated basis set $\tilde{\mathcal{N}} = \{\mathfrak{n}_1, ..., \mathfrak{n}_{n+1}\}$ is given by

$$\begin{pmatrix} \lambda_1(\mathfrak{n}_1) & \dots & \lambda_1(\mathfrak{n}_{n+1}) \\ \vdots & \ddots & \vdots \\ \lambda_{n+1}(\mathfrak{n}_1) & \dots & \lambda_{n+1}(\mathfrak{n}_{n+1}) \end{pmatrix} = \tilde{L} = \begin{pmatrix} L & 0 \\ (L^{-1} \cdot R)^T & P_\Lambda(\lambda_{n+1}) \end{pmatrix},$$

see Remark 6.7. Hence, we get the following equation (see also [98, Korollar 5.1.1]):

$$\lambda_{n+1}(\mathfrak{n}_{n+1}) = P_\Lambda(\lambda_{n+1}) \tag{6.5}$$

(iii) During the update of Lemma 6.9, we need to divide by the term $P_\Lambda(\lambda_{n+1})$, which represents the normalization factor in the Gram-Schmidt orthogonalization process. In order to avoid numerical problems, the power function value needs to be rather high, as dividing by values near zero leads to severe instabilities. Generally, the power function value affects the overall stability of the interpolation method. More details follow in Section 7.1.

Finally, we are in the position to efficiently update the new interpolant with respect to the updated set $\tilde{\Lambda} = \Lambda \cup \{\lambda_{n+1}\}$ (cf. [99, Lemma 5]).

Corollary 6.10. *Under the assumptions of Lemma 6.9, let $f \in \mathcal{H}_K$ and $\mathcal{N} = \{\mathfrak{n}_1, ..., \mathfrak{n}_{n+1}\}$ denote the updated Newton basis of $S_{K,\tilde{\Lambda}}$. Then, we can write the updated interpolant as*

$$I_{K,\tilde{\Lambda}}(f) = I_{K,\Lambda}(f) + \frac{\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda}(f))}{P_\Lambda(\lambda_{n+1})} \cdot \mathfrak{n}_{n+1}.$$

Proof. It is clear that we have

$$s := I_{K,\Lambda}(f) + \frac{\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda}(f))}{P_\Lambda(\lambda_{n+1})} \cdot \mathfrak{n}_{n+1} \in S_{K,\tilde{\Lambda}}.$$

We verify that s interpolates f on $\tilde{\Lambda}$. For $1 \leq i < n+1$, we have $\lambda_i(\mathfrak{n}_{n+1}) = 0$ due to Theorem 6.6 part (2) and therefore

$$\lambda_i(s) = \lambda_i(I_{K,\Lambda}(f)) = \lambda_i(f).$$

For $i = n+1$, we can use equation (6.5) to get

$$\begin{aligned}\lambda_{n+1}(s) &= \lambda_{n+1}(I_{K,\Lambda}(f)) + \frac{\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda}(f))}{P_\Lambda(\lambda_{n+1})} \cdot \lambda_{n+1}(\mathfrak{n}_{n+1}) \\ &= \lambda_{n+1}(I_{K,\Lambda}(f)) + \frac{\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda}(f))}{P_\Lambda(\lambda_{n+1})} \cdot P_\Lambda(\lambda_{n+1}) \\ &= \lambda_{n+1}(f).\end{aligned}$$

Since the interpolation problem is uniquely solveable in $S_{K,\tilde{\Lambda}}$, we must have

$$I_{K,\tilde{\Lambda}}(f) = I_{K,\Lambda}(f) + \frac{\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda}(f))}{P_\Lambda(\lambda_{n+1})} \cdot \mathfrak{n}_{n+1}.$$

□

In total, the previous discussion delivers a recursive algorithm to compute the interpolant as

$$I_{K,\Lambda_{n+1}}(f) = I_{K,\Lambda_1}(f) + \sum_{i=2}^{n+1} \frac{\lambda_i(f) - \lambda_i(I_{K,\Lambda_{i-1}}(f))}{P_{\Lambda_{i-1}}(\lambda_i)} \cdot \mathfrak{n}_i, \tag{6.6}$$

where $\Lambda_i = \{\lambda_1, ..., \lambda_i\}$ for $i = 1, ..., n+1$ and the initial interpolant is given by

$$I_{K,\Lambda_1}(f) = \frac{\lambda_1(f)}{\|\lambda_1\|_K} \cdot \mathfrak{n}_1.$$

Recall that the power function can be computed recursively using the formula

$$P_{\Lambda_{i+1}}(\lambda)^2 = \langle \lambda, \lambda \rangle_K - \sum_{j=1}^{i+1} \lambda(\mathfrak{n}_j)^2 = P_{\Lambda_i}(\lambda)^2 - \lambda(\mathfrak{n}_{i+1})^2 \qquad \text{for all } \lambda \in \mathcal{H}_K^* \tag{6.7}$$

and $i = 1, ..., n$, see Corollary 5.4, part (4). As it was shown in [105], the algorithm can be implemented more efficiently such that it only stores the evaluations

$$\lambda(\mathfrak{n}_i) \qquad \text{for } i = 1, ..., n+1$$

for a certain subset of functionals $\lambda \in \mathcal{H}_K^*$ in order to compute the interpolant, without storing the coefficients of the Newton basis in the representation of the standard basis. This implementation is based on the following update formula (cf. [105, Section 10]).

Theorem 6.11. *In the setting of Corollary 6.10, we have the update formula*

$$\lambda(\mathfrak{n}_{n+1}) = P_\Lambda(\lambda_{n+1})^{-1} \cdot \left(\langle \lambda, \lambda_{n+1} \rangle_K - \sum_{i=1}^{n} \lambda(\mathfrak{n}_i) \cdot \lambda_{n+1}(\mathfrak{n}_i) \right) \qquad \textit{for all } \lambda \in \mathcal{H}_K^*. \tag{6.8}$$

Proof. As we have already discussed after the proof of Theorem 6.6, we can also compute the basis element $\mathfrak{n}_{n+1}$ via the update formula from the Gram-Schmidt orthogonalization process with the additional element $\lambda_{n+1}^y K(\cdot, y)$:

$$\begin{aligned}\mathfrak{n}_{n+1} &= \frac{1}{\|\lambda_{n+1}^y K(\cdot,y) - I_{K,\Lambda}(\lambda_{n+1}^y K(\cdot,y))\|_K} \cdot \left(\lambda_{n+1}^y K(\cdot,y) - \sum_{i=1}^{n} \langle \lambda_{n+1}^y K(\cdot,y), \mathfrak{n}_i\rangle_K \cdot \mathfrak{n}_i\right) \\ &= P_\Lambda(\lambda_{n+1})^{-1} \cdot \left(\lambda_{n+1}^y K(\cdot,y) - \sum_{i=1}^{n} \lambda_{n+1}(\mathfrak{n}_i) \cdot \mathfrak{n}_i\right).\end{aligned}$$

Inserting this representation into $\lambda \in \mathcal{H}_K^*$ yields the desired update formula (6.8). □

Remark 6.12. The update formula leads to a more efficient algorithm in terms of storage and computational complexity, as it can easily be vectorized to treat large sets of data. However, there remains a huge drawback in this recursion. Note that the computation of the new basis element $\mathfrak{n}_{n+1}$ and its evaluations depend on all previous basis elements of the Newton basis. Thus, the update formula (6.8) gets more costly as the number of data points grows larger. Until now, no general improvement is known to avoid the quadratic growth of the cost for evaluating a fixed functional in all Newton basis functions. For example, a concept like a *3-term-recursion* as known from univariate orthogonal polynomials (see, e.g., [70, Section 4.4]) would significantly improve the computational efficiency.

We close this chapter with a result regarding the expansion of functions into orthogonal series. Under the convergence conditions from Chapter 5, the recursively constructed Newton basis forms a complete orthonormal system in the restricted space $\mathcal{H}_{K,\Gamma}$ (cf. [98, Korollar 4.4.3 & 4.4.5]).

Theorem 6.13. *Let $\Gamma \subset \mathcal{H}_K^*$ and $(\Lambda_n)_{n\in\mathbb{N}}$ be a nested sequence of finite linearly independent subsets of Γ with $|\Lambda_n| = n$ for all $n \in \mathbb{N}$ that satisfies the property*

$$P_{\Lambda_n}(\lambda) \xrightarrow{n\to\infty} 0 \quad \text{for all } \lambda \in \Gamma \qquad \text{or} \qquad h_{\Lambda_n,\Gamma} \xrightarrow{n\to\infty} 0.$$

Moreover, let $\mathcal{N} = \{\mathfrak{n}_i \mid i \in \mathbb{N}\}$ denote the respective set of Newton basis functions. Then, every function $f \in \mathcal{H}_{K,\Gamma}$ can be written as

$$f = \sum_{i=1}^{\infty} \langle f, \mathfrak{n}_i\rangle_K \cdot \mathfrak{n}_i = \frac{\lambda_1(f)}{\|\lambda_1\|_K} \cdot \mathfrak{n}_1 + \sum_{i=2}^{\infty} \frac{\lambda_i(f) - \lambda_i(I_{K,\Lambda_{i-1}}(f))}{P_{\Lambda_{i-1}}(\lambda_i)} \cdot \mathfrak{n}_i.$$

In particular, we then have

$$\lambda^y K(\cdot,y) = \sum_{i=1}^{\infty} \lambda(\mathfrak{n}_i) \cdot \mathfrak{n}_i \qquad \text{and} \qquad \langle \lambda, \tilde{\lambda}\rangle_K = \sum_{i=1}^{\infty} \lambda(\mathfrak{n}_i) \cdot \tilde{\lambda}(\mathfrak{n}_i) \qquad \text{for all } \lambda, \tilde{\lambda} \in \Gamma.$$

Proof. The first equality follows from Theorem 5.6, 5.9 and the representation (6.6). If $f = \lambda^y K(\cdot y)$ for a functional $\lambda \in \Gamma$, we get

$$\lambda^y K(\cdot y) = \sum_{i=1}^{\infty} \langle \lambda^y K(\cdot y), \mathfrak{n}_i\rangle_K \cdot \mathfrak{n}_i = \sum_{i=1}^{\infty} \lambda(\mathfrak{n}_i) \cdot \mathfrak{n}_i.$$

Given another $\tilde{\lambda} \in \Gamma$, we use its continuity on $\mathcal{H}_K$ to derive

$$\langle \lambda, \tilde{\lambda}\rangle_K = \tilde{\lambda}(\lambda^y K(\cdot y)) = \tilde{\lambda}\left(\sum_{i=1}^{\infty} \lambda(\mathfrak{n}_i) \cdot \mathfrak{n}_i\right) = \sum_{i=1}^{\infty} \lambda(\mathfrak{n}_i) \cdot \tilde{\lambda}(\mathfrak{n}_i).$$

□

Remark 6.14. Even if the previously assumed convergence conditions do not hold, we still have

$$\left(\frac{\lambda_1(f)}{\|\lambda_1\|_K}\right)^2 + \sum_{i=2}^{\infty} \left(\frac{\lambda_i(f) - \lambda_i(I_{K,\Lambda_{i-1}}(f))}{P_{\Lambda_{i-1}}(\lambda_i)}\right)^2 = \sum_{i=1}^{\infty} \langle f, \mathfrak{n}_i\rangle_K^2 = \lim_{n\to\infty} \|I_{K,\Lambda_n}\|_K^2 \le \|f\|_K^2$$

for each $f \in \mathcal{H}_{K,\Gamma}$. Hence,

$$\frac{\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda_n}(f))}{P_{\Lambda_n}(\lambda_{n+1})} \xrightarrow{n\to\infty} 0.$$

This shows that the interpolation error at the update functional decays faster than the respective power function value, both with respect to the previously selected functionals. We use this fact later in the proof of Theorem 7.19.

7. Greedy Data Selection Algorithms

Throughout the last chapter, we have assumed that the new functional λ_{n+1} is always given and we just have to proceed with the update of the orthogonal basis and the interpolant. But this is not a realistic scenario, as we are usually confronted with a mathematical problem, e.g. a partial differential equation, and need to choose the data points on our own. To get an idea of how to choose well-suited data points for the interpolation process, we generalize stability estimates from standard interpolation and explain the *uncertainty relation* in the context of generalized interpolation in Section 7.1. With our general convergence results from Chapter 5, we can then derive generalized versions of greedy algorithms known from the standard case (cf. [41], [43], [119], [145]) in Section 7.2. Most of these greedy algorithms for generalized interpolation were already introduced and analyzed in [146]. We build on that work and prove convergence for the so-called *β-greedy algorithms* under rather mild assumptions on the kernel and the interpolation domain. Besides the β-greedy approach, we analyze generalized *geometric* approaches that rely on the distance function within the dual space or the parameter space in case a parametrization of the functional set exists (cf. Section 5.3). However, one has to keep in mind that we only deal with finite data in practice. As an outlook, we mention the concept of *data thinning* and discuss how greedy algorithms can help to thin out large, redundant data point sets to improve the computational stability and efficiency of the reconstruction method. In particular, this idea can be combined with known regularization tools from kernel-based approximation, which we discuss in Section 7.3. To this end, the penalization of the native space norm is generalized by penalizing the evaluation of certain regularizing functionals from the native dual space. We close the first part of this thesis with a pseudo code that summarizes the derived framework for kernel-based generalized interpolation.

Following Chapter 5, we select the data points from a given superset $\Gamma \subset \mathcal{H}_K^*$ that depends on the problem we want to solve. This results in nested sequences of subsets from Γ. It is well-known that the choice of data points has a strong effect on the accuracy and numerical stability of the interpolation method, see e.g. [37]. Hence, we are interested in suitable selection strategies for new data points that focus on at least one of these two properties. Concerning the numerical stability, we are particularly interested in limiting the growth of the spectral condition numbers

$$\operatorname{cond}_2\left(A_{K,\Lambda_n}\right) = \frac{\sigma_{\max}(A_{K,\Lambda_n})}{\sigma_{\min}(A_{K,\Lambda_n})} \qquad \text{for } n \in \mathbb{N}, \tag{7.1}$$

where $\sigma_{\min}$ and $\sigma_{\max}$ denote the lowest and highest singular values. If the condition number is too high, the reconstruction gets highly sensitive to noise and therefore useless for practical applications.

7.1. Uncertainty relation

For the stability analysis, we want to derive estimates for the smallest and highest eigenvalues of the updated interpolation matrix depending on the newly added functional. This gives an insight into how to select new functionals in order to obtain a sufficiently high numerical stability.

We start this section with a general result on the behavior of the condition numbers, which is a direct consequence of *Cauchy's interlacing theorem* (see, e.g., [65, Theorem 4.3.17]).

Theorem 7.1. *Let $\Lambda_n \subset \mathcal{H}_K^*$ be linearly independent and $\lambda_{n+1} \in \mathcal{H}_K^* \setminus \Lambda_n$. Set $\Lambda_{n+1} = \Lambda_n \cup \{\lambda_{n+1}\}$, then we have*

$$\sigma_{\min}(A_{K,\Lambda_{n+1}}) \leq \sigma_{\min}(A_{K,\Lambda_n}) \qquad \text{and} \qquad \sigma_{\max}(A_{K,\Lambda_{n+1}}) \geq \sigma_{\max}(A_{K,\Lambda_n}).$$

In combination with the formula (7.1), Theorem 7.1 shows that the sequence of condition numbers is increasing. Hence, adding new functionals always decreases the numerical stability to some degree. Similar to the standard interpolation case (cf. [143, Chapter 12]), the highest singular value grows at most linearly depending on the number of data points if the underlying set of functionals is bounded.

Proposition 7.2. *Let $\Gamma \subset \mathcal{H}_K^*$ be bounded and $\Lambda_n = \{\lambda_1, ..., \lambda_n\} \subset \Gamma$ be linearly independent. Then, there is a constant $C > 0$, independent of n, such that the highest singular value of the interpolation matrix can be bounded by*

$$\sigma_{\max}(A_{K,\Lambda_n}) \leq C \cdot n.$$

Proof. Since Γ is bounded, we have

$$C := \sup_{\lambda \in \Gamma} \|\lambda\|_K^2 < \infty.$$

Due to the *Gershgorin circle theorem* (see, e.g., [65, Theorem 6.1.1]), there is $i \in \{1, ..., n\}$ with

$$\sigma_{\max}(A_{K,\Lambda_n}) - \langle \lambda_i, \lambda_i \rangle_K \leq |\sigma_{\max}(A_{K,\Lambda_n}) - \langle \lambda_i, \lambda_i \rangle_K| \leq \sum_{i \neq j} |\langle \lambda_i, \lambda_j \rangle_K|.$$

Note that, due to Cauchy's inequality, we also have

$$|\langle \lambda_i, \lambda_j \rangle_K| \leq \|\lambda_i\|_K \cdot \|\lambda_j\|_K \leq C \qquad \text{for } i \neq j,$$

and therefore

$$\sigma_{\max}(A_{K,\Lambda_n}) \leq \langle \lambda_i, \lambda_i \rangle_K + \sum_{i \neq j} |\langle \lambda_i, \lambda_j \rangle_K| \leq C \cdot n.$$

□

Additionally, we can find an upper bound for the smallest singular value of the updated interpolation matrix to estimate its decay. This upper bound mainly involves the power function value at the additional functional, see [143, Theorem 12.1] for the standard interpolation case.

Theorem 7.3. *Under the assumptions of Theorem 7.1 with $\Lambda_n = \{\lambda_1, ..., \lambda_n\}$, we have the estimate*

$$\sigma_{\min}(A_{K,\Lambda_{n+1}}) \leq P_{\Lambda_n}(\lambda_{n+1})^2 \cdot \left(1 + \sum_{j=1}^{n} \lambda_{n+1}(l_j)^2\right)^{-1},$$

where $\mathcal{L} = \{l_1, ..., l_n\}$ denotes the Lagrangian basis of S_{K,Λ_n}.

Proof. From Corollary 6.4, we know that

$$P_{\Lambda_n}(\lambda_{n+1})^2 = \langle \lambda_{n+1}, \lambda_{n+1} \rangle_K - l^T \cdot A_{K,\Lambda_n} \cdot l,$$

where $l = (\lambda_{n+1}(l_1), ..., \lambda_{n+1}(l_n))^T$. Setting $z = (-l, 1)^T \in \mathbb{R}^{n+1}$, we can use the identity

$$A_{K,\Lambda_n} \cdot l = R = (\langle \lambda_1, \lambda_{n+1} \rangle_K, ..., \langle \lambda_n, \lambda_{n+1} \rangle_K)^T$$

to compute

$$z^T \cdot A_{K,\Lambda_{n+1}} \cdot z = l^T \cdot A_{K,\Lambda_n} \cdot l - 2 \cdot R^T \cdot l + \langle \lambda, \lambda \rangle_K = \langle \lambda, \lambda \rangle_K - l^T \cdot A_{K,\Lambda_n} \cdot l = P_{\Lambda_n}(\lambda)^2.$$

By applying the *Courant-Fischer theorem* (see, e.g., [65, Theorem 4.2.6]), we get the desired result

$$\sigma_{\min}(A_{K,\Lambda_{n+1}}) \leq \frac{z^T \cdot A_{K,\Lambda_{n+1}} \cdot z}{z^T \cdot z} = P_{\Lambda_n}(\lambda_{n+1})^2 \cdot \left(1 + \sum_{j=1}^{n} \lambda_{n+1}(l_j)^2\right)^{-1}.$$

□

Remark 7.4. According to Proposition 7.2, the largest eigenvalue of the interpolation matrix grows at most at a linear rate, which means that it is rather unproblematic for the numerical stability. In contrast, the smallest eigenvalue usually causes numerical instabilities, as it is bounded by the power function value due to Theorem 7.3. Depending on the regularity of the kernel, the power function value can decay very fast, which then translates to the smallest eigenvalue. In the context of standard interpolation, upper bounds for the power function of specific kernels can be found in [143, Table 11.1], whereas lower bounds for the smallest eigenvalue in terms of the separation distance are given in [143, Table 12.1].

Moreover, we want to remark that the estimate from Theorem 7.3 matches with Remark 5.5, where we observed that the power function measures the linear dependence of given functionals. If the power function value is small, the set of functionals is nearly linearly dependent, which means that the columns of the interpolation matrix are nearly linearly dependent. Therefore, the interpolation matrix is almost singular and its smallest eigenvalue is close to zero.

For the sake of numerical stability, we want the power function values $P_{\Lambda_n}(\lambda)$ to be large, as this leads to a large upper bound in Theorem 7.3. A large power function value also stabilizes the update formulas from Corollary 6.10 and Theorem 6.11. But, if we focus on the approximation quality, a fast decay of the power function is desirable to ensure fast (pointwise) convergence, see Theorem 5.6 and Remark 5.7. This tradeoff between approximation quality and stability is called the *uncertainty principle* in kernel-based approximation (cf. [116]), and we cannot have arbitrarily good approximation quality and arbitrarily good stability at the same time within this approach.

7.2. Greedy algorithms

In the following, we discuss generalizations of single-point selection algorithms for kernel-based interpolation methods that have been developed in the last two decades. These algorithms belong to the class of *greedy algorithms* (cf. [135]), which solve a specific optimization problem in each step. In our case, we select the new functional as a maximizer of a given error functional $\eta : \Gamma \to \mathbb{R}$, i.e. we apply the selection rule

$$\eta(\lambda_{n+1}) = \sup_{\lambda \in \Gamma} \eta(\lambda).$$

The idea of this selection rule is to select a new data point from an area where the error is still large so that the error around this new data point significantly decreases with the update of the interpolant. Usually, the error functional $\eta \equiv \eta_{f,\Lambda_n}$ depends on the currently selected data set Λ_n and the function f that we want to interpolate. Thus, it adapts to the current state of the interpolation process. Of course, there are several choices for the selection rule η, which are motivated by the wishes and needs of the particular problem.

Remark 7.5. In most cases, the error functional η is bounded and depends continuously on the input argument λ. But this does not guarantee that η attains a maximum on Γ, as Γ does not have to be compact in general. This problem can be fixed (theoretically) by introducing a parameter $\gamma \in (0, 1)$ and changing the selection rule to

$$\eta(\lambda_{n+1}) \geq \gamma \cdot \sup_{\lambda \in \Gamma} \eta(\lambda)$$

see also [42]. Note that these methods lead to an iterative point selection, where we usually have to choose an initial value $\lambda_0 \in \Gamma$. Since the analysis of the greedy methods (e.g. convergence results) does not depend on the initial value, we ignore this initial selection throughout this chapter. We just have to keep in mind that we have to make an initial choice if we implement the algorithms.

According to the convergence results from Chapter 5, we aim to construct sequences of nested subsets that satisfy

$$P_{\Lambda_n}(\lambda) \xrightarrow{n\to\infty} 0 \quad \text{for all } \lambda \in \Gamma \qquad \text{or} \qquad h_{\Lambda_n,\Gamma} \xrightarrow{n\to\infty} 0.$$

To this end, we restrict ourselves to the case that the superset Γ is totally bounded (cf. Definition A.3), as the notion of the fill distance is closely related to this kind of set. In fact, the geometric convergence criterion can only be satisfied for totally bounded sets.

Proposition 7.6. *Let (Z, d_Z) be a metric space. There is a nested sequence of finite subsets $(A_n)_{n\in\mathbb{N}}$ of Z with*

$$h_{A_n,Z} \xrightarrow{n\to\infty} 0 \tag{7.2}$$

if and only if Z is totally bounded.

Proof. Let $\varepsilon > 0$ and $(A_n)_{n\in\mathbb{N}}$ be a nested sequence of finite subsets satisfying (7.2). Then we can find $N \in \mathbb{N}$ with

$$\sup_{z\in Z} \operatorname{dist}(z, A_N) = h_{A_N,Z} < \varepsilon,$$

where A_N is finite. This immediately gives

$$Z \subset \bigcup_{a\in A_N} B_\varepsilon(a),$$

so that Z is totally bounded. Conversely, let Z be totally bounded. For every $n \in \mathbb{N}$, we can find a finite set $\tilde{A}_n$ with

$$Z \subset \bigcup_{a\in \tilde{A}_n} B_{1/n}(a),$$

or equivalently, $h_{\tilde{A}_n,Z} \leq 1/n$. Set $A_n := \bigcup_{i=1}^n \tilde{A}_i$ for each $n \in \mathbb{N}$, then $(A_n)_{n\in\mathbb{N}}$ is a nested sequence of finite subsets that satisfies

$$h_{A_n,Z} \leq h_{\tilde{A}_n,Z} \leq 1/n \xrightarrow{n\to\infty} 0.$$

□

Remark 7.7. For given supersets $\Gamma \subset \mathcal{H}_K^*$, it is not a trivial task to decide whether it is totally bounded or not. But, if Γ can be parametrized by the pair (Ω, ϱ) (cf. Section 5.3), where Ω is totally bounded and ϱ is uniformly continuous, Lemma A.4 assures that Γ is also totally bounded. If Ω is even compact, the requirements can be lowered to ϱ being continuous, so that Γ is compact as well.

Our further analysis relies on the decay of the separation terms $P_{\Lambda_n}(\lambda_{n+1})$ and $\operatorname{dist}(\lambda_{n+1}, \Lambda_n)$ for $n \in \mathbb{N}$. Due to the assumption that Γ is totally bounded, the additional functional λ_{n+1} cannot be separated from the set Λ_n as $n \to \infty$.

Lemma 7.8. *Let (Z, d_Z) be totally bounded and $(a_n)_{n\in\mathbb{N}}$ be a sequence in Z. Set $A_n = \{a_1, ..., a_n\}$ for $n \in \mathbb{N}$, then we have the convergence*

$$\operatorname{dist}(a_{n+1}, A_n) \xrightarrow{n\to\infty} 0.$$

In the special case that $Z = \Gamma \subset \mathcal{H}_K^*$ and $A_n = \Lambda_n = \{\lambda_1, ..., \lambda_n\}$ is linearly independent for each $n \in \mathbb{N}$, we have

$$P_{\Lambda_n}(\lambda_{n+1}) \xrightarrow{n\to\infty} 0.$$

Proof. Given $\varepsilon > 0$, we can find $z_1, ..., z_M \in Z$ such that

$$Z \subset \bigcup_{j=1}^M B_{\varepsilon/2}(z_j).$$

We set

$$I_j = \{n \in \mathbb{N} \mid a_n \in B_{\varepsilon/2}(z_j)\} \qquad \text{for } j = 1, ..., M$$

and

$$N_j = \begin{cases} \min(I_j) & \text{if } I_j \neq \emptyset \\ 0 & \text{if } I_j = \emptyset \end{cases} \qquad \text{for } j = 1, ..., M.$$

For $N = \max\{N_j \mid j = 1, ..., M\}$ and $n > N$, there is $1 \leq j \leq M$ and $a \in A_N$ such that

$$a_n, a \in B_{\varepsilon/2}(z_j)$$

due to the choice of N. Using the triangle inequality, this implies

$$\operatorname{dist}(a_n, A_{n-1}) \leq \operatorname{dist}(a_n, A_N) \leq d_Z(a_n, a) \leq d_Z(a_n, z_j) + d_Z(z_j, a) < \varepsilon.$$

The second part follows from the estimate

$$P_{\Lambda_n}(\lambda_{n+1}) \leq \operatorname{dist}(\lambda_{n+1}, \Lambda_n) \qquad \text{for all } n \in \mathbb{N}.$$

□

Besides the pointwise convergence with respect to Γ, we are also interested in showing convergence with respect to the native space norm. In general, weak convergence in $\mathcal{H}_{K,\Gamma}$ does not necessarily result in normwise convergence in infinite-dimensional Hilbert spaces. But, in the case that we have a sequence of projections, the sequence must have a normwise limit, so that weak convergence with respect to Γ is sufficient.

Lemma 7.9. *Let $\Gamma \subset \mathcal{H}_K^*$. For $f, g \in \mathcal{H}_{K,\Gamma}$, we have the equivalence*

$$f = g \qquad \Longleftrightarrow \qquad \lambda(f) = \lambda(g) \quad \text{for all } \lambda \in \Gamma.$$

Moreover, if $(\Lambda_n)_{n\in\mathbb{N}}$ is a nested sequence of finite linearly independent subsets of Γ satisfying

$$\lambda(I_{K,\Lambda_n}(f)) \xrightarrow{n\to\infty} \lambda(f) \qquad \text{for all } \lambda \in \Gamma,$$

and a fixed $f \in \mathcal{H}_{K,\Gamma}$, we also have normwise convergence

$$\|f - I_{K,\Lambda_n}(f)\|_K \xrightarrow{n\to\infty} 0.$$

Proof. For the first part, notice that we have

$$\lambda(f) = \lambda(g) \quad \text{for all } \lambda \in \Gamma \qquad \Longleftrightarrow \qquad f - g \perp \operatorname{span}_{\mathbb{R}} \{\lambda^y K(\cdot, y) \mid \lambda \in \Gamma\}$$

for $f, g \in \mathcal{H}_{K,\Gamma}$ due to the generalized reproduction property from Theorem 3.12. With the continuity of the inner product, this is equivalent to $f - g \perp \mathcal{H}_{K,\Gamma}$ and $f = g$. For the second part, consider the sequence

$$\left(\|I_{K,\Lambda_n}(f)\|_K^2\right)_{n\in\mathbb{N}}$$

for a fixed $f \in \mathcal{H}_{K,\Gamma}$. Due to the properties

$$I_{K,\Lambda_n}(I_{K,\Lambda_m}(f)) = I_{K,\Lambda_n}(f) \quad \text{for } m > n \qquad \text{and} \qquad \|I_{K,\Lambda_n}(f)\|_K \leq \|f\|_K \quad \text{for } m, n \in \mathbb{N},$$

this sequence is bounded and monotonically increasing. Hence, it is convergent and a Cauchy sequence. Using the projection properties

$$\begin{aligned}\|I_{K,\Lambda_m}(f)\|_K^2 &= \|I_{K,\Lambda_m}(f) - I_{K,\Lambda_n}(I_{K,\Lambda_m}(f))\|_K^2 + \|I_{K,\Lambda_n}(I_{K,\Lambda_m}(f))\|_K^2 \\ &= \|I_{K,\Lambda_m}(f) - I_{K,\Lambda_n}(f)\|_K^2 + \|I_{K,\Lambda_n}(f)\|_K^2,\end{aligned}$$

of the interpolation operator, we get the identity

$$\|I_{K,\Lambda_m}(f) - I_{K,\Lambda_n}(f)\|_K^2 = \|I_{K,\Lambda_m}(f)\|_K^2 - \|I_{K,\Lambda_n}(f)\|_K^2$$

for $m > n$, so that $(I_{K,\Lambda_n}(f))_{n\in\mathbb{N}} \subset \mathcal{H}_{K,\Gamma}$ is a Cauchy sequence as well. Since $\mathcal{H}_{K,\Gamma}$ is complete, it has a normwise limit $g \in \mathcal{H}_{K,\Gamma}$. Then g is also the weak limit of this sequence so that we get

$$\lambda(g) = \lim_{n\to\infty} \lambda(I_{K,\Lambda_n}(f)) = \lambda(f) \qquad \text{for all } \lambda \in \Gamma.$$

due to our assumptions. The assertion now follows from the first part. □

For the introduction of the greedy selection strategies, we distinguish between target-independent and target-dependent algorithms. Large parts of the convergence analysis are shifted to Subsection 7.2.3, which unifies several approaches. Note that these selection rules only perform an infinite number of steps if the superset Γ contains an infinite linearly independent subset. Otherwise, the algorithms terminate after a finite number of iteration steps, as no more information can be added to the interpolation process. In our theoretical analysis, we always assume that the algorithms perform an infinite number of iteration steps, which of course does not apply in practical cases. We discuss a realistic use case in Subsection 7.2.5.

7.2.1. Target-independent algorithms

We start our discussion on greedy algorithms with the target-independent approaches. Here, the term *target-independent* means that the selection rule depends solely on the chosen kernel K, the superset Γ and the current data set Λ_n in the n-th iteration step, $n \in \mathbb{N}$. It does not depend on the given data values of the considered target function $f \in \mathcal{H}_{K,\Gamma}$ so that we can reuse the selected points when approximating different target functions. These methods mainly aim to maintain high numerical stability throughout the interpolation process while keeping sufficient approximation quality.

In Theorem 7.3, we deduced that the decay of the power function is a good indicator for the numerical stability of the reconstruction method. Moreover, the power function value represents the normalization factor in the Gram-Schmidt process and could therefore cause severe numerical problems in the orthogonalization process. Hence, a reasonable approach is to maximize the power function in each iteration step to optimize the upper bound of the smallest eigenvalue. This selection rule, which is given by

$$P_{\Lambda_n}(\lambda_{n+1}) = \sup_{\lambda \in \Gamma} P_{\Lambda_n}(\lambda) =: \|P_{\Lambda_n}\|_{\infty,\Gamma} \qquad \text{for } n \in \mathbb{N}, \tag{7.3}$$

is called the *P-greedy algorithm* (cf. [37], [41], [118]). In other words, we select the functional that has the largest distance to the span of the current data point set and therefore has the worst approximation error in terms of the native space norm. We can immediately observe that the sequence

$$(\|P_{\Lambda_n}\|_{\infty,\Gamma})_{n\in\mathbb{N}} = (P_{\Lambda_n}(\lambda_{n+1}))_{n\in\mathbb{N}}$$

is monotonically decreasing, as we have

$$\|P_{\Lambda_{n+1}}\|_{\infty,\Gamma} = P_{\Lambda_{n+1}}(\lambda_{n+2}) \leq P_{\Lambda_n}(\lambda_{n+2}) \leq \|P_{\Lambda_n}\|_{\infty,\Gamma} = P_{\Lambda_n}(\lambda_{n+1}).$$

Moreover, if λ_{n+1} is chosen via the P-greedy selection rule, it maximizes the evaluation at the updated Newton basis element (cf. [98, Korollar 5.2.1]).

Proposition 7.10. *Let $\lambda_{n+1} \notin \mathrm{span}_{\mathbb{R}}(\Lambda_n)$ be chosen via the P-greedy algorithm (7.3) and $\mathfrak{n}_{n+1}$ be the additional basis element of the updated Newton basis, see Lemma 6.9. Then we have*

$$|\lambda(\mathfrak{n}_{n+1})| \leq \lambda_{n+1}(\mathfrak{n}_{n+1}) = \|P_{\Lambda_n}\|_{\infty,\Gamma} \qquad \textit{for all } \lambda \in \Gamma.$$

Proof. With equation (6.7), we have

$$\lambda(\mathfrak{n}_{n+1})^2 = P_{\Lambda_n}(\lambda)^2 - P_{\Lambda_{n+1}}(\lambda)^2 \leq P_{\Lambda_n}(\lambda)^2 \leq \|P_{\Lambda_n}\|^2_{\infty,\Gamma} \qquad \text{for all } \lambda \in \Gamma.$$

The claim follows from equation (6.5) combined with the selection criterion

$$\lambda_{n+1}(\mathfrak{n}_{n+1}) = P_{\Lambda_n}(\lambda_{n+1}) = \|P_{\Lambda_n,\Gamma}\|_{\infty,\Gamma}.$$

□

Another geometric approach is given by the *geometric greedy algorithm*, which selects the new functional via

$$\mathrm{dist}(\lambda_{n+1}, \Lambda_n) = \sup_{\lambda \in \Gamma} \mathrm{dist}(\lambda, \Lambda_n) = h_{\Lambda_n,\Gamma} \qquad \text{for } n \in \mathbb{N} \tag{7.4}$$

to create an optimal separation between the new functional and the current data set (cf. [37, Section 3.1]). This selection rule can be viewed as a simplified version of the P-greedy algorithm, as we try to maintain numerical stability by the separation with less computational complexity within the selection rule.

Since the geometric greedy selection only depends on the distance function $\operatorname{dist}(\cdot, \Lambda_n)$, it does not include any information about the linear dependence of the functionals. Hence, we have to make sure that the superset Γ is already linearly independent so that we cannot choose linearly dependent functionals. If Γ is totally bounded, the geometric greedy selection induces a convergent interpolation method.

Theorem 7.11. *Let $\Gamma \subset \mathcal{H}_K^*$ be totally bounded and linearly independent. Moreover, let $(\Lambda_n)_{n\in\mathbb{N}}$ be chosen via the geometric greedy algorithm (7.4). Then we have $h_{\Lambda_n,\Gamma} \xrightarrow{n\to\infty} 0$ and the convergence*

$$\|f - I_{K,\Lambda_n}(f)\|_K \xrightarrow{n\to\infty} 0 \qquad \text{for all } f \in \mathcal{H}_{K,\Gamma}.$$

Proof. With the result of Lemma 7.8, we conclude

$$h_{\Lambda_n,\Gamma} = \operatorname{dist}(\lambda_{n+1}, \Lambda_n) \xrightarrow{n\to\infty} 0.$$

The convergence of the interpolation method then follows from Theorem 5.9. □

Not only does the geometric greedy selection lead to a convergent interpolation method, but it also produces *quasi-uniform* data point sets in Γ (cf. [41, Lemma 5.1]).

Definition 7.12. Let (Z, d_Z) be a metric space and $A \subset Z$ with $|A| \geq 2$. The `separation distance` of A is given by

$$q_A := \frac{1}{2} \inf_{\substack{x,y\in A \\ x\neq y}} d_Z(x, y).$$

Moreover, we call A `quasi-uniform` if there is a constant $c > 0$ such that

$$h_{A,Z} \leq c \cdot q_A.$$

Lemma 7.13. *In the setting of Theorem 7.11, we have*

$$2 \cdot q_{\Lambda_n} = h_{\Lambda_{n-1},\Gamma} \geq h_{\Lambda_n,\Gamma} \qquad \text{for } n \geq 2.$$

Proof. It is clear that $h_{\Lambda_{n-1},\Gamma} \geq h_{\Lambda_n,\Gamma}$ holds for all $n \geq 2$. We prove the remaining equality via induction. For the initial value $n = 2$, we have

$$2 \cdot q_{\Lambda_2} = \|\lambda_1 - \lambda_2\|_K = \operatorname{dist}(\lambda_2, \Lambda_1) = h_{\Lambda_1,\Gamma}$$

due to the geometric greedy choice. Assuming that $2 \cdot q_{\Lambda_n} = h_{\Lambda_{n-1},\Gamma}$ holds for a fixed $n \in \mathbb{N}$, we have

$$\begin{aligned} 2 \cdot q_{\Lambda_{n+1}} &= \min\left(2 \cdot q_{\Lambda_n}, \operatorname{dist}(\lambda_{n+1}, \Lambda_n)\right) \\ &= \min\left(h_{\Lambda_{n-1},\Gamma}, h_{\Lambda_n,\Gamma}\right) \\ &= h_{\Lambda_n,\Gamma}, \end{aligned}$$

where we used the monotonicity of the fill distance again. □

Note that the constant for the quasi-uniformity of Λ_n does not depend on the iteration step $n \in \mathbb{N}$. This is particularly useful if we want to derive error estimates for the interpolation method.

Remark 7.14. If a metric space (Z, d_Z) is connected, we can also show that

$$q_A \leq h_{A,Z} \tag{7.5}$$

holds for $A = \{a_1, ..., a_n\} \subset Z$ with $n \geq 2$ (cf. [107, Section 1]). By definition of q_A, the open balls

$$B_{q_A}(a_i) = \{z \in Z \mid d_Z(z, a_i) < q_A\} \qquad i = 1, ..., n$$

are disjoint, and since Z is connected, these sets cannot cover Z. Hence, there is $z \in Z$ which is not an element of the union of the balls with radius q_A so that we get

$$h_{A,Z} \geq \operatorname{dist}(z, A) \geq q_A.$$

The inequality (7.5) can also be proven if Z satisfies an interior cone condition, see also [143, Section 14.1].

Following the idea of Section 5.3, we can also perform a geometric greedy selection in the parameter space. If Γ is parametrized by (Ω, ϱ), where (Ω, d_Ω) is a metric space, we can apply the selection rule

$$\operatorname{dist}(x_{n+1}, X_n) = \sup_{x \in \Omega} \operatorname{dist}(x, X_n) = h_{X_n,\Omega} \qquad \text{for } n \in \mathbb{N}. \tag{7.6}$$

Under certain regularities of the mapping ϱ, the distances in the parameter space Ω reflect the distances in Γ. Recall that, in most cases, the treatment of the parameter space is much easier and less costly in comparison to the dual space. The convergence of this algorithm for totally bounded sets Ω is a direct conclusion of Lemma 7.8 and Theorem 5.13, where we assumed ϱ to be uniformly continuous.

Theorem 7.15. *Let (Ω, ϱ) be a uniformly continuous parametrization of $\Gamma \subset \mathcal{H}_K^*$, where Ω is totally bounded and Γ is linearly independent. Moreover, let $(X_n)_{n\in\mathbb{N}}$ be chosen via the selection rule (7.6). Then we have $h_{X_n,\Omega} \xrightarrow{n\to\infty} 0$ and the convergence*

$$\|f - I_{K,\varrho(X_n)}(f)\|_K \xrightarrow{n\to\infty} 0 \qquad \text{for all } f \in \mathcal{H}_{K,\Gamma}.$$

In the special case that $\Omega \subset \mathbb{R}^m$, $m \in \mathbb{N}$, we can even state decay rates for the fill distances $h_{X_n,\Omega}$ in dependence of the dimension $m \in \mathbb{N}$ (cf. [143, Proposition 14.1]). The result is based on the fact that the geometric greedy algorithm produces quasi-uniform data sets.

Lemma 7.16. *In the setting of Theorem 7.15, let $\Omega \subset \mathbb{R}^m$ be measurable and bounded with positive Lebesgue measure* $\operatorname{vol}(\Omega) > 0$. *Then we can find constants $C_1, C_2 > 0$ such that*

$$C_1 \cdot n^{-1/m} \leq h_{X_n,\Omega} \leq C_2 \cdot n^{-1/m} \qquad \text{for } n \geq 2.$$

We can combine the results for this special case with our findings from Corollary 5.14 to derive the following estimate for the interpolation error.

Corollary 7.17. *Under the assumptions of Corollary 5.14 and Lemma 7.16, there is a constant $C > 0$ such that we can estimate*

$$|\lambda(f) - \lambda(I_{K,\varrho(X_n)}(f))| \leq C \cdot n^{-a/m} \cdot \|f\|_K \qquad \text{for all } f \in \mathcal{H}_{K,\Gamma}, \lambda \in \Gamma, n \geq 2,$$

where $a \in (0, 1]$ is the Hölder exponent of the parametrization mapping ϱ.

To end the theoretical discussion about this geometric approach, we remark that the decay rate $n^{-a/m}$ of the upper bound is not optimal to describe the decay of the interpolation error in most cases, as it does not take into account the regularity of the kernel. In particular, the upper bound decays very slowly for large dimensions $m \in \mathbb{N}$. Nevertheless, this approach provides an error estimate under rather mild assumptions.

7.2.2. Target-dependent algorithms

Contrary to the target-independent algorithms, we can incorporate the target function $f \in \mathcal{H}_{K,\Gamma}$ into the selection rule. The most obvious algorithm might be to choose the functional that has the largest pointwise interpolation error regarding f, as we want to select data points in areas where we do not have high interpolation accuracy in terms of functional evaluation. Hence, we select the new functional via

$$|\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda_n}(f))| = \sup_{\lambda \in \Gamma} |\lambda(f) - \lambda(I_{K,\Lambda_n}(f))| \qquad \text{for } n \in \mathbb{N}, \tag{7.7}$$

which is known as the *f-greedy algorithm* (cf. [119]). Recall that Corollary 6.10 and Theorem 6.11 provide update formulas for the residual terms as the number of data points increases.

Besides measuring the pointwise error, we can use the native space norm as our primary error indicator. In particular, we are interested in minimizing the distance to the target function f in each step, i.e.

$$\|f - I_{K,\Lambda_n \cup \{\lambda_{n+1}\}}(f)\|_K = \inf_{\lambda \in \Gamma} \|f - I_{K,\Lambda_n \cup \{\lambda\}}(f)\|_K \qquad \text{for } n \in \mathbb{N}. \tag{7.8}$$

This approach is called the *f/P-greedy algorithm* (cf. [98, Definition 3.1.1 ff.], [120, Theorem 6 ff.]). In the format (7.8), it is not clear how we can solve the minimization problem in each iteration step. Instead, we can rewrite the updated error into simpler terms for each $\lambda \in \Gamma \setminus \operatorname{span}(\Lambda_n)$:

$$\begin{aligned}
\|f - I_{K,\Lambda_n}(f)\|_K^2 &= \|f - I_{K,\Lambda_n \cup \{\lambda\}}(f) + I_{K,\Lambda_n \cup \{\lambda\}}(f) - I_{K,\Lambda_n}(f)\|_K^2 \\
&= \|f - I_{K,\Lambda_n \cup \{\lambda\}}(f)\|_K^2 + \|I_{K,\Lambda_n \cup \{\lambda\}}(f) - I_{K,\Lambda_n}(f)\|_K^2 \\
&= \|f - I_{K,\Lambda_n \cup \{\lambda\}}(f)\|_K^2 + \frac{|\lambda(f) - \lambda(I_{K,\Lambda_n}(f))|^2}{P_{\Lambda_n}(\lambda)^2},
\end{aligned}$$

where we used the update formula from Corollary 6.10. Therefore, minimizing the term

$$\|f - I_{K,\Lambda_n \cup \{\lambda\}}(f)\|_K^2 = \|f - I_{K,\Lambda_n}(f)\|_K^2 - \frac{|\lambda(f) - \lambda(I_{K,\Lambda_n}(f))|^2}{P_{\Lambda_n}(\lambda)^2}$$

is equivalent to maximizing the subtrahend on the right side of the equation, so that the selection rule (7.8) can be rewritten as

$$\frac{|\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda_n}(f))|}{P_{\Lambda_n}(\lambda_{n+1})} = \sup_{\lambda \in \Gamma \setminus \operatorname{span}(\Lambda_n)} \frac{|\lambda(f) - \lambda(I_{K,\Lambda_n}(f))|}{P_{\Lambda_n}(\lambda)} \qquad \text{for } n \in \mathbb{N}. \tag{7.9}$$

We observe that the power function occurs in the denominator this time, so the selection rule favors smaller values of the power function. In terms of numerical stability, this leads to bad performances in practical cases, see e.g. [98, Section 6.5].

Similarly, the f-greedy selection rule neglects the numerical stability of the interpolation process, as it only takes the pointwise error into account. If we wish to have a good tradeoff between approximation quality and numerical stability, we could complement the pointwise error with the respective power function values, which leads to the *psr-greedy algorithm* (cf. [43])

$$|\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda_n}(f))| \cdot P_{\Lambda_n}(\lambda_{n+1}) = \sup_{\lambda \in \Gamma} |\lambda(f) - \lambda(I_{K,\Lambda_n}(f))| \cdot P_{\Lambda_n}(\lambda) \qquad \text{for } n \in \mathbb{N}. \tag{7.10}$$

Here, the abbreviation *psr* stands for *power-scaled residual*. Due to the multiplication, we require both the pointwise error and the power function value to be sufficiently high. The obvious drawback of the target-dependent algorithm is that for each $f \in \mathcal{H}_K$, the sequence of data points has to be computed individually, whereas the target-independent algorithms generalize to all functions from the native space. On the other hand, the target-dependent algorithms are expected to deliver better approximation quality, as the data points are optimized for the given function that we want to approximate.

Remark 7.18. Throughout Chapter 5, we have collected several measurements for the approximation error and numerical stability. Of course, we can combine these in various other ways. For example, we could use the pointwise error and the distance function via the selection rule

$$|\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda_n}(f))| \cdot \operatorname{dist}(\lambda_{n+1}, \Lambda_n) = \sup_{\lambda \in \Gamma} |\lambda(f) - \lambda(I_{K,\Lambda_n}(f))| \cdot \operatorname{dist}(\lambda, \Lambda_n) \qquad \text{for } n \in \mathbb{N}.$$

7.2.3. Generalization: β-greedy algorithms

In the papers [145] and [146], the authors were able to summarize several approaches from the last subsection as one special class of greedy algorithms depending on a parameter $\beta \in [0, \infty)$. Given this fixed parameter, the *β-greedy algorithm* selects the new functional via

$$|\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda_n}(f))|^\beta \cdot P_{\Lambda_n}(\lambda_{n+1})^{1-\beta} = \sup_{\lambda \in \Gamma \setminus \operatorname{span}(\Lambda_n)} |\lambda(f) - \lambda(I_{K,\Lambda_n}(f))|^\beta \cdot P_{\Lambda_n}(\lambda)^{1-\beta} \tag{7.11}$$

for $n \in \mathbb{N}$. It is easy to see that $\beta = 0$ leads to the P-greedy algorithm (7.3) and $\beta = 1$ yields the f-greedy algorithm (7.4). For $\beta = 1/2$, we get the selection rule

$$|\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda_n}(f))|^{1/2} \cdot P_{\Lambda_n}(\lambda_{n+1})^{1/2} = \sup_{\lambda \in \Gamma \setminus \operatorname{span}(\Lambda_n)} |\lambda(f) - \lambda(I_{K,\Lambda_n}(f))|^{1/2} \cdot P_{\Lambda_n}(\lambda)^{1/2}$$

for $n \in \mathbb{N}$, which is equivalent to the psr-greedy algorithm (7.10). We can even consider the limit case $\beta \to \infty$ for the term

$$|\lambda(f) - \lambda(I_{K,\Lambda_n}(f))|^{\beta} \cdot P_{\Lambda_n}(\lambda)^{1-\beta} = \left(\frac{|\lambda(f) - \lambda(I_{K,\Lambda_n}(f))|}{P_{\Lambda_n}(\lambda)} \right)^{\beta} \cdot P_{\Lambda_n}(\lambda).$$

If two functionals $\lambda, \mu \in \Gamma \setminus \operatorname{span}(\Lambda_n)$ satisfy

$$\frac{|\lambda(f) - \lambda(I_{K,\Lambda_n}(f))|}{P_{\Lambda_n}(\lambda)} < \frac{|\mu(f) - \mu(I_{K,\Lambda_n}(f))|}{P_{\Lambda_n}(\mu)},$$

there is $\beta_0 > 0$ such that

$$|\lambda(f) - \lambda(I_{K,\Lambda_n}(f))|^{\beta} \cdot P_{\Lambda_n}(\lambda)^{1-\beta} < |\mu(f) - \mu(I_{K,\Lambda_n}(f))|^{\beta} \cdot P_{\Lambda_n}(\mu)^{1-\beta} \qquad \text{for } \beta > \beta_0.$$

Hence, it makes sense to define the β-greedy algorithm as the f/P-greedy algorithm for $\beta = \infty$. We can prove the convergence of the β-greedy algorithm for every $\beta \in [0, \infty]$.

Theorem 7.19. *Let $\beta \in [0, \infty]$ and $\Gamma \subset \mathcal{H}_K^*$ be totally bounded. If $f \in \mathcal{H}_{K,\Gamma}$ and $(\Lambda_n)_{n \in \mathbb{N}}$ is chosen via the respective β-greedy algorithm, we have the convergence*

$$\|f - I_{K,\Lambda_n}(f)\|_K \xrightarrow{n \to \infty} 0.$$

Proof. According to Lemma 7.9, it is sufficient to show pointwise convergence. Therefore, let $\lambda \in \Gamma$. If $\lambda \in \bigcup_{n \in \mathbb{N}} \operatorname{span}(\Lambda_n)$ holds, there is $n_0 \in \mathbb{N}$ such that $P_{\Lambda_n}(\lambda) = 0$ for $n \geq n_0$. The standard power function estimate (5.2) then gives

$$|\lambda(f) - \lambda(I_{K,\Lambda_n}(f))| \leq P_{\Lambda_n}(\lambda) \cdot \|f\|_K = 0 \qquad \text{for } n \geq n_0.$$

Hence, let us assume that $\lambda \notin \bigcup_{n \in \mathbb{N}} \operatorname{span}(\Lambda_n)$. We distinguish between several cases regarding the parameter β:

- $\underline{\beta = 0}$: In this case, we have

  $$P_{\Lambda_n}(\lambda) \leq P_{\Lambda_n}(\lambda_{n+1}) \xrightarrow{n \to \infty} 0$$

  due to the selection rule and Lemma 7.8. Again, the standard power function estimate gives

  $$|\lambda(f) - \lambda(I_{K,\Lambda_n}(f))| \leq P_{\Lambda_n}(\lambda) \cdot \|f\|_K \xrightarrow{n \to \infty} 0.$$

- $\underline{\beta \in (0,1)}$: Consider the monotonic and bounded sequence

  $$(P_{\Lambda_n}(\lambda))_{n \in \mathbb{N}} \qquad \text{and its limit} \qquad C := \lim_{n \to \infty} P_{\Lambda_n}(\lambda).$$

  If $C = 0$, we already have pointwise convergence. Otherwise, we have $C > 0$ and

  $$\begin{aligned} |\lambda(f) - \lambda(I_{K,\Lambda_n}(f))|^{\beta} &= |\lambda(f) - \lambda(I_{K,\Lambda_n}(f))|^{\beta} \cdot P_{\Lambda_n}(\lambda)^{1-\beta} \cdot P_{\Lambda_n}(\lambda)^{\beta-1} \\ &\leq |\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda_n}(f))|^{\beta} \cdot P_{\Lambda_n}(\lambda_{n+1})^{1-\beta} \cdot P_{\Lambda_n}(\lambda)^{\beta-1} \\ &\leq C^{\beta-1} \cdot \|f\|_K^{\beta} \cdot P_{\Lambda_n}(\lambda_{n+1}) \xrightarrow{n \to \infty} 0. \end{aligned}$$

  This yields

  $$|\lambda(f) - \lambda(I_{K,\Lambda_n}(f))| \xrightarrow{n \to \infty} 0.$$

- $\beta \in [1, \infty)$: Similarly, we can estimate

$$\begin{aligned} |\lambda(f) - \lambda(I_{K,\Lambda_n}(f))|^\beta &\leq |\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda_n}(f))|^\beta \cdot P_{\Lambda_n}(\lambda_{n+1})^{1-\beta} \cdot P_{\Lambda_n}(\lambda)^{\beta-1} \\ &\leq \|\lambda\|_K^{\beta-1} \cdot \|f\|_K^\beta \cdot P_{\Lambda_n}(\lambda_{n+1}) \xrightarrow{n\to\infty} 0. \end{aligned}$$

- $\beta = \infty$: From Remark 6.14, we know that

$$\frac{|\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda_n}(f))|}{P_{\Lambda_n}(\lambda_{n+1})} \xrightarrow{n\to\infty} 0.$$

  As in the case before, we estimate

$$|\lambda(f) - \lambda(I_{K,\Lambda_n}(f))| \leq \frac{|\lambda_{n+1}(f) - \lambda_{n+1}(I_{K,\Lambda_n}(f))|}{P_{\Lambda_n}(\lambda_{n+1})} \cdot \|\lambda\|_K \xrightarrow{n\to\infty} 0.$$

In total, we have the desired pointwise convergence for every $\beta \in [0, \infty]$. □

Remark 7.20. In our general convergence analysis, we have not mentioned any convergence rates for the β-greedy algorithms, as these depend on the set Γ and the regularity of the kernel K. Convergence rates for the standard interpolation method in the case of a *Sobolev kernel* can be found in [145], similar results for linear elliptic differential operators of second order are given in [146].

7.2.4. Vectorial approach

In many cases, we have a set $\{f_1, ..., f_M\} \subset \mathcal{H}_K$ of target functions instead of a single target, which we refer to as a *training set*, and we want to select a sequence of data points that leads to a good approximation quality for all target functions simultaneously. Note that a suitable method was already proposed in [149] that adapts the f/P-greedy algorithm, and we further generalize this approach to derive *weighted vectorial versions* of the β-greedy algorithms for all $\beta \in [0, \infty]$. We remark that this approach can be seen as a special application of interpolation via matrix-valued kernels. For detailed information on matrix-valued kernels, we refer to [150].

First, we need to find a suitable Hilbert space setting for this simultaneous approximation problem. We consider the collection of target functions as a vector

$$f = (f_1, ..., f_M) \in \mathcal{H}_K^M = \mathcal{H}_K \times ... \times \mathcal{H}_K.$$

Given a fixed vector $w = (w_1, ..., w_M)^T \in \mathbb{R}^M$ with $w_i > 0$ for $i = 1, ..., M$ and $\sum_{i=1}^M w_i = 1$, we define an inner product on $\mathcal{H}_K^M$ by

$$\langle f, g \rangle_{K,w} := \sum_{i=1}^M w_i \cdot \langle f_i, g_i \rangle_K \qquad \text{for } f = (f_1, ..., f_M), g = (g_1, ..., g_M) \in \mathcal{H}_K^M.$$

The idea of the weight vector w is to control the influence of the single target functions on the data selection. If a target function from the training set is considered a good representation of the resulting application, its weight should be sufficiently high. We can embed the space $\mathcal{H}_K$ into $\mathcal{H}_K^M$ via the linear mapping

$$\mathfrak{i} : \mathcal{H}_K \to \mathcal{H}_K^M, \ s \mapsto (s, ..., s),$$

which satisfies

$$\langle \mathfrak{i}(s_1), \mathfrak{i}(s_2) \rangle_{K,w} = \sum_{i=1}^M w_i \cdot \langle s_1, s_2 \rangle_K = \langle s_1, s_2 \rangle_K \qquad \text{for all } s_1, s_2 \in \mathcal{H}_K. \tag{7.12}$$

Thus, given a set of functionals $\Lambda_n = \{\lambda_1, ..., \lambda_n\} \subset \mathcal{H}_K^*$, we consider the embedded space $\mathfrak{i}(S_{K,\Lambda_n})$. If $\mathcal{N}$ is the Newton basis of S_{K,Λ_n}, then $\mathfrak{i}(\mathcal{N})$ is again an orthonormal system due to (7.12). Following the interpolation of a single target function, we define the simultaneous approximation to the target functions

f_i as the orthogonal projection $\mathcal{P}_{\mathfrak{i}(S_{K,\Lambda_n})}f$ of $f = (f_1, ..., f_M)$ onto the space $\mathfrak{i}(S_{K,\Lambda_n})$ (cf. Theorem A.9), which satisfies

$$\|f - \mathcal{P}_{\mathfrak{i}(S_{K,\Lambda_n})}f\|_{K,w}^2 = \inf_{s\in\mathfrak{i}(S_{K,\Lambda_n})} \|f - s\|_{K,w}^2 = \inf_{s\in S_{K,\Lambda_n}} \sum_{i=1}^M w_i \cdot \|f_i - s\|_K^2$$

and is given by

$$\begin{aligned}\mathcal{P}_{\mathfrak{i}(S_{K,\Lambda_n})}f = \sum_{j=1}^n \langle f, \mathfrak{i}(\mathfrak{n}_j)\rangle_{K,w} \cdot \mathfrak{i}(\mathfrak{n}_j) &= \sum_{j=1}^n \left(\sum_{i=1}^M w_i \cdot \langle f_i, \mathfrak{n}_j\rangle_K\right) \cdot \mathfrak{i}(\mathfrak{n}_j)\\ &= \sum_{i=1}^M w_i \cdot \mathfrak{i}\left(\sum_{j=1}^n \langle f_i, \mathfrak{n}_j\rangle_K \cdot \mathfrak{n}_j\right)\\ &= \mathfrak{i}\left(\sum_{i=1}^M w_i \cdot I_{K,\Lambda_n}(f_i)\right).\end{aligned}$$

Hence, the simultaneous reconstruction is the weighted sum of the single interpolants to the target functions and solves a weighted least squares problem within this approach. This means that we can use the update formulas from Section 6 for an efficient implementation, and we can parallelize the computation of the single interpolants. But we still have to find suitable point selection algorithms for this case.

Let $\lambda_{n+1} \notin \operatorname{span}(\Lambda_n)$ and set $\Lambda_{n+1} = \Lambda_n \cup \{\lambda_{n+1}\}$. The discussion about the f/P-greedy algorithms from Subsection 7.2.2 can be adapted to our setting, since

$$\begin{aligned}\|f - \mathcal{P}_{\mathfrak{i}(S_{K,\Lambda_{n+1}})}f\|_{K,w}^2 &= \sum_{i=1}^M w_i \cdot \|f_i - I_{K,\Lambda_{n+1}}(f_i)\|_K^2\\ &= \sum_{i=1}^M w_i \cdot \left(\|f_i - I_{K,\Lambda_n}(f_i)\|_K^2 - \frac{|\lambda_{n+1}(f_i) - \lambda_{n+1}(I_{K,\Lambda_n}(f_i))|^2}{P_{\Lambda_n}(\lambda_{n+1})^2}\right)\\ &= \|f - \mathcal{P}_{\mathfrak{i}(S_{K,\Lambda_n})}f\|_{K,w}^2 - \sum_{i=1}^M w_i \cdot \frac{|\lambda_{n+1}(f_i) - \lambda_{n+1}(I_{K,\Lambda_n}(f_i))|^2}{P_{\Lambda_n}(\lambda_{n+1})^2}\end{aligned}$$

holds in the vectorial case (cf. [149, Lemma 2.1]). Given a superset $\Gamma \subset \mathcal{H}_K^*$, we need to choose λ_{n+1} such that

$$\sum_{i=1}^M w_i \cdot \frac{|\lambda_{n+1}(f_i) - \lambda_{n+1}(I_{K,\Lambda_n}(f_i))|^2}{P_{\Lambda_n}(\lambda_{n+1})^2} = \sup_{\lambda\in\Gamma\setminus\operatorname{span}(\Lambda_n)} \sum_{i=1}^M w_i \cdot \frac{|\lambda(f_i) - \lambda(I_{K,\Lambda_n}(f_i))|^2}{P_{\Lambda_n}(\lambda)^2} \tag{7.13}$$

in order to minimize the approximation error in the next iteration, which adapts the f/P-greedy algorithm to the case of multiple target functions (cf. [149, Algorithm 3]). In the same way, we define the *vectorial β-greedy algorithm* for $\beta \in [0, \infty)$ as the selection rule

$$\begin{aligned}&\left(\sum_{i=1}^M w_i \cdot |\lambda_{n+1}(f_i) - \lambda_{n+1}(I_{K,\Lambda_n}(f_i))|^2\right)^\beta \cdot P_{\Lambda_n}(\lambda_{n+1})^{2-2\beta}\\ &\qquad = \sup_{\lambda\in\Gamma\setminus\operatorname{span}(\Lambda_n)} \left(\sum_{i=1}^M w_i \cdot |\lambda(f_i) - \lambda(I_{K,\Lambda_n}(f_i))|^2\right)^\beta \cdot P_{\Lambda_n}(\lambda)^{2-2\beta},\end{aligned} \tag{7.14}$$

and for $\beta = \infty$ as the selection rule (7.13). For totally bounded domains, the vectorial versions converge as well.

Theorem 7.21. *Let $\beta \in [0, \infty]$ and $\Gamma \subset \mathcal{H}_K^*$ be totally bounded. If $f_1, ..., f_M \in \mathcal{H}_{K,\Gamma}$, $w \in \mathbb{R}^M$ with*

$$\sum_{i=1}^M w_i = 1, \qquad w_i > 0 \quad \text{for } i = 1, ..., M$$

and $(\Lambda_n)_{n\in\mathbb{N}}$ *is chosen via the respective vectorial* β*-greedy algorithm (7.13, 7.14), we have the convergence*

$$\|f_i - I_{K,\Lambda_n}(f_i)\|_K \xrightarrow{n\to\infty} 0 \qquad \text{for } i = 1, ..., M.$$

Proof. We can recycle large parts from the proof of Theorem 7.19. Again, it is sufficient to show pointwise convergence on Γ and we only need to check the case

$$\lambda \notin \bigcup_{n\in\mathbb{N}} \operatorname{span}(\Lambda_n).$$

- $\underline{\beta = 0}$: This is the same as the single-target P-greedy algorithm, and we have

$$|\lambda(f_i) - \lambda(I_{K,\Lambda_n}(f_i))| \le P_{\Lambda_n}(\lambda) \cdot \|f_i\|_K \le P_{\Lambda_n}(\lambda_{n+1}) \cdot \|f_i\|_K \xrightarrow{n\to\infty} 0 \qquad \text{for } i = 1, ...M.$$

- $\underline{\beta = \infty}$: With Remark 6.14, we get

$$\begin{aligned}\sum_{i=1}^{M} w_i \cdot |\lambda(f_i) - \lambda(I_{K,\Lambda_n}(f_i))|^2 &= \left(\sum_{i=1}^{M} w_i \cdot \frac{|\lambda(f_i) - \lambda(I_{K,\Lambda_n}(f_i))|^2}{P_{\Lambda_n}(\lambda)^2}\right) \cdot P_{\Lambda_n}(\lambda)^2 \\ &\le \left(\sum_{i=1}^{M} w_i \cdot \frac{|\lambda_{n+1}(f_i) - \lambda_{n+1}(I_{K,\Lambda_n}(f_i))|^2}{P_{\Lambda_n}(\lambda_{n+1})^2}\right) \cdot \|\lambda\|_K^2 \xrightarrow{n\to\infty} 0.\end{aligned}$$

  The desired pointwise convergence then follows from the estimate

$$|\lambda(f_j) - \lambda(I_{K,\Lambda_n}(f_j))| \le w_j^{-1/2} \cdot \left(\sum_{i=1}^{M} w_i \cdot |\lambda(f_i) - \lambda(I_{K,\Lambda_n}(f_i))|^2\right)^{1/2} \qquad \text{for } j = 1, ..., M.$$

- $\underline{\beta \in (0, \infty)}$: We can use the same estimates as in the proof of Theorem 7.19 to derive

$$\left(\sum_{i=1}^{M} w_i \cdot |\lambda(f_i) - \lambda(I_{K,\Lambda_n}(f_i))|^2\right)^{\beta} \xrightarrow{n\to\infty} 0,$$

  which results in pointwise convergence similar to the case before.

□

7.2.5. Data thinning

As we have already pointed out, we mostly deal with finite supersets $\Gamma \subset \mathcal{H}_K^*$ in practical cases, so that the greedy algorithms only perform a finite number of iteration steps anyways. In this scenario, the greedy algorithms can be interpreted as a method to thin out the given superset (cf. [44], [69, Chapter 4]). Here, data thinning means that we want to get rid of redundancies within the data set.

For example, the power function value $P_{\Lambda_n}(\lambda)$ of a functional $\lambda \in \Gamma$ indicates the approximation error between λ and $\operatorname{span}(\Lambda_n)$, and therefore how well λ is already represented in the current data point set Λ_n. This implies that functionals with small power function values represent redundant data, whereas functionals with high power function values contain new information that should be added to the interpolation process. Hence, the P-greedy algorithm can be used to reduce the full superset to a subset $\Lambda_N \subset \Gamma$ with $|\Lambda_N| = N \ll |\Gamma|$ consisting of the most relevant functionals in terms of linear independence.

Similarly, if the pointwise error at the functional $\lambda \in \Gamma \setminus \operatorname{span}(\Lambda_n)$ is already low, we might not want to add this functional to our data point set if a further reduction of the pointwise error is not required. The presented greedy algorithms from this section can be used to remove redundant data in Γ, which can have significant positive effects on the computational complexity and the numerical stability of the interpolation process without losing too much information. This process is visualized in Figure 7.1 for the geometric greedy selection, where the initial scattered point set $X \subset \mathbb{R}^2$ with $|X| = 10^4$ (left) is reduced to a subset $Y \subset X$ with $|Y| = 10^3$ (right) by applying $10^3 - 1$ steps of the geometric greedy selection from (7.6). The selected data points are marked in red.

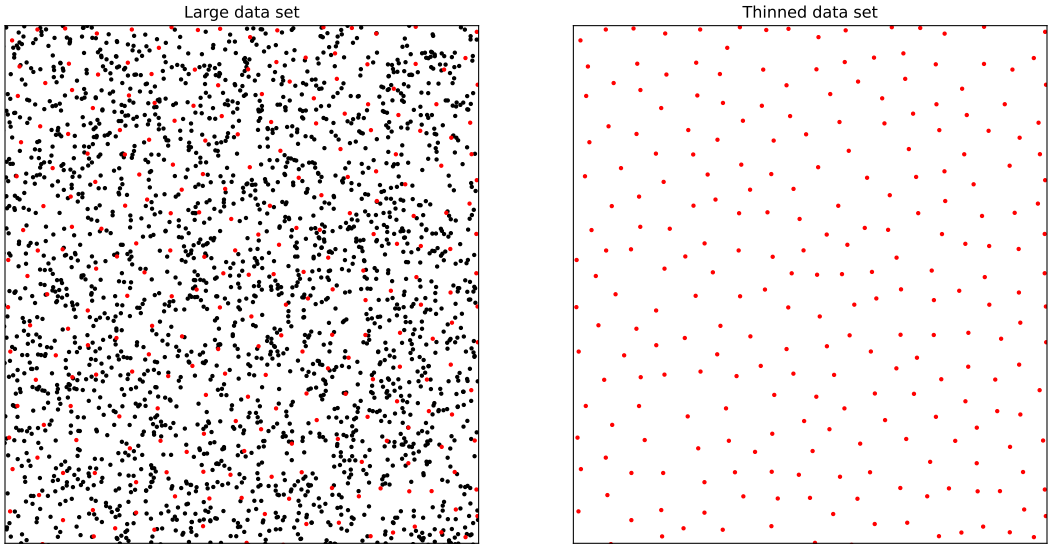

Figure 7.1.: Visualization of data thinning via geometric greedy selection

7.3. Regularization methods

Instead of simply solving the interpolation equations, we can include a regularization method to enforce additional properties on the solution of the approximation problem. Regularization methods have become popular in many applications, as they can help to reduce the effects of *overfitting* when dealing with large data sets and stabilize the solution to *ill-posed problems* (cf. Section 8.3). More details on regularization methods can be found in [45].

The main idea is to substitute the interpolation equations with a modified error functional

$$J(s) := D(f,s) + \gamma \cdot R(s) \qquad \text{for } s \in \mathcal{H}_K, \tag{7.15}$$

where $f \in \mathcal{H}_K$ is the target function, $D : \mathcal{H}_K \times \mathcal{H}_K \to \mathbb{R}$ is a map that measures the data consistency and $R : \mathcal{H}_K \to \mathbb{R}$ is a regularization functional with regularization parameter $\gamma \geq 0$. The approximation to f is then given by the minimizer $s^* \in \mathcal{H}_K$ of the functional, i.e.

$$J(s^*) = \inf_{s \in \mathcal{H}_K} J(s).$$

Popular choices for the data consistency term D are norms on the underlying vector space or least squares functionals. For our interpolation problem, we use the *mean squared error*

$$D(f,s) := \frac{1}{N} \cdot \sum_{i=1}^{N} |\lambda_i(f) - \lambda_i(s)|^2 \qquad \text{for } f, s \in \mathcal{H}_K$$

on a set $\Gamma = \{\lambda_1, ..., \lambda_N\} \subset \mathcal{H}_K^*$. In the special case that R has the form

$$R(s) = \|s\|_K^2 \qquad \text{for } s \in \mathcal{H}_K$$

and $\gamma > 0$, the optimization problem (7.15) can be reduced to the domain $S_{K,\Gamma} \subset \mathcal{H}_K$. This result is a generalized version of the *representer theorem* (see, e.g., [122]).

Theorem 7.22 (Generalized representer theorem)**.** *Let $\Gamma = \{\lambda_1, ..., \lambda_N\} \subset \mathcal{H}_K^*$ and $\gamma > 0$. If $s^* \in \mathcal{H}_K$ satisfies the optimality condition*

$$\frac{1}{N} \cdot \sum_{i=1}^{N} |\lambda_i(f) - \lambda_i(s^*)|^2 + \gamma \cdot \|s^*\|_K^2 = \inf_{s \in \mathcal{H}_K} \frac{1}{N} \cdot \sum_{i=1}^{N} |\lambda_i(f) - \lambda_i(s)|^2 + \gamma \cdot \|s\|_K^2, \tag{7.16}$$

then we must have $s^* \in S_{K,\Gamma}$.

Proof. Let s^* satisfy (7.16). There is a unique decomposition

$$s^* = s_1^* + s_2^*, \qquad \text{where } s_1^* \in S_{K,\Gamma} \text{ and } s_2^* \in S_{K,\Gamma}^{\perp}.$$

Due to Lemma 4.9, we get the equality

$$\lambda_i(s^*) = \lambda_i(s_1^*) + \lambda_i(s_2^*) = \lambda_i(s_1^*) + 0 = \lambda_i(s_1^*) \qquad \text{for } i = 1, ..., N.$$

If $s_2^* \neq 0$, then

$$\|s^*\|_K^2 = \|s_1^*\|_K^2 + \|s_2^*\|_K^2 > \|s_1^*\|_K^2$$

and therefore

$$\frac{1}{N} \cdot \sum_{i=1}^{N} |\lambda_i(f) - \lambda_i(s^*)|^2 + \gamma \cdot \|s^*\|_K^2 > \frac{1}{N} \cdot \sum_{i=1}^{N} |\lambda_i(f) - \lambda_i(s_1^*)|^2 + \gamma \cdot \|s_1^*\|_K^2.$$

But this contradicts (7.16). Hence, we must have $s^* = s_1^* \in S_{K,\Gamma}$. □

The consequence is that we can substitute the initial infinite-dimensional problem (7.16) with a finite-dimensional regularized least-squares problem. Further analysis and numerical examples for the resulting *regularized interpolation problem* can be found in [115].

Here, we want to extend the problem by the idea of data thinning from Section 7.2.5, where we only consider a subdomain $S_{K,\Lambda} \subset S_{K,\Gamma}$ with $\Lambda \subset \Gamma = \{\lambda_1, ..., \lambda_N\}$ and set $n = |\Lambda|$. Moreover, instead of penalizing the native space norm of a function, we regularize with respect to evaluations of additional functionals

$$\Xi = \{\xi_1, ..., \xi_M\} \subset \mathcal{H}_K^*.$$

A possible realization of this approach is the minimization problem

$$\frac{1}{N} \cdot \sum_{i=1}^{N} |\lambda_i(f) - \lambda_i(s^*)|^2 + \gamma \cdot \sum_{i=1}^{M} |\xi_i(s^*)|^2 = \inf_{s \in S_{K,\Lambda}} \frac{1}{N} \cdot \sum_{i=1}^{N} |\lambda_i(f) - \lambda_i(s)|^2 + \gamma \cdot \sum_{i=1}^{M} |\xi_i(s)|^2 \qquad (7.17)$$

for $\gamma \geq 0$. For our analysis, we denote the solution as $s^{(\gamma)} \equiv s^*$ for a given regularization parameter γ. If $\mathcal{N} = \{\mathfrak{n}_1, ..., \mathfrak{n}_n\}$ is the Newton basis of $S_{K,\Lambda}$, we have the identity

$$\|s\|_K^2 = \sum_{i=1}^{n} |\langle s, \mathfrak{n}_i \rangle_K|^2 \qquad \text{for all } s \in S_{K,\Lambda}.$$

This shows that the native space norm regularization is just a special case of (7.17), where

$$\Xi \equiv \Xi_{\mathcal{N}} = \{\langle \cdot, \mathfrak{n}_i \rangle_K \mid i = 1, ..., n\} \subset \mathcal{H}_K^*.$$

If Λ is linearly independent, there is a unique solution.

Theorem 7.23. *Let $\Lambda \subset \Gamma$ be linearly independent, $f \in \mathcal{H}_K$ and $\Xi \subset \mathcal{H}_K^*$. For any parameter $\gamma \geq 0$, there is a unique solution $s^{(\gamma)} \in S_{K,\Lambda}$ to the regularized problem (7.17).*

Proof. Without loss of generality, we assume that $\Lambda = \{\lambda_1, ..., \lambda_n\}$ and write the elements $s \in S_{K,\Lambda}$ in terms of the standard basis, i.e.

$$s = \sum_{i=1}^{n} c_i \cdot \lambda_i^y K(\cdot, y), \qquad \text{with } c = (c_1, ..., c_n)^T \in \mathbb{R}^n.$$

In this case, the energy functional can be rewritten as

$$\frac{1}{N} \cdot \sum_{i=1}^{N} |\lambda_i(f) - \lambda_i(s)|^2 + \gamma \cdot \sum_{i=1}^{M} |\xi_i(s)|^2 = \frac{1}{N} \cdot \|A_{K,\Gamma,\Lambda} \cdot c - f_\Gamma\|_2^2 + \gamma \cdot \|B_\Xi \cdot c\|_2^2, \qquad (7.18)$$

where the matrices are given by

$$A_{K,\Gamma,\Lambda} := (\langle \lambda_i, \lambda_j \rangle_K)_{\substack{1 \le i \le N \\ 1 \le j \le n}} \in \mathbb{R}^{N \times n} \qquad \text{and} \qquad B_\Xi := \big(\xi_i(\lambda_j^y K(\cdot, y))\big)_{\substack{1 \le i \le M \\ 1 \le j \le n}} \in \mathbb{R}^{M \times n}.$$

Note that the right side of (7.18) is a standard *Tikhonov regularization problem* on $\mathbb{R}^n$ (cf. [70, Section 2.2 & Subsection 8.6.1]). It is well-known that $c \in \mathbb{R}^n$ minimizes the right side of (7.18) if and only if it solves the resulting normal equations

$$\left(\frac{1}{N} \cdot A_{K,\Gamma,\Lambda}^T \cdot A_{K,\Gamma,\Lambda} + \gamma \cdot B_\Xi^T \cdot B_\Xi \right) \cdot c = \frac{1}{N} \cdot A_{K,\Gamma,\Lambda}^T \cdot f_\Gamma. \tag{7.19}$$

Since $A_{K,\Lambda}$ is a regular submatrix of $A_{K,\Gamma,\Lambda}$, the rectangular matrix $A_{K,\Gamma,\Lambda}$ must have full rank so that the system matrix of (7.19) is positive definite for any $\gamma \ge 0$. Consequently, there is exactly one solution $c^{(\gamma)} \in \mathbb{R}^n$, and

$$s^{(\gamma)} = \sum_{i=1}^n c_i^{(\gamma)} \cdot \lambda_i^y K(\cdot, y)$$

is the unique solution to (7.17) due to the identity (7.18). □

For theoretical purposes, we provide an alternative characterization of the minimizer, which is a generalization of [70, Theorem 8.54]. To this end, let us recall the notation

$$s_\Gamma := (\lambda_1(s), ..., \lambda_N(s))^T \in \mathbb{R}^N \qquad \text{and} \qquad s_\Xi := (\xi_1(s), ..., \xi_M(s))^T \in \mathbb{R}^M \qquad \text{for all } s \in S_{K,\Lambda}.$$

Lemma 7.24. *Under the assumptions of Theorem 7.23, the element $s^{(\gamma)} \in S_{K,\Lambda}$ solves the minimization problem (7.17) if and only if it satisfies*

$$\frac{1}{N} \cdot \left\langle f_\Gamma - s_\Gamma^{(\gamma)}, s_\Gamma \right\rangle_2 = \gamma \cdot \left\langle s_\Xi^{(\gamma)}, s_\Xi \right\rangle_2 \qquad \textit{for all } s \in S_{K,\Lambda}. \tag{7.20}$$

Proof. Since all expressions within (7.20) are linear in the argument $s \in S_{K,\Lambda}$, the property is equivalent to the system of equations

$$\frac{1}{N} \cdot \left\langle f_\Gamma - s_\Gamma^{(\gamma)},\ \lambda_j^y K(\cdot, y)_\Gamma \right\rangle_2 = \gamma \cdot \left\langle s_\Xi^{(\gamma)},\ \lambda_j^y K(\cdot, y)_\Xi \right\rangle_2 \qquad \text{for } j = 1, ..., n. \tag{7.21}$$

We write the element $s^{(\gamma)}$ as

$$s^{(\gamma)} = \sum_{i=1}^n c_i^{(\gamma)} \cdot \lambda_i^y K(\cdot, y)$$

and compute

$$\frac{1}{N} \cdot \left\langle f_\Gamma - s_\Gamma^{(\gamma)},\ \lambda_j^y K(\cdot, y)_\Gamma \right\rangle_2 = {e^{(j)}}^T \cdot \left(\frac{1}{N} \cdot A_{K,\Gamma,\Lambda}^T \cdot f_\Gamma - \frac{1}{N} \cdot A_{K,\Gamma,\Lambda}^T \cdot A_{K,\Gamma,\Lambda} \cdot c^{(\gamma)} \right)$$

and

$$\gamma \cdot \left\langle s_\Xi^{(\gamma)},\ \lambda_j^y K(\cdot, y)_\Xi \right\rangle_2 = {e^{(j)}}^T \cdot \left(\gamma \cdot B_\Xi^T \cdot B_\Xi \cdot c^{(\gamma)} \right)$$

for $j = 1, ..., n$, where $e^{(j)} \in \mathbb{R}^n$ is again the j-th standard basis vector. Hence, the system (7.21) is equivalent to the normal equations (7.19), which proves the assertion. □

Remark 7.25. Alternatively, we can prove Lemma 7.24 by considering the *Gateaux derivative* of

$$J_\gamma(s) := \frac{1}{N} \cdot \|f_\Gamma - s_\Gamma\|_2^2 + \gamma \cdot \|s_\Xi\|_2^2 \qquad \text{for } s \in S_{K,\Lambda}, \gamma \ge 0, \tag{7.22}$$

which is given by

$$D_v J_\gamma(s) = \lim_{t\to 0} \frac{J_\gamma(s+t\cdot v) - J_\gamma(s)}{t} = 2\cdot\left(\gamma\cdot\langle s_\Xi, v_\Xi\rangle_2 - \frac{1}{N}\cdot\langle f_\Gamma - s_\Gamma, v_\Gamma\rangle_2\right) \qquad \text{for all } v \in S_{K,\Gamma}\setminus\{0\}$$

and $s \in S_{K,\Lambda}$. It is well-known from convex analysis that $s^{(\gamma)} \in S_{K,\Lambda}$ minimizes the convex functional J_γ if and only if the respective Gateaux derivative vanishes for all directions, i.e.

$$0 = D_s J_\gamma(s^{(\gamma)}) = 2\cdot\left(\gamma\cdot\left\langle s_\Xi^{(\gamma)}, s_\Xi\right\rangle_2 - \frac{1}{N}\cdot\left\langle f_\Gamma - s_\Gamma^{(\gamma)}, s_\Gamma\right\rangle_2\right) \qquad \text{for all } s \in S_{K,\Gamma}\setminus\{0\}.$$

The equation is also satisfied for $s = 0$.

It should be mentioned here that finding an optimal regularization parameter γ is a highly non-trivial task. We shift the discussion about the parameter choice to Chapter 11 and just state some basic results in this section, which generalizes [70, Subsection 8.6.2].

Theorem 7.26. *In the setting of Theorem 7.23, the following statements hold:*

(1) For any $\gamma \geq 0$, there is a constant $C_\gamma > 0$ independent of f such that

$$\frac{1}{N}\cdot\left\|f_\Gamma - s_\Gamma^{(\gamma)}\right\|_2^2 + \gamma\cdot\left\|s_\Xi^{(\gamma)}\right\|_2^2 \leq C_\gamma\cdot\|f\|_K^2.$$

(2) If $\gamma_1 \leq \gamma_2$, we have

$$\left\|s_\Xi^{(\gamma_2)}\right\|_2^2 \leq \left\|s_\Xi^{(\gamma_1)}\right\|_2^2.$$

(3) There is $C > 0$ such that

$$\left\|s^{(\gamma_1)} - s^{(\gamma_2)}\right\|_K^2 \leq C\cdot|\gamma_1 - \gamma_2| \qquad \textit{for all } \gamma_1, \gamma_2 \geq 0.$$

Proof. For the proof, we use the notation (7.22).

(1) We simply compare with the interpolant $I_{K,\Lambda}(f)$, i.e.

$$\begin{aligned} J_\gamma(s^{(\gamma)}) \leq J_\gamma(I_{K,\Lambda}(f)) &= \frac{1}{N}\sum_{i=1}^{N}|\lambda_i(f) - \lambda_i(I_{K,\Lambda}(f))|^2 + \gamma\cdot\sum_{i=1}^{M}|\xi_i(I_{K,\Lambda}(f))|^2 \\ &\leq \frac{1}{N}\cdot\sum_{i=n+1}^{N} P_\Lambda(\lambda_i)^2\cdot\|f\|_K^2 + \gamma\cdot\sum_{i=1}^{M}\|\xi_i\|_K^2\cdot\|I_{K,\Lambda}(f)\|_K^2 \\ &\leq \left(\frac{1}{N}\cdot\sum_{i=n+1}^{N} P_\Lambda(\lambda_i)^2 + \gamma\cdot\sum_{i=1}^{M}\|\xi_i\|_K^2\right)\cdot\|f\|_K^2. \end{aligned}$$

(2) Let $\gamma_1 < \gamma_2$ and assume that $\left\|s_\Xi^{(\gamma_2)}\right\|_2^2 > \left\|s_\Xi^{(\gamma_1)}\right\|_2^2$. This results in

$$J_{\gamma_2}(s^{(\gamma_1)}) = J_{\gamma_1}(s^{(\gamma_1)}) + (\gamma_2 - \gamma_1)\cdot\left\|s_\Xi^{(\gamma_1)}\right\|_2^2 < J_{\gamma_1}(s^{(\gamma_2)}) + (\gamma_2 - \gamma_1)\cdot\left\|s_\Xi^{(\gamma_2)}\right\|_2^2 = J_{\gamma_2}(s^{(\gamma_2)}),$$

which is a contradiction to the assumption that $s^{(\gamma_2)}$ minimizes J_{γ_2}.

(3) Let $\gamma_1 \leq \gamma_2$. With Lemma 7.24, we have

$$\frac{1}{N}\cdot\left\langle f_\Gamma - s_\Gamma^{(\gamma_1)}, s_\Gamma\right\rangle_2 = \gamma_1\cdot\left\langle s_\Xi^{(\gamma_1)}, s_\Xi\right\rangle_2$$

and

$$\frac{1}{N} \cdot \left\langle f_\Gamma - s_\Gamma^{(\gamma_2)}, s_\Gamma \right\rangle_2 = \gamma_2 \cdot \left\langle s_\Xi^{(\gamma_2)}, s_\Xi \right\rangle_2 \quad \text{for all } s \in S_{K,\Lambda}.$$

Subtracting the two equations from one another leads to

$$\frac{1}{N} \cdot \left\langle s_\Gamma^{(\gamma_1)} - s_\Gamma^{(\gamma_2)}, s_\Gamma \right\rangle_2 = \gamma_2 \cdot \left\langle s_\Xi^{(\gamma_2)}, s_\Xi \right\rangle_2 - \gamma_1 \cdot \left\langle s_\Xi^{(\gamma_1)}, s_\Xi \right\rangle_2 \quad \text{for all } s \in S_{K,\Lambda}.$$

For $s = s^{(\gamma_1)} - s^{(\gamma_2)} \in S_{K,\Lambda}$, we get

$$\begin{aligned}
\frac{1}{N} \cdot \left\| s_\Gamma^{(\gamma_1)} - s_\Gamma^{(\gamma_2)} \right\|_2^2 &= (\gamma_2 + \gamma_1) \cdot \left\langle s_\Xi^{(\gamma_2)}, s_\Xi^{(\gamma_1)} \right\rangle_2 - \gamma_2 \cdot \left\| s_\Xi^{(\gamma_2)} \right\|_2^2 - \gamma_1 \cdot \left\| s_\Xi^{(\gamma_1)} \right\|_2^2 \\
&\leq (\gamma_2 + \gamma_1) \cdot \left\| s_\Xi^{(\gamma_2)} \right\|_2 \cdot \left\| s_\Xi^{(\gamma_1)} \right\|_2 - \gamma_2 \cdot \left\| s_\Xi^{(\gamma_2)} \right\|_2^2 - \gamma_1 \cdot \left\| s_\Xi^{(\gamma_1)} \right\|_2^2 \\
&= \left(\gamma_2 \cdot \left\| s_\Xi^{(\gamma_2)} \right\|_2 - \gamma_1 \cdot \left\| s_\Xi^{(\gamma_1)} \right\|_2 \right) \cdot \left(\left\| s_\Xi^{(\gamma_1)} \right\|_2 - \left\| s_\Xi^{(\gamma_2)} \right\|_2 \right) \\
&\leq (\gamma_2 - \gamma_1) \cdot \left\| s_\Xi^{(\gamma_1)} \right\|_2 \cdot \left(\left\| s_\Xi^{(\gamma_1)} \right\|_2 - \left\| s_\Xi^{(\gamma_2)} \right\|_2 \right) \\
&\leq (\gamma_2 - \gamma_1) \cdot \left\| s_\Xi^{(0)} \right\|_2^2,
\end{aligned}$$

where we used

$$\left\| s_\Xi^{(\gamma_2)} \right\|_2 \leq \left\| s_\Xi^{(\gamma_1)} \right\|_2 \leq \left\| s_\Xi^{(0)} \right\|_2$$

from part (2) and Cauchy's inequality. Moreover, the functional

$$\| \cdot \|_\Gamma : S_{K,\Lambda} \to \mathbb{R}, \ s \mapsto \|s_\Gamma\|_2$$

is a norm on $S_{K,\Lambda}$, as $\Lambda \subset \Gamma$ holds and generalized interpolation problems with respect to Λ are uniquely solveable. Since all norms are equivalent on the finite-dimensional space $S_{K,\Lambda}$, there is a constant $c > 0$ such that

$$\|s\|_K \leq c \cdot \|s\|_\Gamma \qquad \text{for all } s \in S_{K,\Lambda}.$$

In combination with our previous estimate, this gives the desired result

$$\left\| s^{(\gamma_1)} - s^{(\gamma_2)} \right\|_K^2 \leq N c^2 \cdot \left\| s_\Xi^{(0)} \right\|_2^2 \cdot (\gamma_2 - \gamma_1).$$

□

The previous theorem provides important insights into this variational approach:

(1) For fixed $\gamma \geq 0$, the evaluation at the minimizer, and therefore the regularized interpolation error is bounded by

$$J_\gamma(s^{(\gamma)}) \leq C_\gamma \cdot \|f\|_K^2 \qquad \text{for all } f \in \mathcal{H}_K,$$

where C_γ does not depend on f. Hence, the approximation method is stable with respect to the choice of the target function.

(2) If we increase the regularization parameter, the standard norm of the evaluation with respect to the functionals Ξ decreases so that we can enforce additional properties on the solution by choosing appropriate regularizing functionals.

(3) The mapping $\gamma \mapsto s^{(\gamma)}$ is Hölder continuous on $[0, \infty)$ with exponent $1/2$ for fixed $f \in \mathcal{H}_K$. Note that the proof also shows the Hölder continuity of the evaluation mapping

$$\gamma \mapsto s_\Gamma^{(\gamma)}$$

with respect to Γ. In [70, Theorem 8.58], it is shown that even

$$\left\| s_\Gamma^{(\gamma)} - s_\Gamma^{(0)} \right\|_2^2 = o(\gamma) \qquad \text{for } \gamma \to 0$$

holds in the case of norm regularization, where o expresses the order of convergence in terms of the *Landau notation*. The generalization of this convergence result to our case is straightforward.

Pseudo-code

In order to close the first part of this thesis, we give a summary of the discussed methods in the form of a pseudo-code, see Algorithm 1, for treating the generalized interpolation problem with kernel functions. This includes the Newton basis computation from Chapter 6 and the data selection algorithms from this chapter. As we see in the second part, problems based on linear operators can be interpreted as generalized interpolation problems and therefore can be treated with the derived methods.

Remark 7.27. Currently, most of the presented greedy algorithms are only available as single update algorithms, as they only choose one new functional. It is still an open problem to find multi-update versions to improve the numerical performance. For example, regarding the P-greedy algorithm, it would not be sufficient to simply choose several new functionals whose power function values are high, since no information about the relation between the newly chosen functionals is available. This approach could lead to severe numerical problems if we choose functionals that are nearly linearly dependent. In [6, Section 6], a componentwise version of the P-greedy algorithm was proposed for product kernels and grid-like data point sets, resulting in a multi-update. However, this approach is restricted to a special setting and cannot be generalized.

Additionally, a multi-selection of new functionals would require a multi-update of the Newton basis to take full advantage of the increased efficiency. In [4, Subsection 8.2.3], it has already been shown that a multi-update of the Newton basis needs an efficient Cholesky decomposition of the resulting Schur complement, which is not available to the best of our knowledge.

Algorithm 1: Approximation scheme for generalized interpolation problems

Data: Superset $\Gamma \subset \mathcal{H}_K^*$, target values f_Γ of the target function $f \in \mathcal{H}_K$, selection rule η for greedy selection, max. number of iterations $N_{\max}$, tolerance for numerical linear dependence ε, evaluation set $X \subset \mathbb{R}^d$

Choose $\lambda_1 \in \Gamma$ and set $\Lambda_1 = \{\lambda_1\}$;

Initialize evaluations at Newton basis elements via

$$\lambda(\mathfrak{n}_1) = \frac{\langle \lambda, \lambda_1 \rangle_K}{\|\lambda_1\|_K} \qquad \mathfrak{n}_1(x) = \frac{\lambda_1^y K(x,y)}{\|\lambda_1\|_K}$$

for all $\lambda \in \Gamma$, $x \in X$;

Initialize power function values via $P_{\Lambda_1}^2(\lambda) = \langle \lambda, \lambda \rangle_K - \lambda(\mathfrak{n}_1)^2$ for all $\lambda \in \Gamma$;

Initialize residual and evaluation of interpolant via

$$\lambda(f) - \lambda(I_{K,\Lambda_1}(f)) = \lambda(f) - \frac{\lambda_1(f)}{\|\lambda_1\|} \cdot \lambda(\mathfrak{n}_1) \qquad I_{K,\Lambda_1}[f](x) = \frac{\lambda_1(f)}{\|\lambda_1\|} \cdot \mathfrak{n}_1(x)$$

for all $\lambda \in \Gamma$, $x \in X$;

for $i = 1, ..., N_{\max} - 1$ **do**

  Choose λ_{i+1} such that $\eta_{f,\Lambda_i}(\lambda_{i+1}) = \sup_{\lambda \in \Gamma} \eta_{f,\Lambda_i}(\lambda)$;

  if $P_{\Lambda_i}(\lambda_{i+1}) < \varepsilon$ **then**

    break;

  else

    Set $\Lambda_{i+1} = \Lambda_i \cup \{\lambda_{i+1}\}$;

  end

  Update Newton basis evaluation at Γ via

$$\lambda(\mathfrak{n}_{i+1}) = P_{\Lambda_i}(\lambda_{i+1})^{-1} \cdot \left(\langle \lambda, \lambda_{i+1} \rangle_K - \sum_{j=1}^{i} \lambda(\mathfrak{n}_j) \cdot \lambda_{i+1}(\mathfrak{n}_j) \right)$$

  for all $\lambda \in \Gamma$;

  Update residual via

$$\lambda(f) - \lambda(I_{K,\Lambda_{i+1}}(f)) = \lambda(f) - \lambda(I_{K,\Lambda_i}(f)) - \frac{\lambda_{i+1}(f) - \lambda_{i+1}(I_{K,\Lambda_i}(f))}{P_{\Lambda_i}(\lambda_{i+1})} \cdot \lambda(\mathfrak{n}_{i+1})$$

  for all $\lambda \in \Gamma$;

  Update Newton basis evaluation at X via

$$\mathfrak{n}_{i+1}(x) = P_{\Lambda_i}(\lambda_{i+1})^{-1} \cdot \left(\lambda_{i+1}^y K(x,y) - \sum_{j=1}^{i} \mathfrak{n}_j(x) \cdot \lambda_{i+1}(\mathfrak{n}_j) \right)$$

  for all $x \in X$;

  Update evaluation of interpolant via

$$I_{K,\Lambda_{i+1}}[f](x) = I_{K,\Lambda_i}[f](x) + \frac{\lambda_{i+1}(f) - \lambda_{i+1}(I_{K,\Lambda_i}(f))}{P_{\Lambda_i}(\lambda_{i+1})} \cdot \mathfrak{n}_{i+1}(x)$$

  for all $x \in X$;

  Update power function via

$$P_{\Lambda_{i+1}}(\lambda)^2 = P_{\Lambda_i}(\lambda)^2 - \lambda(\mathfrak{n}_{i+1})^2$$

  for all $\lambda \in \Gamma$;

end

Result: Approximate reconstruction $I_{K,\Lambda_{N_{\max}}}(f)$ evaluated at X

Part II.

Application to Computerized Tomography

In the second part, we focus on the application of kernel-based generalized interpolation to *computerized (axial) tomography*. Over the last 50 years, computerized tomography has become one of the standard methods for non-invasive diagnosis with applications in medical imaging, material testing and others. Its history dates back to the discovery of *X-radiation (X-rays)* by Wilhelm C. Röntgen in 1895, which earned him the first Nobel Prize in Physics in 1901. X-radiation is a high-energetic form of radiation that can be used to image the interior of an object. While passing the considered object, the intensity levels of X-rays decrease in dependence on the underlying material, e.g. bones in a human body. Due to the work of August Beer (cf. [19]), this loss of intensity is given by a line integral, where we search for the integrand that describes the structure of the scanned object. As the loss of intensity can be measured in practice, the inverse problem of computerized tomography consists of reconstructing an object from its line integral values.

This mathematical problem was solved by Johann Radon, who published the first theoretical reconstruction formula in his seminal paper [109] from 1917. Nearly 50 years later, Allan M. Cormack „rediscovered“ (cf. [52, page xiii, line 19]) the ideas of Radon in the 1960s and developed reconstruction algorithms based on Fourier transforms (cf. [33], [34]). Independent of Cormack's work, Godfrey N. Hounsfield designed the first operational and commercial scanner in the late 1960s and early 1970s (cf. [66]). Due to their groundbreaking work, Cormack and Hounsfield jointly received the Nobel Prize for Medicine and Physiology in 1979. Although the reconstruction problem has been solved in theory for more than a century, there remain some practical limitations:

- **Finite data samples:** The inversion formula of Radon assumes that the line integral values for all possible lines in the plane are available. Of course, this assumption is not satisfied in practical cases, since we can only perform a finite number of measurements. This means that we need to derive suitable discretizations for the involved operators to obtain an approximate reconstruction from finite data samples.

- **Noise:** In real-world applications, the measured data usually contains additional noise and therefore does not represent the exact line integral values of the object. Hence, (approximate) reconstruction methods are required that are not oversensitive to this noise, i.e. small changes in the measured data should only result in small changes in the reconstruction image. This requirement is associated with the notion of *ill-posed inverse problems* (cf. Section 8.3).

- **Limited view:** In certain applications, we can only measure the integral values for a restricted subset of all possible lines in the plane due to limitations in the data collection process. For example, the directions of the lines in the plane can be limited to certain angles, or the distances from the lines to the origin might be restricted (cf. [108]). These settings require particular care since the methods and results from the non-restricted case might not be applicable anymore (see, e.g., [102, Section 6.2]).

- **Limiting X-ray dosage:** Nowadays, it is well-known that a high dosage of X-rays can severely harm the human body, e.g. by causing cancer. To reduce the health risks associated with computerized tomography, one is usually interested in keeping the radiation dosage as low as possible, in particular when scanning very sensitive parts of the body. This requires reconstruction methods

that are able to obtain a reasonable reconstruction from a small number of measurements. For further reading, we refer to [29, Chapter 11].

The limitations of the reconstruction formula and the introduction of commercial scanners have led to a wide research field regarding reconstruction algorithms for computerized tomography. Recently, the reconstruction via kernel-based generalized interpolation was proposed by Stefano De Marchi, Armin Iske and Gabriele Santin in [38], which is based on the initial results from the master thesis [126] of Amos Sironi. The main goal of the second part is to explain and further elaborate the proposed kernel-based reconstruction method. To this end, we explain the mathematics of computerized tomography and show that the discretization of the reconstruction problem fits into the framework of the first part for suitable kernel functions, which allows us to apply all the derived results and tools. Furthermore, we want to make comparisons with other reconstruction methods to evaluate the utility of the kernel-based reconstruction method.

8. Computerized Tomography

In this chapter, we introduce the mathematics of computerized tomography, which is the foundation of the following chapters. We start by formulating the underlying mathematical problem, which leads to the definition of the main mathematical operator behind computerized tomography in Section 8.1, the *Radon transform*. While discussing relevant properties, we provide several examples that are used in our numerical tests (cf. Chapter 11). The explicit reconstruction formula, known as the *filtered back projection formula*, is explained in Section 8.2. Besides its derivation, we highlight the numerical challenges it brings along in practical implementations. Before we explain suitable reconstruction methods in Chapter 9, we want to briefly discuss the notion of *ill-posed inverse problems* in Section 8.3, where we list two examples that demonstrate the ill-posedness of the reconstruction problem. For further reading on the fundamentals of computerized tomography, we refer to [18], [52] and [101], which have mainly inspired this introduction.

To introduce the computerized tomography problem, consider an object in the plane $\mathbb{R}^2$, whose photon absorption, or less precisely, whose density is described by its unknown attenuation (coefficient) function $f : \mathbb{R}^2 \to \mathbb{R}$. Moreover, we consider a finite line

$$[x_S, x_E] := \left\{ x_s := x_S + s \cdot \frac{x_E - x_S}{\|x_E - x_S\|_2} \;\middle|\; 0 \leq s \leq \|x_E - x_S\|_2 =: L \right\} \subset \mathbb{R}^2$$

in the plane, where $x_S, x_E \in \mathbb{R}^2$ denote the start and end point. The line represents the path of an X-ray passing through the object, and the behavior of its intensity values along this path is denoted by the function $I : [x_S, x_E] \to \mathbb{R}$. In this setting, *Beer's law* states that the change of intensity is proportional to the attenuation coefficient (cf. [19], [52, Section 1.2]).

Assumption 8.1 (Beer's law)**.** For our analysis of the computerized tomography problem, we assume that the X-ray beams are monochromatic, non-refractive and have zero width. In this case, `Beer's law` states that the decay of the intensity while passing the object is given by

$$\frac{\partial I(x_s)}{\partial s} = -f(x_s) \cdot I(x_s) \qquad \text{for all } s \in [0, L]\,. \tag{8.1}$$

Further details on the physics of computerized tomography can be found in [29, Chapter 2].

Usually, we cannot measure the intensity at each point on the line $[x_S, x_E]$, but only at the start and end point, yielding the intensity values $I(x_S)$ and $I(x_E)$. With Beer's law (Assumption 8.1), we get the identity

$$\int\limits_{[x_S, x_E]} f(x)\, dx = \int\limits_{[0,L]} f(x_s)\, ds = \int\limits_{[0,L]} -I(x_s)^{-1} \cdot \frac{\partial I(x_s)}{\partial s}\, ds = -\left[\ln(I(x_s))\right]_0^L = \ln\left(\frac{I(x_S)}{I(x_E)}\right)$$

for the respective line integral of the attenuation function. Hence, we can compute the line integral of the unknown function f over $[x_S, x_E]$ from the total loss of intensity along the line. Note that the start point and the end point of the line should be chosen such that they lie outside the object in order to make these measurements possible. In this case, we could also write the line integral as

$$\int\limits_{[0,L]} f(x_s)\, ds = \int\limits_{\mathbb{R}} f(x_s)\, ds \tag{8.2}$$

if the infinite version of the line does not hit the object somewhere outside $[x_S, x_E]$ and therefore no more attenuation of the X-ray beam takes place. The goal of computerized tomography is to reconstruct the function f from all possible line integral values of the form (8.2).

8.1. Radon transform

Before we continue our analysis, we want to find a suitable, non-redundant parametrization of the set of all infinite lines (affine geometry) in the plane, denoted by

$$AG(\mathbb{R}^2) := \Big\{ a + \mathbb{R} \cdot b \;\Big|\; a \in \mathbb{R}^2, b \in \mathbb{R}^2 \setminus \{0\} \Big\},$$

where

$$a + \mathbb{R} \cdot b := \{a + s \cdot b \mid s \in \mathbb{R}\} \qquad \text{for } a \in \mathbb{R}^2, b \in \mathbb{R}^2 \setminus \{0\}.$$

To this end, we use the orthogonal polar coordinate vectors

$$n_\theta := \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \quad \text{and} \quad v_\theta := \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix} \qquad \text{for } \theta \in [0, \pi)$$

to describe the lines in the plane. Note that the directional vectors v_θ cover all possible directions up to multiplication by -1, and n_θ then represents the shortest path between the line and the zero vector, see also [52, Section 1.3].

Proposition 8.2. *The following mapping between the parameter space $\mathbb{R} \times [0, \pi)$ and $AG(\mathbb{R}^2)$ is bijective:*

$$\varrho : \mathbb{R} \times [0, \pi) \to AG(\mathbb{R}^2),\ (r, \theta) \mapsto \ell_{r,\theta} := r \cdot n_\theta + \mathbb{R} \cdot v_\theta$$

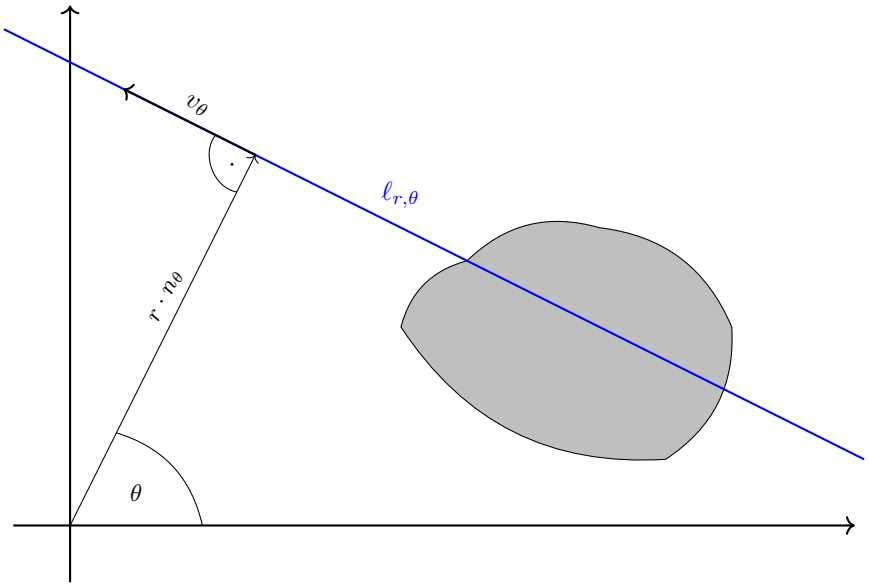

Figure 8.1.: X-ray beam $\ell_{r,\theta}$ passing through an object described by its *attenuation function* $f : \mathbb{R}^2 \to \mathbb{R}$

The previous proposition delivers a suitable parametrization of all lines in the plane, which is visualized in Figure 8.1. With this, we define the main operator behind the problem of computerized tomography, the *Radon tranform* (cf. [52, Definition 2.1]).

Definition 8.3 (Radon transform)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function. The `Radon transform` of f is then given by the collection of line integrals

$$\mathcal{R}f(r, \theta) := \int\limits_{\ell_{r,\theta}} f(x)\, dx = \int\limits_{\mathbb{R}} f(r \cdot n_\theta + s \cdot v_\theta)\, ds \qquad \text{for } (r, \theta) \in \mathbb{R} \times [0, \pi)\,.$$

At this point, it is not yet clear for which class of functions the Radon transform is well-defined. One possible choice for the domain is the space $L^1(\mathbb{R}^2)$. Given a function $f \in L^1(\mathbb{R}^2)$ and an angle $\theta \in [0, \pi)$, we consider the rotation

$$T_\theta : \mathbb{R}^2 \to \mathbb{R}^2,\ \begin{pmatrix} r \\ s \end{pmatrix} \mapsto \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \cdot \begin{pmatrix} r \\ s \end{pmatrix} = r \cdot n_\theta + s \cdot v_\theta. \tag{8.3}$$

Using substitution, we get

$$\int_{\mathbb{R}^2} |f(r \cdot n_\theta + s \cdot v_\theta)|\, d(r,s) = \int_{\mathbb{R}^2} |f(x)|\, dx = \|f\|_{L^1(\mathbb{R}^2)} < \infty, \tag{8.4}$$

which implies that $\mathcal{R}f(\cdot, \theta) \in L^1(\mathbb{R})$ with

$$\|\mathcal{R}f(\cdot,\theta)\|_{L^1(\mathbb{R})} = \int_{\mathbb{R}} |\mathcal{R}f(r,\theta)|\, dr \le \int_{\mathbb{R}} \int_{\mathbb{R}} |f(r \cdot n_\theta + s \cdot v_\theta)|\, ds\, dr = \|f\|_{L^1(\mathbb{R}^2)} \tag{8.5}$$

due to the *Fubini-Tonelli theorem*. In particular, $\mathcal{R}f(r,\theta)$ exists and is finite for almost every $r \in \mathbb{R}$. Combining all previous rotations via the continuous, and therefore measurable mapping

$$T : \mathbb{R}^2 \times [0,\pi)\,,\ (s,r,\theta) \mapsto T_\theta(r,s),$$

the equality (8.4) leads to

$$\int_{[0,\pi)} \int_{\mathbb{R}^2} |f(T(r,s,\theta))|\, d(r,s)\ d\theta = \int_{[0,\pi)} \int_{\mathbb{R}^2} |f(T_\theta(r,s))|\, d(r,s)\ d\theta = \pi \cdot \|f\|_{L^1(\mathbb{R}^2)} < \infty.$$

Another application of the Fubini-Tonelli theorem finally shows that $\mathcal{R}f \in L^1(\mathbb{R} \times [0,\pi))$ holds with

$$\|\mathcal{R}f\|_{L^1(\mathbb{R}\times[0,\pi))} \le \pi \cdot \|f\|_{L^1(\mathbb{R}^2)}. \tag{8.6}$$

Note that inequality (8.6) in combination with the linearity of the operator also shows that $\mathcal{R}f = \mathcal{R}g$ holds in $L^1(\mathbb{R} \times [0,\pi))$ if $f = g$ holds in $L^1(\mathbb{R}^2)$. In total, we get the following result (cf. [18, Proposition 2.2.2]).

Proposition 8.4. *The Radon transform $\mathcal{R} : L^1(\mathbb{R}^2) \to L^1(\mathbb{R} \times [0,\pi))$ is a well-defined, bounded linear operator.*

Example 8.5. We give some examples of Radon transforms, which are important for our numerical tests in Chapter 11. These and more examples can be found in [18, Section 5.2] and [52, Chapter 2].

(i) For given shape parameter $\nu > 0$, consider the Gaussian function

$$G_\nu(x) := e^{-\nu \cdot \|x\|_2^2} \qquad \text{for } x \in \mathbb{R}^2.$$

The Radon value at $(r,\theta) \in \mathbb{R} \times [0,\pi)$ is then given by

$$\mathcal{R}G_\nu(r,\theta) = \int_{\mathbb{R}} G_\nu(r \cdot n_\theta + s \cdot v_\theta)\, ds = \int_{\mathbb{R}} e^{-\nu \cdot (r^2+s^2)}\, ds = e^{-\nu \cdot r^2} \cdot \int_{\mathbb{R}} e^{-\nu \cdot s^2}\, ds = \sqrt{\frac{\pi}{\nu}} \cdot e^{-\nu \cdot r^2},$$

where we used that n_θ, v_θ are orthonormal. Hence, the Radon transform of a Gaussian is again a Gaussian, but only with respect to the radial variable. In general, the Radon transform of a radial symmetric function only depends on the radial variable (see, e.g., [18, Proposition 2.2.8]).

(ii) For a shifted and rotated elliptical region (Figure 8.2)

$$E_{R_1,R_2,y,\tau} = \Big\{ x \in \mathbb{R}^2 \;\Big|\; \frac{((x_1-y_1)\cdot\cos(\tau) + (x_2-y_2)\cdot\sin(\tau))^2}{R_1^2} + \frac{(-(x_1-y_1)\cdot\sin(\tau) + (x_2-y_2)\cdot\cos(\tau))^2}{R_2^2} \le 1 \Big\},$$

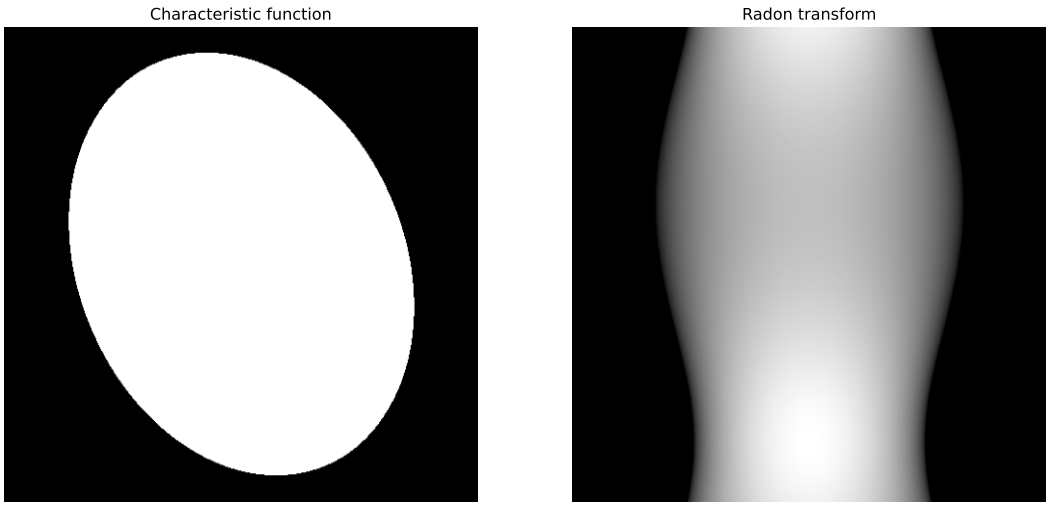

Figure 8.2.: Characteristic function of a rotated elliptical region (left) and its Radon transform (right)

where $R_1, R_2 > 0, y = (y_1, y_2)^T \in \mathbb{R}^2$ and $\tau \in [0, \pi)$, the Radon transform is given by

$$\mathcal{R}\chi_{E_{R_1,R_2,y,\tau}}(r,\theta) = \frac{2R_1R_2 \cdot \sqrt{R_2^2 \cdot \sin(\theta-\tau)^2 + R_1^2 \cdot \cos(\theta-\tau)^2 - (r - y_1 \cdot \cos(\theta) - y_2 \cdot \sin(\theta))^2}}{R_2^2 \cdot \sin(\theta-\tau)^2 + R_1^2 \cdot \cos(\theta-\tau)^2}$$

if $(r - y_1 \cdot \cos(\theta) - y_2 \cdot \sin(\theta))^2 \leq R_2^2 \cdot \sin(\theta-\tau)^2 + R_1^2 \cdot \cos(\theta-\tau)^2$ holds for $(r,\theta) \in \mathbb{R} \times [0,\pi)$. Otherwise, we have $\mathcal{R}\chi_{E_{R_1,R_2,y,\tau}}(r,\theta) = 0$. We do not go further into detail here and refer to [52, Subsection 2.5.2] for the computations.

(iii) Consider the shifted rectangle

$$E_{R_1,R_2,y} = [y_1 - R_1, y_1 + R_1] \times [y_2 - R_2, y_2 + R_2] \subset \mathbb{R}^2$$

with halved side lengths $R_1, R_2 > 0$ and center $y = (y_1, y_2)^T \in \mathbb{R}^2$ (cf. Figure 8.3). The Radon transform of the respective characteristic function is given by

$$R_{\chi_{E_{R_1,R_2,y}}}(r,\theta) = \begin{cases} 2R_2 \cdot \chi_{[-R_1,R_1]}(r - y_1), & \text{for } \theta = 0,\ r \in \mathbb{R} \\ 2R_1 \cdot \chi_{[-R_2,R_2]}(r - y_2), & \text{for } \theta = \pi/2,\ r \in \mathbb{R} \\ \max\left(\min\left(s_{V,1}, s_{H,2}\right) - \max\left(s_{V,2}, s_{H,1}\right),\ 0\right), & \text{for } \theta \in (0,\pi/2),\ r \in \mathbb{R} \\ \max\left(\min\left(s_{V,1}, s_{H,1}\right) - \max\left(s_{V,2}, s_{H,2}\right),\ 0\right), & \text{for } \theta \in (\pi/2,\pi),\ r \in \mathbb{R}, \end{cases}$$

where the intersection points with the rectangle in dependence of r, θ are given by

$$s_{V,1}, s_{V,2} = \frac{r \cdot \cos(\theta) - y_1 \pm R_1}{\sin(\theta)}$$
$$s_{H,1}, s_{H,2} = -\frac{r \cdot \sin(\theta) - y_2 \pm R_2}{\cos(\theta)},$$

see also [52, Subsection 2.5.4]. Note that the formula can easily be adapted to the case of (half-)open rectangles.

(iv) Given a shape parameter $\nu \geq 0$, we consider the function

$$f_\nu(x) = \begin{cases} \left(1 - \|x\|_2^2\right)^\nu, & \text{for } \|x\|_2 \leq 1 \\ 0, & \text{for } \|x\|_2 > 1. \end{cases}$$

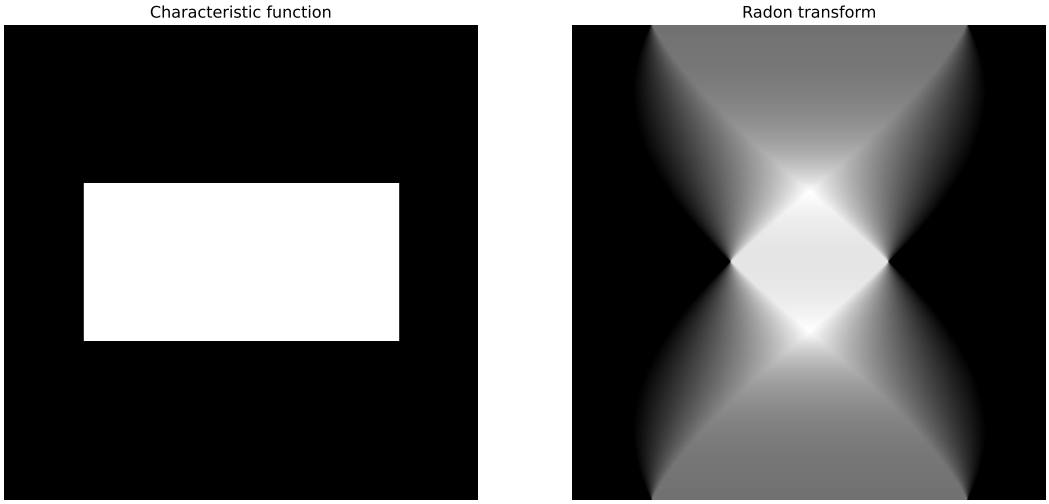

Figure 8.3.: Characteristic function of a rectangular region (left) and its Radon transform (right)

For $(r, \theta) \in \mathbb{R} \times [0, \pi)$, the Radon value is given by

$$\mathcal{R}f_\nu(r,\theta) = \begin{cases} \frac{\sqrt{\pi}\cdot\Gamma(\nu+1)}{\Gamma(\nu+\frac{3}{2})} \cdot \left(1-r^2\right)^{\nu+\frac{1}{2}}, & \text{if } |r| \leq 1 \\ 0, & \text{if } |r| > 1, \end{cases}$$

where Γ denotes the *gamma function* in this particular case (cf. [18, Proposition 5.2.5]). Similar to the rotated ellipse from (ii), we can modify the function via rotations, shifts and stretching as well, see Section 11.1. The function f_ν and its Radon transform are visualized in Figure 8.4 for $\nu = 3$.

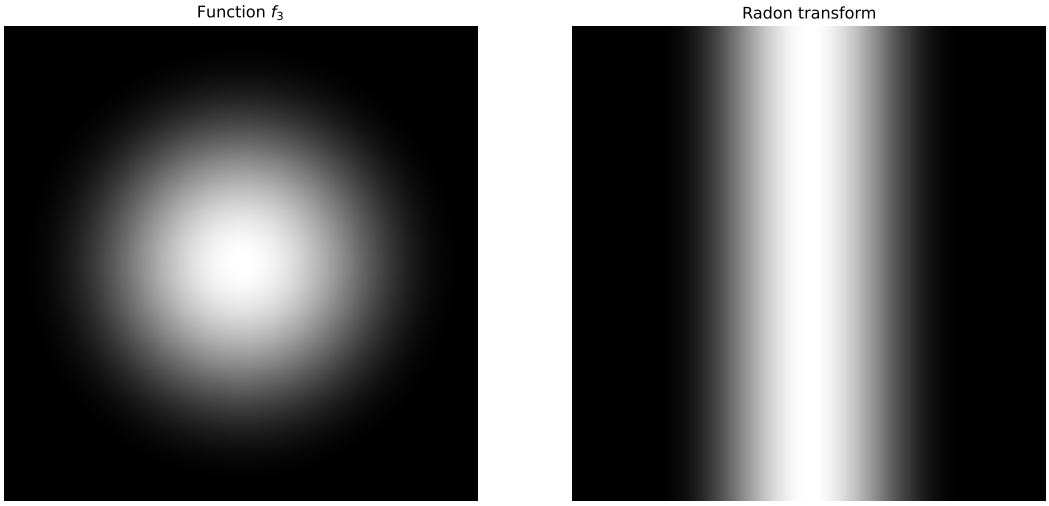

Figure 8.4.: Visualization of f_ν (left) and its Radon transform (right) for $\nu = 3$

In the previous example, we have seen that the Radon transform of a Gaussian function is again a Gaussian function, which only depends on the radial argument $r \in \mathbb{R}$. Thus, one is usually interested in which properties of the function f are inherited by its Radon transform. A discussion regarding the differentiability can be found in [60, Section 1.2]. Here, we want to focus on the support of the resulting Radon transform (cf. [18, Proposition 2.2.9]).

Proposition 8.6. *Let $f \in L^1(\mathbb{R}^2) \cap \mathscr{C}(\mathbb{R}^2)$ have compact support and $R > 0$ such that*

$$f(x) = 0 \qquad \text{for all } x \in \mathbb{R}^2 \setminus B_R(0).$$

Then $\mathcal{R}f$ is compactly supported with respect to the radial variable, i.e.

$$\mathcal{R}f(r,\theta) = 0 \qquad \text{for } |r| \geq R, \theta \in [0,\pi)\,.$$

Proof. For $|r| \geq R$ and $\theta \in [0,\pi)$, we have

$$\|r \cdot n_\theta + s \cdot v_\theta\|_2^2 = r^2 + s^2 \geq r^2 \geq R^2 \qquad \text{for all } s \in \mathbb{R},$$

which already implies that the respective line integral of f must vanish. □

With Proposition 8.2 and Definition 8.3, the reconstruction problem can be written as the operator equation

$$\mathcal{R}f = g, \tag{8.7}$$

where g contains the measured Radon data and f is the unknown function that we want to reconstruct. In order to have a unique solution, the Radon transform is required to be injective. This is a consequence of the *central slice theorem* (cf. [52, Theorem 6.1]).

Theorem 8.7 (Central slice theorem)**.** *Let $\mathcal{F}_i$ denote the Fourier transform on $L^1(\mathbb{R}^i)$ for $i = 1, 2$. If $f \in L^1(\mathbb{R}^2)$, we have the equality*

$$\mathcal{F}_1\left[\mathcal{R}f(\cdot,\theta)\right](S) = (2\pi)^{1/2} \cdot \mathcal{F}_2 f(S\cdot\cos(\theta), S\cdot\sin(\theta)) \qquad \text{for all } S \in \mathbb{R}$$

and fixed angle $\theta \in [0,\pi)$.

Proof. Given $\theta \in [0,\pi)$, we have $\mathcal{R}f(\cdot,\theta) \in L^1(\mathbb{R})$ due to (8.5). Hence, we compute

$$\begin{aligned}
\mathcal{F}_1\left[\mathcal{R}f(\cdot,\theta)\right](S) &= (2\pi)^{-1/2} \cdot \int_{\mathbb{R}} \mathcal{R}f(r,\theta)\cdot e^{-iSr}\,dr \\
&= (2\pi)^{-1/2} \cdot \int_{\mathbb{R}}\int_{\mathbb{R}} f(r\cdot n_\theta + s\cdot v_\theta)\cdot e^{-iSr}\,ds\,dr \\
&= (2\pi)^{-1/2} \cdot \int_{\mathbb{R}}\int_{\mathbb{R}} f(x)\cdot e^{-iS\cdot(x_1\cdot\cos(\theta)+x_2\cdot\sin(\theta))}\,dx \\
&= (2\pi)^{1/2} \cdot \mathcal{F}_2 f(S\cdot\cos(\theta), S\cdot\sin(\theta)),
\end{aligned}$$

where we again used the transformation from (8.3). □

Corollary 8.8. *The Radon transform $\mathcal{R} : L^1(\mathbb{R}^2) \to L^1(\mathbb{R}\times[0,\pi))$ is injective.*

Proof. Let $f \in L^1(\mathbb{R}^2)$ satisfy $\mathcal{R}f = 0$. Then we have

$$0 = \|\mathcal{R}f\|_{L^1(\mathbb{R}\times[0,\pi))} = \int_0^\pi \|\mathcal{R}f(\cdot,\theta)\|_{L^1(\mathbb{R})}\,d\theta,$$

and therefore $\mathcal{R}f(\cdot,\theta) = 0$ in L^1-sense for almost every $\theta \in [0,\pi)$. With Theorem 8.7, we conclude that

$$\mathcal{F}_2 f(S\cdot\cos(\theta), S\cdot\sin(\theta)) = 0 \qquad \text{for all } S \in \mathbb{R}$$

holds for almost every $\theta \in [0,\pi)$. Since $\mathcal{F}_2 f$ is continuous, this property must hold for all $\theta \in [0,\pi)$. Writing the elements of $\mathbb{R}^2$ in polar coordinates gives $\mathcal{F}_2 f = 0$, and therefore $f = 0$ due to the injectivity of the Fourier transform. □

8.2. Filtered back projection formula

Due to Corollary 8.8, there is a unique reconstruction of a function $f \in L^1(\mathbb{R}^2)$ from its Radon transform. Besides the injectivity of the Radon transform, an explicit reconstruction formula is desirable for practical implementations. Hence, we derive the reconstruction formula in this section, including the numerical problems and challenges it brings along.

As a first reasonable approach for the reconstruction, we consider the mean Radon value of all lines that go through a fixed point $x = (x_1, x_2)^T \in \mathbb{R}^2$. Note that for any $\theta \in [0, \pi)$, we can find exactly one $r \in \mathbb{R}$ such that $x \in \ell_{r,\theta}$ holds. In order to identify the unique radius parameter, we observe that $x \in \ell_{r,\theta}$ holds if and only if there is $s \in \mathbb{R}$ with

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x = r \cdot \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} + s \cdot \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \cdot \begin{pmatrix} r \\ s \end{pmatrix}.$$

Applying the inverse matrix and extracting the first entry of the vector shows that the set of all lines passing through the point x is given by

$$AG(\mathbb{R}^2)_x := \left\{ \ell_{r,\theta} \;\middle|\; \theta \in [0, \pi),\ r = x_1 \cdot \cos(\theta) + x_2 \cdot \sin(\theta) \right\}.$$

Consequently, the mean of all respective Radon values is given by

$$\frac{1}{\pi} \int_0^\pi \mathcal{R}f(x_1 \cdot \cos(\theta) + x_2 \cdot \sin(\theta), \theta)\, d\theta$$

for all functions $f \in L^1(\mathbb{R}^2)$ and points $x \in \mathbb{R}^2$. The above integral gives rise to the definition of the *back projection* (cf. [52, Definition 3.2]).

Definition 8.9 (Back projection)**.** For $g : \mathbb{R} \times [0, \pi) \to \mathbb{R}$, its `back projection` is defined as

$$\mathcal{B}g(x) := \frac{1}{\pi} \int_0^\pi g(x_1 \cdot \cos(\theta) + x_2 \cdot \sin(\theta), \theta)\, d\theta \qquad \text{for all } x = (x_1, x_2)^T \in \mathbb{R}^2.$$

The back projection satisfies the following mapping properties (cf. [18, Proposition 2.2.16 ff.]).

Proposition 8.10. *For $g \in L^1(\mathbb{R} \times [0, \pi))$, its back projection $\mathcal{B}g$ is defined almost everywhere on $\mathbb{R}^2$ and locally integrable, i.e. for any compact set $E \subset \mathbb{R}^2$, we have $\mathcal{B}g \in L^1(E)$. Moreover, the back projection*

$$\mathcal{B} : L^\infty(\mathbb{R} \times [0, \pi)) \to L^\infty(\mathbb{R}^2)$$

is a well-defined, bounded linear operator.

Unlike the Radon transform and Example 8.5, it is not easy to find relevant non-constant functions whose back projection can be computed explicitly. Some artificial examples are given in [52, Section 3.3]. But one can at least estimate the integrals to show that the back projection does not invert the Radon transform. Counterexamples are given in [18, Observation 2.2.20] and [52, Example 3.5].

Although the back projection is not the inverse operator, it still plays a key role in the actual reconstruction formula, where the Radon transform is pre-filtered by the modulus of the radial variable. Due to this pre-filtering, the formula is known as the *filtered back projection formula*. Here, we state the version of [18, Theorem 2.2.22], where we use the result and notation of Theorem 8.7.

Theorem 8.11 (Filtered back projection formula)**.** *Let $f \in L^1(\mathbb{R}^2) \cap \mathscr{C}(\mathbb{R}^2)$ such that $\mathcal{F}_2 f \in L^1(\mathbb{R}^2)$. Then, f can be reconstructed via the formula*

$$f(x) = \frac{1}{2} \cdot \mathcal{B}\left[\mathcal{F}_1^{-1}\left(|\cdot| \cdot \mathcal{F}_1\left[\mathcal{R}f\right]\right)\right](x) \qquad \textit{for all } x \in \mathbb{R}^2,$$

where $\mathcal{F}_1\left[\mathcal{R}f\right](S, \theta) := \mathcal{F}_1\left[\mathcal{R}f(\cdot, \theta)\right](S)$ for $(S, \theta) \in \mathbb{R} \times [0, \pi)$.

Proof. With the Fourier inversion formula from Theorem A.15, the function f can be written as

$$f(x) = \mathcal{F}_2^{-1}[\mathcal{F}_2 f](x) = \frac{1}{2\pi} \cdot \int_{\mathbb{R}^2} \mathcal{F}_2 f(\omega) \cdot e^{i \cdot \langle x, \omega \rangle_2} \, d\omega \qquad \text{for all } x \in \mathbb{R}^2.$$

Using the polar coordinates diffeomorphism

$$T_P : \mathbb{R} \setminus \{0\} \times (0, \pi) \to \mathbb{R}^2 \setminus \{(x_1, 0) \mid x_1 \in \mathbb{R}\}, \ (S, \theta) \mapsto (S \cdot \cos(\theta), S \cdot \sin(\theta))$$

with $|\det(DT_P(S, \theta))| = |S|$ for $(S, \theta) \in \mathbb{R} \times (0, \pi)$, we get

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \cdot \int_0^\pi \int_{\mathbb{R}} \mathcal{F}_2 f(T_P(S, \theta)) \cdot e^{i \cdot \langle x, T_P(S,\theta) \rangle_2} \cdot |S| \, dS \, d\theta \\ &= \frac{1}{2\pi} \cdot \int_0^\pi (2\pi)^{-1/2} \cdot \int_{\mathbb{R}} |S| \cdot \mathcal{F}_1[\mathcal{R}f](S, \theta) \cdot e^{iS \cdot (x_1 \cdot \cos(\theta) + x_2 \cdot \sin(\theta))} \, dS \, d\theta \\ &= \frac{1}{2} \cdot \mathcal{B}\left[\mathcal{F}_1^{-1}\left(|\cdot| \cdot \mathcal{F}_1\left[\mathcal{R}f\right]\right)\right](x) \end{aligned}$$

for every $x \in \mathbb{R}^2$, where we used the result from Theorem 8.7 in the second line. □

Remark 8.12. The first reconstruction formula from the complete set of line integral values goes back to Radon and his seminal paper [109] from 1917. In the paper, he proved that f can be reconstructed via

$$f(x) = -\frac{1}{\pi} \cdot \int_0^\infty \frac{F_x'(q)}{q} \, dq \qquad \text{for all } x \in \mathbb{R}^d \tag{8.8}$$

under suitable regularity assumptions, where $F_x' = \partial F_x / \partial q$ is the derivative of

$$F_x(q) := \frac{1}{2\pi} \cdot \int_0^{2\pi} \mathcal{R}f(x_1 \cdot \cos(\theta) + x_2 \cdot \sin(\theta) + q, \theta) \, d\theta \qquad \text{for } x = (x_1, x_2)^T \in \mathbb{R}^2, q \in [0, \infty).$$

Note that the value $F_x(q)$ is the mean value of the line integrals along all tangents of the closed ball with center $x \in \mathbb{R}^2$ and radius $q \geq 0$. Alternatively, a derivation of the formula (8.8) using *Riesz potentials* and the *Hilbert transform* can be found in [101, Section II.2].

The filtered back projection formula solves the reconstruction problem in theory but is problematic when it comes to practical implementations. On the one hand, there is only a finite set of Radon data measurements available in practical cases. This means that we have to use adequate discretizations of the occuring operators to avoid large discretization errors. On the other hand, the reconstruction formula is highly sensitive to noise, since the high-frequency part of $\mathcal{R}f$, which is usually associated with noise, is amplified by the filter function $|\cdot|$ (cf. [18, Observation 2.2.24]). Hence, modifications are needed to improve the stability of the reconstruction. We discuss modifications for the filtered back projection formula in Section 9.1 that lead to a suitable reconstruction method.

8.3. Ill-posed inverse problems

To close this chapter, we briefly introduce the notion of *ill-posed inverse problems* and further comment on the well-posedness of the computerized tomography problem. Recall that, due to equation (8.7), the reconstruction problem can be interpreted as a linear operator equation. In a more general setting, we assume that $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ are normed linear spaces and $A : X \to Y$ is a bounded linear operator. According to Hadamard, the linear inverse problem

$$Ax = y \qquad \text{for } y \in Y \tag{8.9}$$

is *well-posed* if the following conditions are satisfied (cf. [45, page 31]):

(i) The operator A is surjective, i.e. for every $y \in Y$, there is at least one $x \in X$ that solves (8.9).

(ii) The operator A is injective, i.e. for every $y \in Y$, there is at most one $x \in X$ that solves (8.9).

(iii) The inverse operator $A^{-1} : Y \to X$ is bounded as well, i.e. the solution of (8.9) depends continuously on $y \in Y$.

The problem (8.9) is called *ill-posed* if at least one of the conditions (i)-(iii) does not hold. Given an injective operator like the Radon transform, it is easy to achieve the first property by restricting the co-domain of A to the image $\mathrm{im}(A)$. However, the third property, which is highly relevant in practice since it reflects the effect of noise on the reconstruction quality, is not satisfied in many cases. A collection of ill-posed inverse problems is given in [74, Section 3]. In this section, we want to discuss two concepts that demonstrate the ill-posedness of the computerized tomography problem. The following discussion is mainly based on [101, Chapter IV].

Singular values and ill-posedness

Let X and Y be Hilbert spaces. A representation of the bounded linear mapping $A : X \to Y$ in the form

$$Ax = \sum_{i \in I} \sigma_i \cdot \langle x, x_i \rangle_X \cdot y_i \qquad \text{for } x \in X, \tag{8.10}$$

where $\Sigma := \{\sigma_i \mid i \in I\} \subset \mathbb{R}$ are positive real numbers and $\{x_i \mid i \in I\} \subset X$, $\{y_i \mid i \in I\} \subset Y$ are orthonormal, is called a *singular value decomposition* of A (cf. [101, page 86 ff.]). Moreover, the elements of Σ are called *singular values*. The equation (8.10) implies that

$$AA^* y = \sum_{i \in I} \sigma_i^2 \cdot \langle y, y_i \rangle_Y \cdot y_i \qquad \text{for } y \in Y,$$

where $A^* : Y \to X$ is the adjoint operator of A. Hence, the singular values and orthonormal vectors can be determined with an eigenvalue decomposition of the self-adjoint mapping AA^* and the relation

$$x_i = \sigma_i^{-1} \cdot A^* y_i \qquad \text{for } i \in I.$$

It can be shown that a singular value decomposition exists for certain classes of linear operators, e.g. for compact operators (see, e.g., [147, Satz VI.3.6]). Given the decomposition (8.10) of an injective operator, the inverse is given by

$$A^{-1} y = \sum_{i \in I} \sigma_i^{-1} \cdot \langle y, y_i \rangle_Y \cdot x_i \qquad \text{for } y \in \mathrm{im}(A).$$

In the case of a non-injective operator, the representation holds for the *Moore-Penrose generalized inverse* A^+ (cf. [101, Theorem IV.1.2]). Due to the equality

$$\|A^{-1} y_i\|_X = \sigma_i^{-1} \qquad \text{for } i \in I,$$

the inverse operator is unbounded and therefore discontinuous if zero is an accumulation point of Σ. Consequently, the singular values reflect the stability of the linear operator equation (8.9) in terms of $y \in Y$. To demonstrate the utility of this theory in the context of computerized tomography, we consider the Radon transform

$$\mathcal{R} : L^2(\overline{B_1(0)}) \to L^2_w([-1,1] \times [0,\pi))$$

between the two Hilbert spaces

$$L^2(\overline{B_1(0)}) \simeq \left\{ f \in L^2(\mathbb{R}^2) \;\middle|\; \mathrm{supp}(f) \subset \overline{B_1(0)} \right\}$$

and

$$L^2_w([-1,1] \times [0,\pi)) \simeq \left\{ [g : \mathbb{R} \times [0,\pi) \to \mathbb{R}] \;\middle|\; \mathrm{supp}(g) \subset [-1,1] \times [0,\pi) \text{ and } \|g\|_{L^2_w([-1,1]\times[0,\pi))} < \infty \right\}$$

with norm

$$\|g\|_{L^2_w([-1,1]\times[0,\pi))} = \left(\int_0^\pi \int_{-1}^1 |g(r,\theta)|^2 \cdot \left(1-r^2\right)^{-1/2} \, dr \, d\theta \right)^{1/2}.$$

This particular version of the Radon transform is a bounded linear operator (cf. [101, Theorem II.1.6]) and its positive singular values are given by

$$\sigma_i = \frac{4\pi}{i+1} \qquad i \in \mathbb{N}_0.$$

For the derivation of the singular value decomposition, we refer to [101, Section IV.3]. Since $\sigma_i \xrightarrow{i\to\infty} 0$, the (generalized) inverse operator is not bounded according to our previous discussion.

Degree of smoothing

For linear operators between L^2-spaces, we can use its smoothing properties to classify the ill-posedness of the problem (8.9). To this end, let $\Omega_1 \subset \mathbb{R}^{d_1}, \Omega_2 \subset \mathbb{R}^{d_2}$ be suitable domains and assume that

$$A : L^2(\Omega_1) \to L^2(\Omega_2).$$

We call A a *smoothing operator of degree* $s > 0$ if for any $a \in \mathbb{R}$ and any bounded open subset $\Omega \subset \Omega_1$, there are constants $c_{a,\Omega}, C_{a,\Omega} > 0$ such that the Sobolev norm estimates

$$c_{a,\Omega} \cdot \|f\|_{H^a(\Omega_1)} \le \|Af\|_{H^{a+s}(\Omega_2)} \le C_{a,\Omega} \cdot \|f\|_{H^a(\Omega_1)} \qquad \text{for all } f \in H_0^a(\Omega) \tag{8.11}$$

hold. Here, $H_0^a(\Omega)$ denotes the subspace

$$H_0^a(\Omega) := \left\{ f \in H^a(\Omega_1) \;\middle|\; \operatorname{supp}(f) \subset \overline{\Omega} \right\}.$$

Further information on fractional Sobolev spaces is provided in Section A.5. Note that (8.11) implies that $A : H_0^a(\Omega) \to H^{a+s}(\Omega_2)$ is an injective, bounded operator and its inverse operator is continuous. Setting $a = -s$, we get

$$c_{-s,\Omega} \cdot \|f\|_{H^{-s}(\Omega_1)} \le \|Af\|_{L^2(\Omega_2)} \le C_{-s,\Omega} \cdot \|f\|_{H^{-s}(\Omega_1)} \qquad \text{for all } f \in H_0^{-s}(\Omega_1). \tag{8.12}$$

Since $-s < 0$, the space $H_0^{-s}(\Omega)$ is a Sobolev space of negative order and its norm is (strictly) weaker than the L^2-norm, i.e. there is a constant $C > 0$ such that

$$\|f\|_{H^{-s}(\Omega_1)} \le C \cdot \|f\|_{L^2(\Omega_1)} \qquad \text{for all } f \in L^2(\Omega)$$

and a sequence $(f_n)_{n\in\mathbb{N}} \subset L^2(\Omega)$ such that

$$\|f_n\|_{L^2(\Omega_1)} > n \cdot \|f_n\|_{H^{-s}(\Omega_1)} \qquad \text{for } n \in \mathbb{N}.$$

In combination with (8.12), this yields

$$\|f_n\|_{L^2(\Omega_1)} > n \cdot C_{-s,\Omega}^{-1} \cdot \|Af_n\|_{L^2(\Omega_2)} \qquad \text{for } n \in \mathbb{N},$$

so that A^{-1} is not continuous in the L^2-setting (cf. [18, Remark 2.2.29], [85, Section III.B]). The degree of smoothing can also be used to analyze worst-case errors of the reconstruction, see [101, Theorem IV.1.3]. For the analysis of the Radon transform, we set $\Omega_1 = \mathbb{R}^2$ and $\Omega_2 = \mathbb{R} \times [0,\pi)$. In particular, we set

$$H^a(\mathbb{R}\times[0,\pi)) = \left\{ [g : \mathbb{R}\times[0,\pi) \to \mathbb{R}] \;\middle|\; g(\cdot,\theta) \in H^a(\mathbb{R}) \text{ for all } \theta \in [0,\pi) \text{ and } \|g\|_{H^a(\mathbb{R}\times[0,\pi)} < \infty \right\}$$

for $a \in \mathbb{R}$, where the Sobolev norm is defined as

$$\|g\|_{H^a(\mathbb{R}\times[0,\pi))} := \left(\int_0^\pi \|g(\cdot,\theta)\|^2_{H^a(\mathbb{R})} \, d\theta \right)^{1/2}.$$

Based on [101, Theorem II.5.1], it was shown in [18, Theorem 2.2.28] that the Radon transform is a smoothing operator of degree $s = 1/2$, i.e. there are constants $c_{a,\Omega}, C_{a,\Omega} > 0$ such that

$$c_{a,\Omega} \cdot \|f\|_{H^a(\mathbb{R}^2)} \leq \|\mathcal{R}f\|_{H^{a+1/2}(\mathbb{R}\times[0,\pi))} \leq C_{a,\Omega} \cdot \|f\|_{H^a(\mathbb{R}^2)} \qquad \text{for all } f \in H_0^a(\Omega) \cap L^1(\mathbb{R}^2)$$

given a smoothness level $a \in \mathbb{R}$ and a bounded open set $\Omega \subset \mathbb{R}^2$. This is a slight modification of (8.11) due to the intersection with $L^1(\mathbb{R}^2)$, but our previous argumentation still applies. Hence, the Radon transform does not have a continuous inverse operator in the L^2-setting.

The singular values of the Radon transform and its smoothing properties show that the problem of reconstructing a function from its line integral values is ill-posed. Hence, so-called *regularization methods* are usually required for a stable reconstruction (cf. [45]). Common regularization methods are also listed in [101, pages 86-89], for example.

9. Reconstruction Methods

As we have already pointed out, the filtered back projection formula from Theorem 8.11 is problematic when it comes to practical implementations due to incomplete data and its sensitivity to noise. Consequently, an exact reconstruction is not possible and approximation methods are needed to produce a reconstruction that is similar to the original. Over the last decades, several reconstruction methods have been developed that rely on different simplifications of the reconstruction problem. An overview of reconstruction methods is given in [21], [61], [78], [102].

Our main focus in this chapter lies on the *method of filtered back projection (FBP method)* and *algebraic reconstruction methods*, which are two of the most popular reconstruction approaches. We explain the ideas behind the two approaches as well as important theoretical results in Section 9.1 & 9.2 since the corresponding reconstruction methods serve as a reference for our theoretical discussions in Chapter 10 and numerical comparison in Section 11.3. For the sake of completeness, we provide a selection of more recent reconstruction methods with brief descriptions in Section 9.3.

9.1. Method of filtered back projection

As the name suggests, the FBP method is based on the filtered back projection formula from Theorem 8.11. Recall that the modulus $|\cdot|$ inside the formula is mainly responsible for its instability since it amplifies the noise of the Radon data for high values of $S \in \mathbb{R}$. The main idea of the FBP method is to replace the modulus with a compactly supported function $A : \mathbb{R} \to \mathbb{R}$ that imitates the modulus on a compact domain and cuts off high frequencies. In the language of signal processing, this is called a *low-pass filter*, and the function A serves as a *filter function*. The following section is based on [18, Chapter 3-5], [52, Chapter 8] and [101, Chapter III].

Let us start with the construction of the filter function A. Usually, it has the form

$$A(S) = A_{W,L}(S) = |S| \cdot W(S/L) \qquad \text{for } S \in \mathbb{R}, \tag{9.1}$$

where $W \in L^\infty(\mathbb{R})$ is an even window function with $\operatorname{supp}(W) \subset [-1,1]$ and $L > 0$ is a given bandwidth. In the format (9.1), the bandwidth L determines the length of the support, as the support of A_L is contained in $[-L, L]$. Note that, due to the compact support of $A_L \in L^\infty(\mathbb{R})$, we must have $A_L \in L^p(\mathbb{R})$ for every $p \in [1, \infty]$. Replacing the modulus with A_L inside the filtered back projection formula leads to the approximate reconstruction

$$f_{W,L}(x) = \frac{1}{2} \cdot \mathcal{B}\left[\mathcal{F}_1^{-1}\left(A_{W,L} \cdot \mathcal{F}_1\left[\mathcal{R}f\right]\right)\right](x) = \frac{1}{2} \cdot \mathcal{B}\left[\mathcal{F}_1^{-1}\left(|\cdot| \cdot W(\cdot/L) \cdot \mathcal{F}_1\left[\mathcal{R}f\right]\right)\right](x) \qquad \text{for } x \in \mathbb{R}^2.$$

In the original formula, the term $\mathcal{F}_1^{-1}\left(|\cdot| \cdot \mathcal{F}_1\left[\mathcal{R}f\right]\right)$ cannot be simplified with the convolution property of the Fourier transform because $|\cdot| \notin L^p(\mathbb{R})$ for every $p \in [1, \infty]$. But with the replacement of the modulus, we can write this term as a convolution of functions (cf. [18, Theorem 3.2.2]).

Theorem 9.1. *Let $f \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ and $W \in L^\infty(\mathbb{R})$ be even with $\operatorname{supp}(W) \subset [-1,1]$. For any bandwidth $L > 0$, the approximate reconstruction can be written as*

$$f_{W,L} = \frac{1}{2} \cdot \mathcal{B}\left(q_L \star \mathcal{R}f\right), \tag{9.2}$$

where the equality holds almost everywhere on $\mathbb{R}^2$. Here, the function $q_L : \mathbb{R} \times [0, \pi) \to \mathbb{R}$ is given by

$$q_L(r, \theta) := \mathcal{F}_1^{-1} A_{W,L}(r) \qquad \text{for } (r, \theta) \in \mathbb{R} \times [0, \pi)$$

and the convolution $q_L \star \mathcal{R}f$ is defined as the one-dimensional convolution

$$\left(q_L \star \mathcal{R}f\right)(r, \theta) := \left(q_L(\cdot, \theta) * \mathcal{R}f(\cdot, \theta)\right)(r) = \int_{\mathbb{R}} q_L(t, \theta) \cdot \mathcal{R}f(r - t, \theta)\, dt \qquad \text{for } (r, \theta) \in \mathbb{R} \times [0, \pi).$$

With this representation, we get rid of the Fourier transform in front of $\mathcal{R}f$, which is desirable for implementations, as we are only given finite samples of the Radon data. But we need to determine the inverse Fourier transform of the filter function $A_{W,L}$ in order to use (9.2).

Example 9.2. We provide two common examples of window functions $W \in L^\infty(\mathbb{R})$ including the inverse Fourier transform of the resulting filter functions (cf. [52, Section 8.3]). A visualization is given in Figure 9.1. More examples and a numerical comparison can be found in [31].

(i) The characteristic function

$$W(S) = \chi_{[-1,1]}(S) = \begin{cases} 1, & \text{if } |S| \leq 1 \\ 0, & \text{if } |S| > 1 \end{cases} \qquad \text{for } S \in \mathbb{R}$$

generates the *Ram-Lak filter* (cf. [110])

$$A_L(S) = |S| \cdot W(S/L) = \begin{cases} |S|, & \text{if } |S| \leq L \\ 0, & \text{if } |S| > L \end{cases} \qquad \text{for } S \in \mathbb{R} \tag{9.3}$$

for given bandwidth $L > 0$. Its inverse Fourier transform is given by

$$\mathcal{F}_1^{-1} A_L(r) = \frac{L^2}{2\pi} \cdot \left(2 \cdot \operatorname{sinc}(Lr) - \operatorname{sinc}(Lr/2)^2\right) \qquad \text{for } r \in \mathbb{R},$$

where *sinc* denotes the *cardinal sine*

$$\operatorname{sinc}(t) := \begin{cases} \frac{\sin(t)}{t}, & \text{if } t \neq 0 \\ 1, & \text{if } t = 0. \end{cases}$$

(ii) Consider the window function

$$W(S) = \operatorname{sinc}\left(\frac{\pi}{2} \cdot S\right) \cdot \chi_{[-1,1]}(S) = \begin{cases} \frac{\sin\left(\frac{\pi}{2} \cdot S\right)}{\frac{\pi}{2} \cdot S}, & \text{if } 0 < |S| \leq 1 \\ 1 & \text{if } S = 0 \\ 0 & \text{if } |S| > 1 \end{cases} \qquad \text{for } S \in \mathbb{R},$$

which, given $L > 0$, yields the *Shepp-Logan filter* (cf. [124])

$$A_L(S) = |S| \cdot \operatorname{sinc}\left(\frac{\pi}{2} \cdot \frac{S}{L}\right) \cdot \chi_{[-1,1]}(S/L) = \begin{cases} \frac{2L}{\pi} \cdot \left|\sin\left(\frac{\pi}{2} \cdot \frac{S}{L}\right)\right|, & \text{if } |S| \leq L \\ 0, & \text{if } |S| > L \end{cases} \qquad \text{for } S \in \mathbb{R}.$$

The respective inverse Fourier transform is given by

$$\mathcal{F}_1^{-1} A_L(r) = \frac{4L^2}{\pi^2} \cdot \frac{\pi - 2Lr \cdot \sin(Lr)}{\pi^2 - 4L^2 r^2} \qquad \text{for } r \in \mathbb{R}.$$

Convergence rates

In his doctoral thesis [18], Beckmann analyzed the reconstruction error of the FBP method for target functions in the *fractional Sobolev space*

$$\mathcal{H}^\alpha(\mathbb{R}^2) = \left\{ f \in L^2(\mathbb{R}^2) \mid \|f\|_{\mathcal{H}^\alpha(\mathbb{R}^2)} < \infty \right\}$$

of order $\alpha \geq 0$ (cf. Definition A.29), where the respective norm is defined as

$$\|f\|^2_{\mathcal{H}^\alpha(\mathbb{R}^2)} := \int_{\mathbb{R}^2} \left(1 + \|\omega\|_2^2\right)^\alpha \cdot |\mathcal{F}f(\omega)|^2 \, d\omega.$$

Note that $\mathcal{H}^0(\mathbb{R}^2) = L^2(\mathbb{R}^2)$ holds (cf. Section A.5). The error analysis in [18] is based on the result that the approximate reconstruction from Theorem 9.1 can be written as a convolution of the target function with a band-limited convolution kernel due to the convolution properties of the back projection (cf. [18, Theorem 3.2.3 & 3.2.5]).

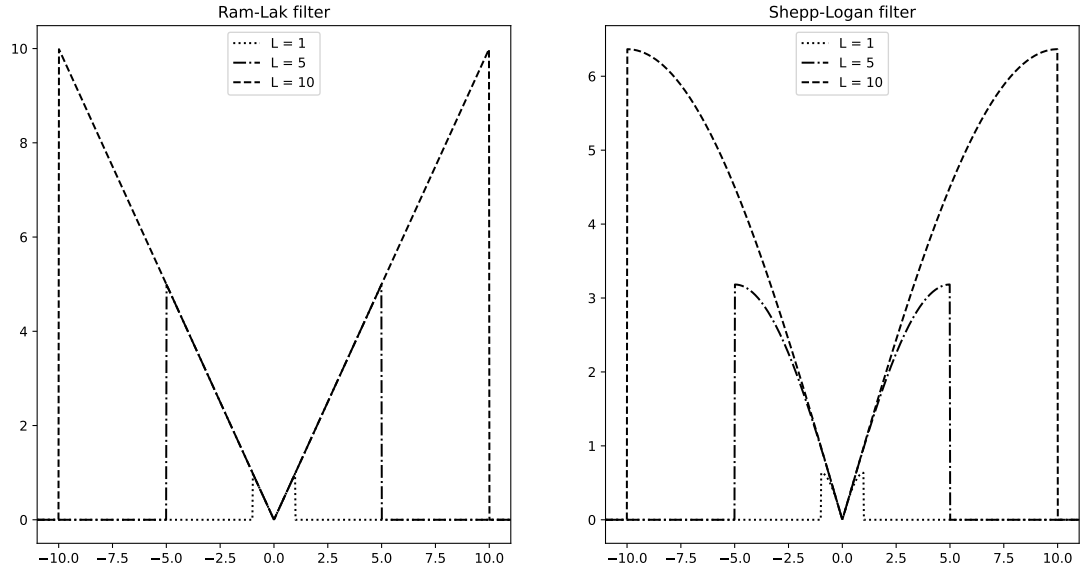

Figure 9.1.: Visualisation of Ram-Lak filter (left) and Shepp-Logan filter (right) for $L \in \{1, 5, 10\}$

Theorem 9.3. *Let $f \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ and $W \in L^\infty(\mathbb{R})$ be even with supp$(W) \subset [-1,1]$. Then, for any $L > 0$, the equality*

$$f_{W,L} = f * K_L$$

holds in the L^2-sense, where $K_L = \frac{1}{2} \cdot \mathcal{B}q_L$ is the inverse Fourier transform of the function

$$W_L(x) := W\left(\frac{\|x\|_2}{L}\right) \qquad \text{for } x \in \mathbb{R}^2.$$

Here, we want to give a short overview of the results from [18, Chapter 4]. We start with the error estimate in terms of the Sobolev norm (cf. [18, Theorem 4.2.3]).

Theorem 9.4. *Let $f \in L^1(\mathbb{R}^2) \cap \mathcal{H}^\alpha(\mathbb{R}^2)$ for $\alpha > 0$ and let $W \in L^\infty(\mathbb{R})$ be even with supp$(W) \subset [-1,1]$. Then, for $0 \le \sigma \le \alpha$, we have*

$$\|f - f_{W,L}\|_{H^\sigma(\mathbb{R}^2)} \le \left(\Phi_{\alpha-\sigma,W}(L)^{1/2} + L^{\sigma-\alpha}\right) \cdot \|f\|_{H^\alpha(\mathbb{R}^2)},$$

where

$$\Phi_{\alpha-\sigma,W}(L) := \operatorname*{ess\,sup}_{S\in[-1,1]} \frac{(1-W(S))^2}{(1+L^2S^2)^{\alpha-\sigma}} \qquad \text{for } L > 0.$$

In order to prove convergence for $L \to \infty$ and to determine convergence rates, the functions

$$\Phi_{\gamma,W}(L) := \operatorname*{ess\,sup}_{S\in[-1,1]} \frac{(1-W(S))^2}{(1+L^2S^2)^{\gamma}} \qquad \text{for } L > 0 \tag{9.4}$$

for $\gamma > 0$ need to be analyzed. If W is the window function of the Ram-Lak filter (cf. Example 9.2 (i)), then $\Phi_{\gamma.W}(L) = 0$ holds for all $L > 0$ so that it represents the optimal window functions for this type of estimate. More generally, the expression (9.4) tends to zero for continuous window functions that are normalized at zero (cf. [18, Theorem 4.2.5]).

Theorem 9.5. *Let $W \in \mathscr{C}([-1,1])$ be even with $W(0) = 1$. Then, for any $\gamma > 0$, we have the convergence*

$$\Phi_{\gamma,W}(L) \xrightarrow{L\to\infty} 0.$$

This in combination with Theorem 9.4 leads to the convergence of the FBP method with respect to the $\mathcal{H}^\sigma$-norm for $0 \leq \sigma < \alpha$ (cf. [18, Corollary 4.2.6]).

Corollary 9.6. *Let $f \in L^1(\mathbb{R}^2) \cap \mathcal{H}^\alpha(\mathbb{R}^2)$ for $\alpha > 0$ and let $W \in L^\infty(\mathbb{R})$ be even with $supp(W) \subset [-1,1]$, such that W is continuous on $[-1,1]$ and $W(0) = 1$. Then, for $0 \leq \sigma < \alpha$, we have the convergence*

$$\|f - f_{W,L}\|_{H^\sigma(\mathbb{R}^2)} \xrightarrow{L\to\infty} 0.$$

It is even possible to extend the convergence result of Corollary 9.6 to the case $\sigma = \alpha$ and to the case that W is only continuous at zero, see [18, Theorem 4.2.7]. Besides these convergence results, Beckmann provided estimates on $\Phi_{\gamma,W}(L)$ for many classes of window functions, see [18, Section 4.3]. Moreover, [18, Section 4.1] gives an overview of related convergence results.

Discretization

For the implementation of the FBP method, we use the representation

$$f_{W,L} = \frac{1}{2} \cdot \mathcal{B}\left(q_L \star \mathcal{R}f\right)$$

from Theorem 9.1. Recall that we can only measure a finite number of Radon data values in practical cases. Hence, scanning schemes are required that lead to a suitable discretization of the involved operators. In our case, we use the *parallel beam geometry* (cf. [101, page 71]), where we assume that the values

$$\mathcal{R}f(j \cdot d,\ k \cdot \pi/N) \qquad j = -M, ..., M, \quad k = 0, ..., N-1 \tag{9.5}$$

are available for parameters $d > 0$ and $M, N \in \mathbb{N}$. This scheme corresponds to a scanning machine that rotates around the target object with stepwidth π/N and sends out $2M+1$ equidistant parallel beams at each considered angle, yielding a total number of $(2M+1) \cdot N$ X-rays (cf. Figure 9.2). Within this approach, we have to determine values for the step width and the number of collected data points.

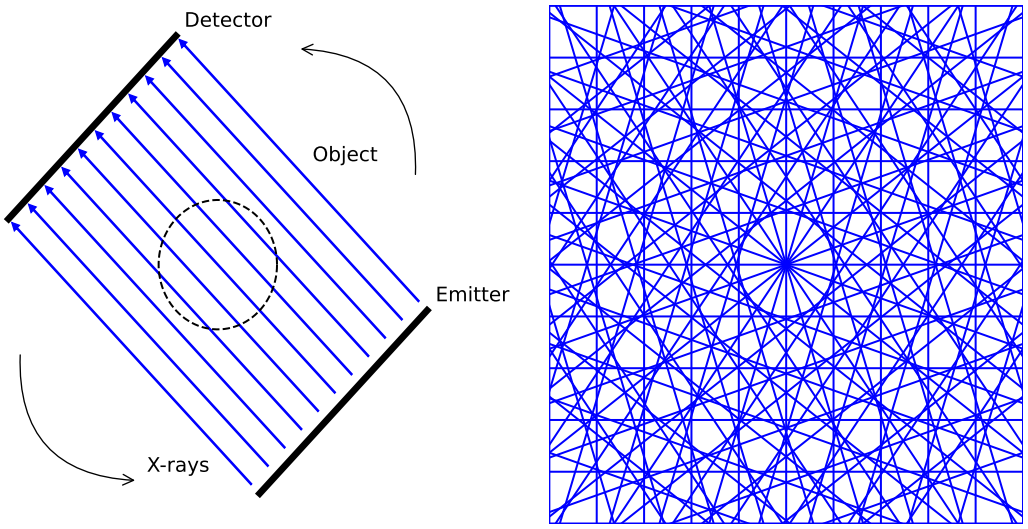

Figure 9.2.: Scanning procedure (left) and respective set of lines (right) for parallel beam geometry with $M = 5$, $N = 10$ and $d = 1/5$

Remark 9.7. We remark that there are other scanning geometries like the *fan beam geometry*, which bring along practical advantages in comparison to the parallel beam geometry. However, these advantages are not relevant to our numerical tests, so we do not cover these here but refer to [52, Section 8.12] and [101, Section III.3] for further reading.

As a starting point, assume that the domain of the index j is unbounded. One possible discrete version of the continuous convolution $(q_L \star \mathcal{R}f)(\cdot, \theta)$ for fixed $\theta \in [0, \pi)$ is then given by

$$(q_L \star_D \mathcal{R}f)(l \cdot d, \theta) := d \cdot \sum_{j \in \mathbb{Z}} \mathcal{R}f(j \cdot d, \theta) \cdot q_L((l-j) \cdot d, \theta) \qquad \text{for } l \in \mathbb{Z}, \tag{9.6}$$

see [18, page 104]. Recall that the objects that we want to reconstruct in practice are bounded, so that the associated attenuation function f has compact support. Hence, we can find a radius $R > 0$ such that $\operatorname{supp}(f) \subset B_R(0)$, which implies $\mathcal{R}f(r, \theta) = 0$ if $|r| > R$ (cf. Theorem 8.6). Setting $M = \lfloor R/d \rfloor$, we can replace (9.6) with the finite sum

$$(q_L \star_D \mathcal{R}f)(l \cdot d, \theta) := d \cdot \sum_{j=-M}^{M} \mathcal{R}f(j \cdot d, \theta) \cdot q_L((l-j) \cdot d, \theta) \qquad \text{for } l \in \mathbb{Z}. \tag{9.7}$$

Note that the discrete convolution $q_L \star_D \mathcal{R}f$ is also evaluated at the rate d with respect to the radial variable. Due to the structure of the low-pass filter, we know that the one-dimensional Fourier transform of $(q_L \star \mathcal{R}f)(\cdot, \theta)$ has compact support for each $\theta \in [0, \pi)$, so that $(q_L \star \mathcal{R}f)(\cdot, \theta)$ is band-limited. Hence, we can make use of the *Shannon sampling theorem* (see, e.g., [101, Theorem III.1.1]) to derive a suitable sampling rate for the radial variable.

Theorem 9.8 (Shannon sampling theorem)**.** *Let $h \in L^2(\mathbb{R})$ and $L > 0$ such that $\operatorname{supp}(\mathcal{F}h) \subset [-L, L]$, where $\mathcal{F}$ denotes the Fourier transform on $L^2(\mathbb{R})$. Then we have*

$$h(t) = \sum_{l \in \mathbb{Z}} h\left(l \cdot \frac{\pi}{L}\right) \cdot \operatorname{sinc}(Lt - l\pi)$$

for almost every $t \in \mathbb{R}$, i.e. h can be reconstructed by the discrete set of values $\{h(l \cdot \pi/L) \mid l \in \mathbb{Z}\}$.

Remark 9.9. In Example 3.5, we have already derived that, under the assumptions of Theorem 9.8, the evaluation of a continuous function h is given by

$$h(t) = \int_{\mathbb{R}} h(s) \cdot \frac{\sin(L \cdot (s-t))}{\pi \cdot (s-t)}\, ds = \frac{L}{\pi} \cdot \int_{\mathbb{R}} h(s) \cdot \operatorname{sinc}(L \cdot (t-s))\, ds \qquad \text{for all } t \in \mathbb{R}.$$

Approximating this integral by a Riemannian sum with separation π/L yields

$$\begin{aligned} \frac{L}{\pi} \cdot \int_{\mathbb{R}} h(s) \cdot \operatorname{sinc}(L \cdot (t-s))\, ds &\approx \frac{L}{\pi} \cdot \sum_{l \in \mathbb{Z}} h\left(l \cdot \frac{\pi}{L}\right) \cdot \operatorname{sinc}(L \cdot (t - l \cdot \pi/L)) \cdot \frac{\pi}{L} \\ &= \sum_{l \in \mathbb{Z}} h\left(l \cdot \frac{\pi}{L}\right) \cdot \operatorname{sinc}(Lt - l\pi) \end{aligned}$$

for $t \in \mathbb{R}$. It is remarkable that Theorem 9.8 establishes equality in the previous calculation, yielding a discretized version of the reproduction property.

As a consequence of Theorem 9.8, we choose the sampling rate with respect to the radial variable as $d = \pi/L$ in our implementation, where L is the bandwidth of the selected low-pass filter.

Example 9.10. For the computation of the discrete convolution (9.7) with $d = \pi/L$, we need to determine the values

$$q_L\left(j \cdot \frac{\pi}{L}, \theta\right) = \mathcal{F}_1^{-1} A_L\left(j \cdot \frac{\pi}{L}\right) \qquad \text{for } j \in \mathbb{Z}, \theta \in [0, \pi)$$

for the selected low-pass filter. Therefore, we can use the formulas from Example 9.2:

(i) For the Ram-Lak filter, we have

$$\mathcal{F}_1^{-1} A_L\left(j \cdot \frac{\pi}{L}\right) = \begin{cases} \frac{L^2}{2\pi}, & \text{for } j = 0 \\ 0, & \text{for } j \neq 0 \text{ even} \\ -\frac{2L^2}{\pi^3 j^2} & \text{for } j \text{ odd.} \end{cases}$$

(ii) In the case of the Shepp-Logan filter, we get

$$\mathcal{F}_1^{-1} A_L\left(j \cdot \frac{\pi}{L}\right) = \frac{4L^2}{\pi^3 \cdot (1-4j^2)} \qquad \text{for all } j \in \mathbb{Z}.$$

To discretize the back projection, we use the *composite trapezoidal rule*

$$Bg(x) \approx \frac{1}{N} \cdot \sum_{k=0}^{N-1} g(x_1 \cdot \cos(k \cdot \pi/N) + x_2 \cdot \sin(k \cdot \pi/N),\ k \cdot \pi/N) \qquad \text{for } x \in \mathbb{R}^2,$$

see [18, page 105], [52, Definition 8.25]. For more details on numerical integration algorithms, we refer to [14, Chapter 5]. Inserting the discretized convolution (9.7) leads to the approximate reconstruction

$$f_{W,L}^{(D)}(x) = \frac{1}{2N} \cdot \sum_{k=1}^{N-1} (q_L \star_D \mathcal{R}f)(x_1 \cdot \cos(k \cdot \pi/N) + x_2 \cdot \sin(k \cdot \pi/N),\ k \cdot \pi/N) \qquad \text{for } x \in \mathbb{R}^2. \tag{9.8}$$

In order to get a good approximation quality within the FBP method, it is recommended to choose $N = \lceil L \rceil$ (cf. [101, page 84, Table III.1]) for target functions whose support is contained in the unit ball $B_1(0)$, so that we have

$$d = \frac{\pi}{L}, \qquad M = \lfloor R \cdot L/\pi \rfloor, \qquad N = \lceil R \cdot L \rceil \tag{9.9}$$

as the parameters of our implementation for a given radius $R > 0$. However, there remains one problem regarding the formula (9.8). For given angle $\theta \in [0, \pi)$, we only evaluate $(q_L \star_D \mathcal{R}f)(\cdot, \theta)$ on the grid

$$\{l \cdot d \mid l \in \mathbb{Z}\} \subset \mathbb{R}$$

according to (9.7). In general, the radial input in the integrand within the back projection formula is not an element of this grid, i.e.

$$x_1 \cdot \cos(\theta) + x_2 \cdot \sin(\theta) \notin \{l \cdot d \mid l \in \mathbb{Z}\}$$

for most $x \in \mathbb{R}^2$. To avoid this problem, we determine the value

$$(q_L \star_D \mathcal{R}f)(x_1 \cdot \cos(\theta) + x_2 \cdot \sin(\theta),\ \theta)$$

via cubic spline interpolation, see [18, page 107]. More details on spline interpolation can be found in [36]. If $x \in B_R(0)$ holds, we can use the Cauchy inequality to estimate

$$|x_1 \cdot \cos(\theta) + x_2 \cdot \sin(\theta)| = |\langle x, n_\theta \rangle_2| \leq \|x\|_2 < R,$$

which means that we only need the finite set of evaluations

$$\left\{ (q_L \star_D \mathcal{R}f)(l \cdot d, \theta) \;\middle|\; l = -\lceil R/d \rceil, ..., \lceil R/d \rceil \right\}$$

for the spline interpolation in this case. At last, we want to remark that the discretized FBP method can be implemented efficiently using appropriate vectorization.

9.2. Algebraic reconstruction methods

In contrast to the FBP method, algebraic reconstruction methods do not rely on an explicit reconstruction formula. The main idea is to discretize the involved operator from the start, resulting in a finite system of equations (cf. [52, Section 9.1]). Suppose that we are given finite Radon data

$$\left\{ \mathcal{R}f(r_i, \theta_i) \;\middle|\; i = 1, ..., M \right\}$$

of a target function $f : \mathbb{R}^2 \to \mathbb{R}$. The respective discretized version of the inverse problem (8.7) is then given by the system of equations

$$\mathcal{R}s(r_i, \theta_i) = \mathcal{R}f(r_i, \theta_i) \qquad \text{for } i = 1, ..., M, \tag{9.10}$$

where $s : \mathbb{R}^2 \to \mathbb{R}$ denotes the approximate reconstruction of f. Of course, we have to make some assumptions on s in order to analyze the system (9.10). As in the case of an interpolation problem, we fix a set of basis functions

$$B = \left\{ b_j : \mathbb{R}^2 \to \mathbb{R} \;\middle|\; j = 1, ..., N \right\}$$

such that the Radon transform is defined everywhere for each basis function and we assume the approximate reconstruction s to have the form

$$s = \sum_{j=1}^{N} c_j \cdot b_j \in \operatorname{span}_{\mathbb{R}}(B),$$

see [62]. Due to the linearity of the Radon transform, this leads to the linear system

$$\begin{pmatrix} \mathcal{R}b_1(r_1, \theta_1) & \dots & \mathcal{R}b_N(r_1, \theta_1) \\ \vdots & \ddots & \vdots \\ \mathcal{R}b_1(r_M, \theta_M) & \dots & \mathcal{R}b_N(r_M, \theta_M) \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_N \end{pmatrix} = \begin{pmatrix} \mathcal{R}f(r_1, \theta_1) \\ \vdots \\ \mathcal{R}f(r_M, \theta_M) \end{pmatrix}, \tag{9.11}$$

where the system matrix is not necessarily square. Before we discuss special realizations, we want to refer to [61, Chapter 11] and [78, Chapter 7] for further reading about algebraic reconstruction methods.

One specific version of this approach is the *algebraic reconstruction technique (ART)* (cf. [56]), which uses characteristic functions of pixel squares as basis functions. Assuming that the target object is contained in the rectangle $[-R_1, R_1] \times [-R_2, R_2]$ with radii $R_1, R_2 > 0$, we choose resolution parameters $N_1, N_2 \in \mathbb{N}$ and divide the domain into the $N := N_1 \cdot N_2$ disjoint, equally sized rectangles

$$E_{j,k} = \begin{cases} [-R_1 + (j-1) \cdot h_1, -R_1 + j \cdot h_1) \times [-R_2 + (k-1) \cdot h_2, -R_2 + k \cdot h_2)\,, & j < N_1,\ k < N_2 \\ [-R_1 + (N_1 - 1) \cdot h_1, R_1] \times [-R_2 + (k-1) \cdot h_2, -R_2 + k \cdot h_2)\,, & j = N_1,\ k < N_2 \\ [-R_1 + (j-1) \cdot h_1, -R_1 + j \cdot h_1) \times [-R_2 + (N_2 - 1) \cdot h_2, R_2]\,, & j < N_1,\ k = N_2 \\ [R_1 - h_1, R_1] \times [R_2 - h_2, R_2]\,, & j = N_1, k = N_2, \end{cases}$$

where $h_1 = 2R_1/N_1$ and $h_2 = 2R_2/h_2$. If s is a linear combination of the characteristic functions, i.e.

$$s = \sum_{j=1}^{N_1} \sum_{k=1}^{N_2} c_{j,k} \cdot \chi_{E_{j,k}},$$

then s is constant on each rectangle with function value $c_{i,j}$, so that the coefficients equal the pixel value of the reconstructed image. Note that the matrix entries

$$\mathcal{R}\chi_{E_{j,k}}(r_i, \theta_i) \qquad \text{for } 1 \le i \le M,\ 1 \le j \le N_1,\ 1 \le k \le N_2$$

can be computed with the formula from Example 8.5 (iii). Moreover, the resulting matrix is sparse, as each X-ray only intersects a small percentage of the total pixels.

Remark 9.11. In the following, we focus on the functions $\chi_{E_{j,k}}$, but the theoretical results also hold for other choices of basis functions. For example, alternative basis functions are given by so-called *blob functions*, which were introduced in [83], [84]. These basis functions are translations of a smooth radial symmetric function with compact support. The main intention of using smooth basis functions is to generate smoother reconstructions that are potentially more resistant to noise (cf. [61, Section 6.5]). Numerical examples can be found in [61, Section 11.5], [92]. We come back to this idea in the context of kernel-based image reconstruction, see Section 10.4.

Kaczmarz's method

In the typical setting of $N_1 = N_2 = 256$, the number of pixels highly exceeds the number of measured Radon values, so that the system (9.11) is heavily underdetermined. To find a solution to this rectangular system, ART uses Kaczmarz's method for linear systems, which was introduced in [75]. We explain the method in the following.

Consider a general linear system

$$A \cdot x = b, \qquad \text{where} \qquad A = \begin{pmatrix} - & a_1^T & - \\ & \vdots & \\ - & a_m^T & - \end{pmatrix} \in \mathbb{R}^{m \times n},\ x \in \mathbb{R}^n, b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m. \tag{9.12}$$

Given an initial guess $x_0 \in \mathbb{R}^n$ and a column index $i \in \{1, ..., m\}$, we update the initial guess such that the i-th equation of the linear system is satisfied, or in other words, we project x_0 onto the solution space

$$S_i := \left\{ x \in \mathbb{R}^n \;\middle|\; \langle a_i, x \rangle_2 = b_i \right\}$$

of the i-th equation. Note that S_i is an affine space in $\mathbb{R}^n$, and the projection of x_0 onto this space is given by

$$x_0^{(i)} = x_0 - \frac{\langle x_0, a_i \rangle_2 - b_i}{\langle a_i, a_i \rangle_2} \cdot a_i,$$

see e.g. [52, Proposition 9.3]. Here, we need to assume that all rows of A are non-zero, which is usually satisfied in the context of computerized tomography. Since we want to solve the whole system of equations, one iteration of Kaczmarz's method consists of applying the projection for each $i \in \{1, ..., m\}$ successively. So, assuming that we have performed $k \in \mathbb{N}_0$ steps of the method resulting in the vector $x_k \in \mathbb{R}^n$, we set

$$x_k^{(0)} = x_k$$

and iteratively compute

$$x_k^{(i)} = x_k^{(i-1)} - \frac{\langle x_k^{(i-1)}, a_i \rangle_2 - b_i}{\langle a_i, a_i \rangle_2} \cdot a_i \qquad \text{for } i = 1, ..., m. \tag{9.13}$$

Then $x_{k+1} = x_k^{(m)}$ gives the next approximate solution to the linear system. Within each update step, a component is only changed if the corresponding component of a_i is non-zero. In the context of ART, this means that the color value of a pixel only changes if the considered X-ray intersects with the pixel. Note that each update step only requires one row of the entire matrix, which can be computed efficiently. Hence, it can be beneficial to not store the entire matrix and compute the necessary row in each update step in order to improve the efficiency of the method. These types of algorithms are also referred to as *row-action methods* (cf. [61, Section 11.1]). If we keep track of the non-zero components, we can use the sparsity of each row to further improve the efficiency.

Under the previous assumptions on the system matrix A, Kaczmarz's method converges to the least squares solution of the linear system (cf. [134, Corollary 9]).

Theorem 9.12. *Let $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ with non-zero rows. Set*

$$S = \left\{ x \in \mathbb{R}^n \;\middle|\; \|b - A \cdot x\|_2 = \inf_{y \in \mathbb{R}^n} \|b - A \cdot y\|_2 \right\},$$

and define $x^* \in S$ as the unique element in S that satisfies

$$\|x^*\|_2 = \inf_{x \in S} \|x\|_2.$$

If $\mathcal{P}_{\ker(A)}$ denotes the orthogonal projection onto the kernel $\ker(A) \subset \mathbb{R}^n$ of A, then for every initial vector $x_0 \in \mathbb{R}^n$, the iteration (9.13) is convergent with

$$x_k \xrightarrow{k \to \infty} \mathcal{P}_{\ker(A)}(x_0) + x^*.$$

In particular, the method converges to x^* if $x_0 \in \ker(A)^\perp$.

Remark 9.13. One significant drawback of the standard Kaczmarz's method is that, when applied to computerized tomography within ART, the resulting reconstruction often suffers from salt-and-pepper-like artifacts (cf. [61, Section 11.5]), in particular when applied to noisy measurements. The update step (9.13) can be modified with a relaxation parameter $\kappa > 0$, i.e.

$$x_k^{(i)} = x_k^{(i-1)} - \kappa \cdot \frac{\langle x_k^{(i-1)}, a_i \rangle_2 - b_i}{\langle a_i, a_i \rangle_2} \cdot a_i \qquad \text{for } i = 1, ..., m. \tag{9.14}$$

Geometrical interpretations for different relaxation parameters are given in [61, Section 11.2]. For noisy data, it is recommended to choose $\kappa \ll 1$ at the cost of convergence speed, which is also referred to as *under-relaxation* (cf. [61, page 212]). If $0 < \kappa < 2$, the convergence result from Theorem 9.12 holds as well if the linear system is consistent, see [102, Theorem 5.1]. Moreover, the convergence of the standard Kaczmarz's method is rather slow. To improve the rate of convergence, a randomized version was introduced in [133]. In each iteration, this version randomly chooses a row index and performs the respective update step, where the probability of each index is proportional to the squared norm of the respective row vector. The randomization yields an exponential convergence rate for the expected error.

Related methods

As we have already pointed out, ART without a relaxation parameter is known for producing non-smooth reconstructions in the case of noisy measurements. In each step (9.13), the color values are corrected such that the corresponding equation of the system is satisfied. But, if the reference value contains noise, this update step has a negative effect on the reconstruction quality. Therefore, further methods based on the concept of ART were developed that incorporate the error for each X-ray equation into the update step in order to reduce the effects of noise. One approach was the *simultaneous iterative reconstruction technique (SIRT)* (cf. [55]), which was followed by the *simultaneous algebraic reconstruction technique (SART)* (cf. [9]).

Here, we want to focus on the modified iterative solver for the system (9.12) within SART. Given the current approximate solution $x_k \in \mathbb{R}^n$, $k \in \mathbb{N}_0$, the j-th component $x_{k,j}$ of x_k is updated via the formula

$$x_{k+1,j} = x_{k,j} - \frac{1}{v_j} \cdot \sum_{i=1}^{m} a_{i,j} \cdot \frac{\langle x_k, a_i \rangle_2 - b_i}{w_i} \qquad \text{for } j = 1, ..., n, \tag{9.15}$$

where $A = (a_{i,j})_{\substack{1 \le i \le m \\ 1 \le j \le n}}$ denote the matrix entries and

$$v_j := \sum_{i=1}^{m} a_{i,j} \quad \text{for } j = 1, ..., n, \qquad w_i := \sum_{j=1}^{n} a_{i,j} \quad \text{for } j = 1, ..., m \tag{9.16}$$

are the column and row sums of A. In most applications, the expressions (9.16) are non-zero, so that (9.15) is well-defined. Note that the update term is a weighted sum of the correction terms

$$\frac{\langle x_k, a_i \rangle_2 - b_i}{w_i}, \tag{9.17}$$

where $w_{i,i}$ represents the length of the i-th X-ray beam and $\langle x_k, a_i \rangle_2$ is the line integral of the current approximate reconstruction in the context of image reconstruction. Hence, the term (9.17) distributes the interpolation error uniformly over the i-th X-ray beam. By setting

$$V = \begin{pmatrix} v_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & v_n \end{pmatrix} \in \mathbb{R}^{n \times n} \qquad \text{and} \qquad W = \begin{pmatrix} w_1^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_m^{-1} \end{pmatrix} \in \mathbb{R}^{m \times m},$$

we can write the iteration (9.15) in terms of matrix vector multiplication. As in Remark 9.13, we can add a relaxation parameter (cf. [10]), which yields the iteration

$$x_{k+1} = x_k - \kappa \cdot V^{-1} A^T W \cdot (A \cdot x_k - b) \qquad \text{for } k \in \mathbb{N}_0. \tag{9.18}$$

For a positive definite matrix $B \in \mathbb{R}^{l\times l}$ with $l \in \mathbb{N}$, we define the respective inner product and resulting B-norm as

$$\langle x, y\rangle_B := x^T \cdot B \cdot y \qquad \text{and} \qquad \|x\|_B := \sqrt{\langle x, x\rangle_B} \qquad \text{for } x, y \in \mathbb{R}^l.$$

With this notation, we get the following convergence result (cf. [72]).

Theorem 9.14. *Let $\kappa \in (0, 2)$, $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$ satisfy the conditions*

(i) $a_{i,j} \geq 0$ *for* $i = 1, ..., m,\ j = 1, ..., n$,

(ii) $v_j \neq 0$ *for* $j = 1, ..., n$,

(iii) $w_i \neq 0$ *for* $i = 1, ..., m$.

Set

$$S = \Big\{ x \in \mathbb{R}^n \;\Big|\; \|b - A \cdot x\|_W = \inf_{y\in\mathbb{R}^n} \|b - A \cdot y\|_W \Big\},$$

and define $x^* \in S$ as the unique element in S that satisfies

$$\|x^*\|_V = \inf_{x\in S} \|x\|_V.$$

If $\mathcal{P}_{\ker(A)}$ denotes the orthogonal projection onto the kernel of A with respect to $\langle\cdot,\cdot\rangle_V$, then for every initial vector $x_0 \in \mathbb{R}^n$, the iteration (9.18) is convergent with

$$x_k \xrightarrow{k\to\infty} \mathcal{P}_{\ker(A)}(x_0) + x^*.$$

In particular, the method converges to x^* if $x_0 \in \ker(A)^\perp$, where the orthogonal complement is again taken with respect to the modified inner product.

Advantages and disadvantages

Finally, we want to highlight the main advantages and disadvantages of algebraic reconstruction methods in comparison to the FBP method from Section 9.1. Recall that the discretization of the FBP method is based on the discrete convolution, composite trapezoidal rule and linear spline interpolation, which can be implemented efficiently (cf. [61, Section 8.7]). With an appropriate choice of filter function and bandwidth, we get an accurate reconstruction due to the respective convergence rates. In contrast, the computational costs of algebraic reconstruction methods are very high, since they involve huge systems of linear equations for high-resolution images. To solve these linear systems, one usually makes use of iterative solvers such as Kaczmarz's method (9.13), which can suffer from slow or even non-convergence. Moreover, many parameters have to be determined within algebraic methods, e.g. the choice of basis functions, the initial value for the iterative solver or the number of iterations. The determination of suitable parameters for the reconstruction is very challenging, as there is no general selection rule available. Due to these challenges, transform methods such as the FBP method have been preferred in commercial scanners for a long time (cf. [21, Introduction], [29, Section 6.1-6.4], [52, Section 9.6]).

However, the FBP method relies on large, regularly distributed data sets. As we have already stated in the introduction to the second part of this thesis, there are situations where regular line distributions like the parallel beam geometry cannot be applied, e.g. due to angle limitations, or where we need to reduce the radiation dose significantly. In these settings, the accuracy of the FBP method is no longer given or it might not be applicable without further modifications. One of the main advantages of algebraic reconstruction methods is their flexibility, since they allow for all possible distributions of lines in the plane, even for scattered data sets (cf. [61, page 214]). It has been shown by numerous examples that algebraic methods are able to outperform the FBP method in certain restricted or low-dose settings, see e.g. [95]. Moreover, it is easy to incorporate prior (physical) knowledge into the reconstruction process when using algebraic methods, e.g. by incorporating a regularization term (see, e.g., [21, page 99 & 101]).

9.3. Other methods

To close this chapter, we provide some further examples of reconstruction methods with brief descriptions, which demonstrate the wide range of the research field concerning computerized tomography:

- **Statistical inversion:** Given an inverse problem $Ax = y$, where $A : X \to Y$ is a linear operator, the involved spaces X, Y are seen as probability spaces with probability distributions p_X, p_Y, and y is is seen as a realization of Y. Moreover, it is assumed that we know the *prior density* p_X and the conditional probability distribution

$$p_Y(y \mid x) \qquad \text{for } x \in X,$$

also referred to as the *likelihood function*. In this case, *Bayes' theorem* (cf. [77, Theorem 3.1]) states that the converse conditional probability distribution is given by

$$p_X(x \mid y) = \frac{p_X(x) \cdot p_Y(y \mid x)}{p_Y(y)} \qquad \text{for } x \in X,$$

where $p_Y(y)$ can be computed via

$$p_Y(y) = \int_X p_Y(y \mid x) \cdot p_X(x)\, dx.$$

With this *posterior probability distribution*, we can then determine an approximate solution to the inverse problem. One common approach is the *maximum a posteriori estimate (MAP)*, which determines the reconstruction via

$$x_{\text{MAP}} = \max_{x \in X} p_X(x \mid y).$$

When applying this method to computerized tomography, the input space X is typically simplified, e.g. by using the pixel basis from Section 9.2. The main challenge of statistical inversion methods is to find good models for the required probability distributions. For further reading, we refer to [77], [80], [125].

- **Interpolation with zonal kernels:** In [17], every parameter pair is identified as an element of the unit sphere $\mathbb{S}^2 = \{x \in \mathbb{R}^3 \mid \|x\|_2\}$ via the injective mapping

$$\psi : \mathbb{R} \times [0, \pi) \to \mathbb{S}^2,\ (r, \theta) \mapsto \frac{1}{\sqrt{1 + r^2}} \cdot (\cos(\theta), \sin(\theta), r)^T. \tag{9.19}$$

The image of this mapping is given by the back half of the unit sphere without the north pole $(0, 0, 1)^T$, and the respective set of antipodal points represents the same set of lines in the plane. Hence, the image is seen as a subset of the projective space

$$\mathbb{P}^2 \cong \mathbb{S}^2/\sim,$$

where $\sim$ is the equivalence relation that identifies antipodal points. In contrast to other reconstruction methods, this approach does not come with a reconstruction model. Instead, given scattered Radon data, standard interpolation on $\mathbb{P}^2$ with *zonal kernels*, i.e. kernels on manifolds, is used to approximate the required Radon data of the target function to apply other reconstruction methods like the FBP method. For the interpolation on $\mathbb{P}^2$, a suitable metric on this manifold is required. We discuss further details in Section 10.3.

- **Neural networks:** Lastly, artificial neural networks have also found their way into the research field of tomography. In [73], the authors propose a method that combines the FBP method from Section 9.1 with a *convolutional neural network*, called *FBPConvNet*. The idea is that, given a limited amount of measurements, the FBP method computes an initial approximate solution which is then refined by the convolutional network. Another example is given in [15], where the *deep image prior approach* is used. But this is just a small glimpse at the wide use of neural networks in computerized tomography, or more generally, in inverse problems.

10. Image Reconstruction with Weighted Kernel Functions

With the preparations from the first part of this thesis and Chapter 8 & 9, we are finally able to analyze and classify the kernel-based reconstruction proposed in [38]. In the introduction of this thesis, we have already mentioned that the theoretical part of this paper did not properly explain the well-posedness of the reconstruction method. To verify this claim, we list and explain the critical parts below:

- For the framework of kernel-based generalized interpolation, it is essential that the considered linear functionals are elements of the dual space. In particular, the boundedness of a functional $\lambda \in \mathcal{H}_K^*$ ensures that the induced basis function $\lambda^y K(\cdot, y)$ is a well-defined function and an element of the native space $\mathcal{H}_K$ (cf. Theorem 3.12). At best, the sections [38, Section 2-3] explain that all entries of the resulting interpolation matrix are finite when using weighted versions of common translation-invariant kernels. In our eyes, this indicates that the incorporation of a weight function results in the boundedness of the associated functionals, but it does not represent a sufficient proof. We provide a full proof in Proposition 10.15, where we derive suitable conditions on the weight function as well.
- As discussed in the introduction of Chapter 4, the generalized interpolation problem has a unique solution for any given data values if and only if the considered functionals are linearly independent. To justify the well-posedness of the reconstruction scheme, the authors refer to the paper [68] (cf. [38, page 11, line 13-14]), which does not make sense. The analysis regarding the linear independence in [68] is concerned with compactly supported distributions, whereas the line integral functionals are not compactly supported due to their unbounded integration domain. Thus, we cannot use the results of [68] in this case. We discuss the linear independence of these functionals in Subsection 10.2.1, where we recall the practical workaround from Remark 5.5 that makes the theoretical analysis of the linear independence almost irrelevant.

The unclear points from above give rise to a theoretical reinvestigation of the reconstruction method, which is the main focus of this chapter. Based on [38, Proposition 1], we explain that commonly used translation-invariant kernels are not suited for the interpolation of Radon data. Due to this incompatibility, the authors of [38] introduced the class of *weighted kernel functions*, which we analyze in Section 10.1. We explain the underlying concept and useful properties. Following our analysis of weighted kernels, we demonstrate their utility in the context of Radon data interpolation. We prove that the associated functionals are indeed elements of the native dual space for suitable choices of weight functions, discuss the linear independence of these functionals and show how we can apply the convergence results and greedy algorithms from Chapter 5 & 7. Regarding the selection of data points, we provide further approaches in Section 10.3 that adapt to the particular reconstruction problem, where we take into account the symmetry of the Radon transform. As an alternative, we briefly explain the non-symmetric version of this reconstruction method in Section 10.4. Thereby, we use translates of common translation-invariant kernels to model the reconstruction, as it is done in *Kansa's method* (cf. [79]).

Let us start by specifying the interpolation setting for the image reconstruction from (scattered) Radon data. In this case, the interpolation functionals are given by the line integral operators

$$\mathcal{R}_{r,\theta}(f) := \int_{\ell_{r,\theta}} f(x)\, dx \qquad \text{for } (r,\theta) \in \mathbb{R} \times [0,\pi)\,,\ f : \mathbb{R}^2 \to \mathbb{R}. \tag{10.1}$$

The functionals from (10.1) are called *Radon functionals* in the following discussion. It is clear that the Radon functionals are linear, but, in order to apply the generalized interpolation framework, the functionals also have to be bounded on a suitable native space. In the initial part of [38], it was shown that this approach is not compatible with common translation-invariant kernels. For this class of kernels, the double integrals with respect to a fixed line in the plane diverge (cf. [38, Proposition 1]).

Proposition 10.1. *Let $K = \Phi(\cdot - \cdot)$ be a translation-invariant kernel with $\Phi \in \mathscr{C}(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$ and $(r, \theta) \in \mathbb{R} \times [0, \pi)$ such that $\mathcal{R}\Phi(0, \theta) > 0$. Then we have*

$$\int_{\ell_{r,\theta}} \int_{\ell_{r,\theta}} K(x, y) \, dy \, dx = \infty$$

i.e. the respective double line integral diverges.

Proof. We rewrite the double integral as

$$\int_{\ell_{r,\theta}} \int_{\ell_{r,\theta}} K(x, y) \, dy \, dx = \int_{\mathbb{R}} \int_{\mathbb{R}} \Phi\left((t - s) \cdot v_\theta\right) \, ds \, dt = \int_{\mathbb{R}} \int_{\mathbb{R}} \Phi(s \cdot v_\theta) \, ds \, dt = \int_{\mathbb{R}} \mathcal{R}\Phi(0, \theta) \, dt$$

by using the mass-preserving diffeomorphism $T_t : \mathbb{R} \to \mathbb{R}, \; s \mapsto t - s$ for every $t \in \mathbb{R}$. This representation proves the assertion. □

Recall from Theorem 3.12 and Remark 3.13 that for any functional $\lambda \in \mathcal{H}_K^*$ from the dual space, its squared norm is given by

$$\|\lambda\|_K^2 = \langle \lambda, \lambda \rangle_K = \lambda \left(\lambda^y K(\cdot, y)\right). \tag{10.2}$$

In the case of a Radon functional $\mathcal{R}_{r,\theta}$, the right side of equation (10.2) coincides with the double integral from Proposition 10.1. But, under the assumptions of the proposition, the double integral diverges and cannot be the squared norm value of a functional from the dual space. Therefore, we can conclude that the respective Radon functional is not an element of the dual space.

Corollary 10.2. *In the setting of Proposition 10.1, we have $\mathcal{R}_{r,\theta} \notin \mathcal{H}_K^*$.*

The consequence of Corollary 10.2 is that we cannot naively use common kernels like the Gaussian kernel or the radial characteristic kernel (cf. Example 2.8). However, the authors of [38] also provided a workaround to fix this problem. We can modify standard kernels by multiplying them with a weight function to enforce desired properties. This construction yields a new class of kernels, called *weighted kernel functions* (cf. [38, Section 3]). Note that we could also try to find other non-translation-invariant kernels, but these are rather rare and the requirement that the Radon functionals have to be elements of the native dual space further complicates the search. Meanwhile, the theoretical analysis of weighted kernels is simplified due to the connections to the original kernel. By choosing appropriate weight functions, we can construct suitable kernels for the interpolation of Radon data quite easily, see Section 10.2.

Remark 10.3. The analysis of kernel-based image reconstruction from Radon data via generalized interpolation dates back to the master thesis of Sironi from 2011, see [126]. There, it was already shown that the double line integrals diverge for Gaussian kernels. In order to fix this problem, the outer line integral was replaced by a truncated or weighted version of the Radon functional. A summary of the theoretical and numerical results from the master thesis was published in [39]. One significant drawback of these two approaches is the missing symmetry in the interpolation matrices since only one application of the Radon functional is replaced by a truncated or weighted version. Hence, the reconstruction method does not fit perfectly in the setting of generalized interpolation, so we cannot use the tools from the first part of this thesis without further modifications.

10.1. Weighted kernel functions

In the paper [38], the respective authors proposed the use of weighted kernel functions to guarantee that the double integrals from Proposition 10.1 are finite. We want to collect relevant properties of weighted kernel functions here, which help us to show the well-posedness of Radon data interpolation with these types of kernels. Note that weighted kernel functions have already been analyzed in the master thesis [57], mainly in the context of standard interpolation. We build on that work and extend some of the theoretical results.

The idea of weighted kernel functions is to symmetrically apply weights to the kernel evaluation via multiplication.

Definition 10.4. Let $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a positive semi-definite kernel function and $w : \mathbb{R}^d \to \mathbb{R}$ be a function. The `weighted kernel function` $K_w : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ of K and w is defined as

$$K_w(x, y) := w(x) \cdot K(x, y) \cdot w(y) \qquad \text{for } x, y \in \mathbb{R}^d.$$

The function w is called the `weight function` of K_w.

In general, a weighted kernel is again positive semi-definite on $\mathbb{R}^d$. For interpolation purposes, we wish to guarantee that the weighted kernel K_w of a positive definite kernel K is again positive definite. The positive definiteness of the kernel is preserved if and only if the weight function w has no zeros on $\mathbb{R}^d$ (cf. [57, Theorem 3.1]).

Proposition 10.5. *Let K be positive semi-definite $\mathbb{R}^d$ and $w : \mathbb{R}^d \to \mathbb{R}$ be a function. Then K_w is again a positive semi-definite kernel. If K is positive definite, then K_w is positive definite if and only if $w(x) \neq 0$ holds for all $x \in \mathbb{R}^d$.*

Proof. It is clear that K_w is symmetric. Let $X = \{x_1, ..., x_n\} \subset \mathbb{R}^d$ be a finite subset and define the diagonal matrix

$$W_X := \begin{pmatrix} w(x_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w(x_n) \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

The matrix $A_{K_w,X}$ is then given by $A_{K_w,X} = W_X^T \cdot A_{K,X} \cdot W_X$, and therefore

$$c^T \cdot A_{K_w,X} \cdot c = (W_X \cdot c)^T \cdot A_{K,X} \cdot (W_X \cdot c) \geq 0 \qquad \text{for all } c \in \mathbb{R}^n,$$

since K is positive semi-definite. This proves that K_w is positive semi-definite as well.

For the second part, we first assume that K_w is positive definite. According to part (1) of Proposition 2.22, we have

$$0 < K_w(x, x) = w(x) \cdot K(x, x) \cdot w(x) = w(x)^2 \cdot K(x, x) \qquad \text{for all } x \in \mathbb{R}^d,$$

which implies $w(x) \neq 0$ for all $x \in \mathbb{R}^d$. Conversely, if $w(x) \neq 0$ holds for all $x \in \mathbb{R}^d$, we have

$$c^T \cdot A_{K_w,X} \cdot c = (W_X \cdot c)^T \cdot A_{K,X} \cdot (W_X \cdot c) > 0 \qquad \text{for all } c \in \mathbb{R}^n \setminus \{0\}$$

for all finite sets $X = \{x_1, ..., x_n\} \subset \mathbb{R}^d$ of pairwise distinct points, as K is positive definite and W_X is regular in this case. Hence, K_w is positive definite on $\mathbb{R}^d$. □

Example 10.6. In order to illustrate the effect of the weight function, we consider a Gaussian kernel

$$K(x, y) := e^{-\nu_1 \cdot \|x-y\|_2^2} \qquad \text{for } x, y \in \mathbb{R}^d$$

with shape parameter $\nu_1 > 0$ and a Gaussian weight function

$$w(x) := e^{-\nu_2 \cdot \|x\|_2^2} \qquad \text{for } x \in \mathbb{R}^d.$$

with parameter $\nu_2 > 0$. The kernel K and its weighted version

$$K_w(x, y) = e^{-\nu_2 \cdot \|x\|_2^2} \cdot e^{-\nu_1 \cdot \|x-y\|_2^2} \cdot e^{-\nu_2 \cdot \|y\|_2^2} \qquad \text{for } x, y \in \mathbb{R}^d$$

are visualised in Figure 10.1 for $d = 1$. In contrast to the original translation-invariant kernel K, the weighted kernel K_w is not constant on the diagonal, as the application of the weight function causes a fast decay. Note that the weighted kernel does not inherit the translation-invariance of the original kernel in general.

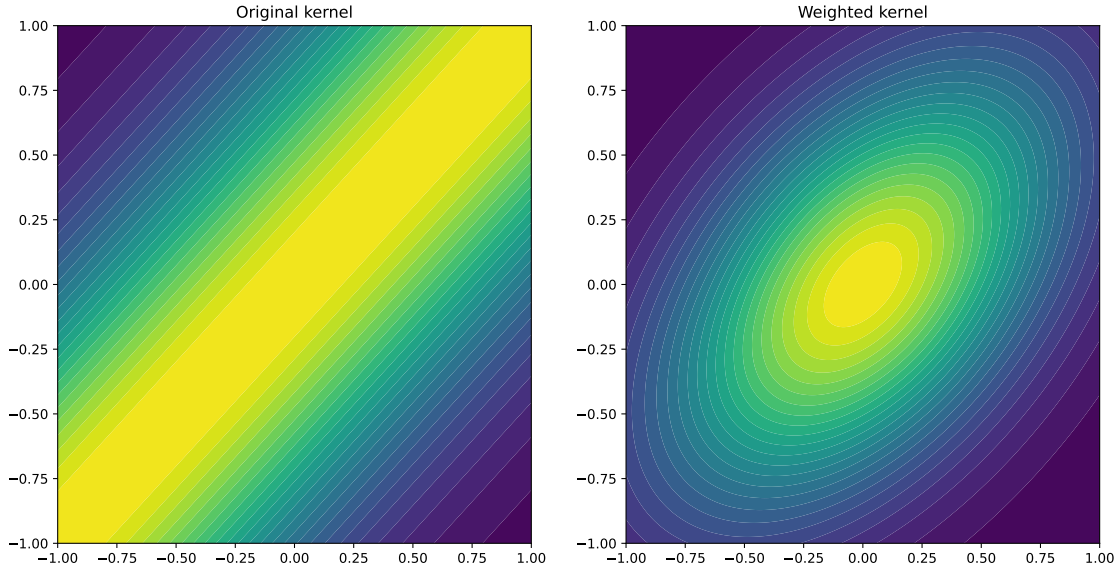

Figure 10.1.: Plot of Gaussian kernel on $\mathbb{R}\times\mathbb{R}$ with shape parameter $\nu_1 = 1$ (left) and plot of the weighted version with Gaussian weight function and additional shape parameter $\nu_2 = 1$ (right)

Due to Proposition 10.5, the weighted kernel K_w again generates a reproducing kernel Hilbert space $\mathcal{H}_{K_w}$ with K_w as its reproducing kernel. In the following, we want to investigate the connection between the reproducing kernel Hilbert space $\mathcal{H}_K$ of the original kernel K and $\mathcal{H}_{K_w}$. As an initial step, we take a look at the standard basis functions of the weighted kernel. For $x \in \mathbb{R}^d$, the standard basis function with respect to K_w is given by

$$K_w(\cdot, x) = w(\cdot) \cdot K(\cdot, x) \cdot w(x). \tag{10.3}$$

Note that the right part $K(\cdot, x) \cdot w(x)$ is the Riesz representer of the weighted Dirac functional

$$\delta_x^{(w)} := w(x) \cdot \delta_x \in \mathcal{H}_K^*,$$

and $\delta_x^{(w)}, \delta_x$ are multiples of each other if $w(x) \neq 0$. In the case that the weight function w does not have any zeros on $\mathbb{R}^d$, we can write the dense subset $\mathcal{S}_K$ from (3.3) as

$$\mathcal{S}_K = \operatorname{span}_{\mathbb{R}} \left\{ {\delta_x^{(w)}}^y K(\cdot, y) \;\middle|\; x \in \mathbb{R}^d \right\}.$$

With equation (10.3), it follows that the linear mapping

$$\Pi_w : \mathcal{S}_K \to \mathcal{S}_{K_w},\ f \mapsto w \cdot f \tag{10.4}$$

is surjective. Moreover, Π_w is isometric since we have

$$\langle K_w(\cdot, x), K_w(\cdot, \tilde{x}) \rangle_{K_w} = w(x) \cdot K(x, \tilde{x}) \cdot w(\tilde{x}) = \left\langle {\delta_x^{(w)}}^y K(\cdot, y), {\delta_{\tilde{x}}^{(w)}}^y K(\cdot, y) \right\rangle_K \qquad \text{for } x, \tilde{x} \in \mathbb{R}^d.$$

By continuous extension of Π_w, we can show that $\mathcal{H}_K$ and $\mathcal{H}_{K_w}$ are isometrically isomorphic. This result is a generalized version of [57, Theorem 3.4], where the assertion was proven for the special case of translation-invariant kernels. Moreover, the kernel K does not have to be positive definite.

Theorem 10.7. *Let K be positive semi-definite on $\mathbb{R}^d$ and $w : \mathbb{R}^d \to \mathbb{R}$ be a weight function such that $w(x) \neq 0$ for all $x \in \mathbb{R}$. Then, the mapping*

$$\Pi_w : \mathcal{H}_K \to \mathcal{H}_{K_w},\ f \mapsto w \cdot f$$

is an isometric isomorphism.

Proof. Consider the restricted mapping Π_w from (10.4). According to our previous discussion, Π_w is an isometric isomorphism between the pre-Hilbert spaces $\mathcal{S}_K$ and $\mathcal{S}_{K_w}$, which are dense in their respective native space. Therefore, Π_w can be extended continuously to an isometric isomorphism

$$\Pi : \mathcal{H}_K \to \mathcal{H}_{K_w}$$

that coincides with Π_w on $\mathcal{S}_K$, i.e.

$$\Pi(f) = \Pi_w(f) = w \cdot f \qquad \text{for all} f \in \mathcal{S}_K.$$

It remains to show that $\Pi_w(f) = w \cdot f$ holds for all $f \in \mathcal{H}_K$. So let $f \in \mathcal{H}_K$ be an arbitrary function. Since $\mathcal{S}_K \subset \mathcal{H}_K$ is dense, we can find a sequence $(s_n)_{n\in\mathbb{N}} \subset \mathcal{S}_K$ that converges normwise to f. Due to Proposition 3.6 part (5), this sequence also converges pointwise to f. Using the same argument and the continuity of Π, we can conclude that $(\Pi(s_n))_{n\in\mathbb{N}}$ converges normwise and pointwise to $\Pi(f)$. In total, we get

$$w(x) \cdot f(x) = \lim_{n\to\infty} w(x) \cdot s_n(x) = \lim_{n\to\infty} \Pi_w[s_n](x) = \lim_{n\to\infty} \Pi[s_n](x) = \Pi[f](x) \qquad \text{for all } x \in \mathbb{R}^d,$$

which proves the assertion. □

Corollary 10.8. *In the setting of Theorem 10.7, let $\tilde{w}$ be another weight function that satisfies $\tilde{w}(x) \neq 0$ for all $x \in \mathbb{R}^d$. Then, the spaces $\mathcal{H}_{K_w}$ and $\mathcal{H}_{K_{\tilde{w}}}$ are isometrically isomorphic via the mapping*

$$\Pi : \mathcal{H}_{K_w} \to \mathcal{H}_{K_{\tilde{w}}},\ f \mapsto \frac{\tilde{w}}{w} \cdot f.$$

As a consequence of Theorem 10.7, the dual mapping of Π_w is an isometric isomorphism as well. This means that we can shift the discussion about linear independence of functionals from $\mathcal{H}^*_{K_w}$ to the original dual space $\mathcal{H}^*_K$. We come back to this idea when we discuss the Radon data interpolation with weighted kernels in Section 10.2.

In addition, we can identify the native space of the weighted kernel as a subspace of a Sobolev space if K satisfies the assumptions of Theorem 3.19 and the weight function w is sufficiently smooth.

Corollary 10.9. *Let $K(\cdot,\cdot) = \Phi(\cdot - \cdot)$ be a translation-invariant kernel on $\mathbb{R}^2 \times \mathbb{R}^2$ that satisfies the assumptions of Corollary 3.19 for $a_1 > 1$. Moreover, let $w \in \mathcal{H}^{a_2}(\mathbb{R}^2)$ for $a_2 \geq 0$ with $w(x) \neq 0$ for all $x \in \mathbb{R}^2$. Then, for any $0 \leq \sigma \leq \min(a_1, a_2)$, the inclusion $\mathcal{H}_{K,w} \subset \mathcal{H}^\sigma(\mathbb{R}^2)$ holds and there is $C > 0$ such that*

$$\|f\|_{\mathcal{H}^\sigma(\mathbb{R}^2)} \leq C \cdot \|f\|_{K_w} \qquad \textit{for all } f \in \mathcal{H}_{K_w}.$$

Proof. Recall from Corollary 3.19 that $\mathcal{H}_K = \mathcal{H}^{a_1}(\mathbb{R}^2)$ holds, and there is $c_1 > 0$ such that

$$\|f\|_{\mathcal{H}^{a_1}(\mathbb{R}^2)} \leq c_1 \cdot \|f\|_K \qquad \text{for all } f \in \mathcal{H}_K. \tag{10.5}$$

Due to Theorem 10.7 and Proposition A.30, we have

$$\mathcal{H}_{K_w} = \Pi_w(\mathcal{H}_K) = \{w \cdot f \mid f \in \mathcal{H}^{a_1}(\mathbb{R}^2)\} \subset \mathcal{H}^\sigma(\mathbb{R}^2)$$

for $0 \leq \sigma \leq \min(a_1, a_2)$, including

$$\|\Pi_w(f)\|_{\mathcal{H}^\sigma(\mathbb{R}^2)} = \|w \cdot f\|_{\mathcal{H}^\sigma(\mathbb{R}^2)} \leq c_2 \cdot \|w\|_{\mathcal{H}^{a_2}(\mathbb{R}^2)} \cdot \|f\|_{\mathcal{H}^{a_1}(\mathbb{R}^2)} \qquad \text{for all } f \in \mathcal{H}_K \tag{10.6}$$

for a suitable constant $c_2 > 0$. Given an arbitrary $f \in \mathcal{H}_{K_w}$, there is $g \in \mathcal{H}_K$ with $f = \Pi_w(g)$, so that we can use (10.5), (10.6) and the isometry property of Π_w to conclude

$$\begin{aligned}\|f\|_{\mathcal{H}^\sigma(\mathbb{R}^2)} = \|\Pi_w(g)\|_{\mathcal{H}^\sigma(\mathbb{R}^2)} \leq c_2 \cdot \|w\|_{\mathcal{H}^{a_2}(\mathbb{R}^2)} \cdot \|g\|_{\mathcal{H}^{a_1}(\mathbb{R}^2)} &\leq c_1 c_2 \cdot \|w\|_{\mathcal{H}^{a_2}(\mathbb{R}^2)} \cdot \|g\|_K \\ &= c_1 c_2 \cdot \|w\|_{\mathcal{H}^{a_2}(\mathbb{R}^2)} \cdot \|f\|_{K_w}.\end{aligned}$$

This proves the assertion. □

Remark 10.10. The previous corollary gives some additional insights on the relation between the considered native spaces if $a_2 \geq a_1$. In this case, we can choose $\sigma = a_1$, so that the native space of the weighted kernel is a subspace of the original kernel's native space, and the norm of the weighted space is stronger than the one of the original space. This implies that any bounded linear functional on $(\mathcal{H}_K, \|\cdot\|_K)$ induces a bounded linear functional on $(\mathcal{H}_{K_w}, \|\cdot\|_{K_w})$, but the converse is not true in general. In Section 10.2, we derive suitable conditions on the weight function, which ensure that the Radon functionals from (10.1) are elements of the dual space $\mathcal{H}^*_{K_w}$, although this does not hold for standard translation-invariant kernels (cf. Corollary 10.2).

Conversion formulas for standard interpolation

To demonstrate useful properties of weighted kernels outside the application in computerized tomography, we derive some conversion formulas for tools like the power function or the Newton basis (cf. Chapter 5, 6) in the case of standard point evaluation functional interpolation. For the sake of simplicity, we set

$$\Lambda_X := \left\{ \delta_x \;\middle|\; x \in X \right\}$$

for every $X \subset \mathbb{R}^d$. In the context of interpolation, $X = \{x_1, ..., x_n\}$ is always a finite subset of pairwise distinct points and K is assumed to be positive definite so that Λ_X is linearly independent. We are interested in the relation between the interpolation spaces S_{K,Λ_X} and S_{K_w,Λ_X} as well as their respective Newton bases. As we show in the following proposition, the Newton basis elements of S_{K_w,Λ_X} can be attained by multiplication with the weight function w. This is again due to the mapping properties of the isometric isomorphism Π_w from Theorem 10.7.

Proposition 10.11. *In the setting of Theorem 10.7, let K be positive definite and $X = \{x_1, ..., x_n\} \subset \mathbb{R}^d$ be a finite subset of pairwise distinct points. Then Π_w is an isometric isomorphism between S_{K,Λ_X} and S_{K_w,Λ_X}. If the weight function w is positive on $\mathbb{R}^d$, i.e. $w(x) > 0$ for all $x \in \mathbb{R}^d$, the Newton basis $\mathcal{N}^{(w)}$ of S_{K_w,Λ_X} is given by*

$$\mathcal{N}^{(w)} = \left\{ w \cdot \mathfrak{n}_1, ..., w \cdot \mathfrak{n}_n \right\} = \left\{ \Pi_w(\mathfrak{n}_1), ..., \Pi_w(\mathfrak{n}_n) \right\},$$

where $\mathcal{N} = \{\mathfrak{n}_1, ..., \mathfrak{n}_n\}$ denotes the Newton basis of S_{K,Λ_X}.

Proof. Similar to the discussion before Theorem 10.7, we observe that

$$S_{K,\Lambda_X} = \operatorname{span}_{\mathbb{R}} \left\{ {\delta_{x_i}^{(w)}}^y K(\cdot, y) \;\middle|\; i = 1, ..., n \right\},$$

so that Π_w isometrically maps basis elements of S_{K,Λ_X} to basis elements of S_{K_w,Λ_X}, i.e.

$$\Pi_w \left({\delta_{x_i}^{(w)}}^y K(\cdot, y) \right) = \delta_{x_i}^y K_w(\cdot, y)$$

and

$$\left\langle {\delta_{x_i}^{(w)}}^y K(\cdot, y), {\delta_{x_j}^{(w)}}^y K(\cdot, y) \right\rangle_K = \left\langle \Pi_w \left({\delta_{x_i}^{(w)}}^y K(\cdot, y) \right), \Pi_w \left({\delta_{x_j}^{(w)}}^y K(\cdot, y) \right) \right\rangle_{K_w}$$

for $i, j = 1, ..., n$. Hence, the restriction of Π_w to S_{K,Λ_X} is an isometric isomorphism between the two interpolation spaces. For the second part, assume that w is positive on $\mathbb{R}^d$. If $\mathcal{N} = \{\mathfrak{n}_1, ..., \mathfrak{n}_n\}$ is the Newton basis of S_{K,Λ_X}, the image

$$\mathcal{B} = \left\{ \Pi_w(\mathfrak{n}_1), ..., \Pi_w(\mathfrak{n}_n) \right\}$$

of $\mathcal{N}$ under Π_w is an orthonormal basis in S_{K_w,Λ_X}. It remains to show that $\mathcal{B}$ is the Newton basis of S_{K_w,Λ_X}. Recall from Remark 6.7 that the evaluations of the Newton basis are given by the Cholesky factor of the interpolation matrix. If L is the Cholesky factor of $A_{K,\Lambda_X} = A_{K,X}$, we have

$$A_{K_w,\Lambda_X} = A_{K_w,X} = W_X \cdot L \cdot L^T \cdot W_X^T = (W_X \cdot L) \cdot (W_X \cdot L)^T,$$

i.e. the triangular matrix $L^{(w)} := W_X \cdot L$ is the Cholesky factor of A_{K_w,Λ_X} as it only contains positive entries on the diagonal (cf. [57, Theorem 3.3 (f)]). If $\mathcal{N}^{(w)} = \{\mathfrak{n}_1^{(w)}, ..., \mathfrak{n}_n^{(w)}\}$ denotes the Newton basis of S_{K_w,Λ_X}, this yields

$$\begin{pmatrix} \mathfrak{n}_1^{(w)}(x_1) & \dots & \mathfrak{n}_n^{(w)}(x_1) \\ \vdots & \ddots & \vdots \\ \mathfrak{n}_1^{(w)}(x_n) & \dots & \mathfrak{n}_n^{(w)}(x_n) \end{pmatrix} = L^{(w)} = W_X \cdot L = \begin{pmatrix} w(x_1) \cdot \mathfrak{n}_1(x_1) & \dots & w(x_1) \cdot \mathfrak{n}_n(x_1) \\ \vdots & \ddots & \vdots \\ w(x_n) \cdot \mathfrak{n}_1(x_n) & \dots & w(x_n) \cdot \mathfrak{n}_n(x_n) \end{pmatrix},$$

or equivalently,

$$\mathfrak{n}_i^{(w)}(x_j) = w(x_j) \cdot \mathfrak{n}_i(x_j) = \Pi_w[\mathfrak{n}_i](x_j) \qquad \text{for } i, j = 1, ..., n.$$

Since the elements in S_{K_w,Λ_X} are uniquely determined by their evaluation at X, we can conclude that $\mathfrak{n}_i^{(w)} = \Pi_w(\mathfrak{n}_i) = w \cdot \mathfrak{n}_i$ for every $i \in \{1, ..., n\}$ and therefore $\mathcal{N}^{(w)} = \mathcal{B}$. □

With the previous results, it is easy to derive a conversion formula for the power function as well (cf. [57, Theorem 3.3 (c)]).

Corollary 10.12. *Under the assumptions of Proposition 10.11, let P_{Λ_X} denote the power function with respect to the kernel K and $P_{\Lambda_X}^{(w)}$ denote the power function with respect to the weighted kernel K_w. In this case, we have the relation*

$$P_{\Lambda_X}^{(w)}(\delta_x) = |w(x)| \cdot P_{\Lambda_X}(\delta_x) \qquad \text{for all } x \in \mathbb{R}^d.$$

Proof. Let $\mathcal{B} = \{b_1, ..., b_n\}$ be an orthonormal basis of S_{K,Λ_X}. Then

$$\tilde{\mathcal{B}} = \Big\{\Pi_w(b_1), ..., \Pi_w(b_n)\Big\}$$

is an orthonormal system of S_{K_w,Λ_X} according to Proposition 10.11. With part (3) of Corollary 5.4, we get

$$P_{\Lambda_x}^{(w)}(\delta_x)^2 = K_w(x,x) - \sum_{i=1}^{n} \delta_x\left(\Pi_w(b_i)\right)^2 = w(x)^2 \cdot K(x,x) - \sum_{i=1}^{n} w(x)^2 \cdot b_i(x)^2 = w(x)^2 \cdot P_{\Lambda_X}(\delta_x)^2$$

for every $x \in \mathbb{R}^d$. Taking the square root on both sides leads to the desired conversion formula. □

The conversion formulas provide a huge computational advantage if the Newton basis and power functions have already been computed for the original kernel, as we only have to multiply with the evaluations of the weight function w. In particular, we can use the conversion formulas to easily switch between different weight functions for numerical comparisons, cf. Corollary 10.8. For the interpolant from $S_{K_w,X}$, a similar conversion formula does not exist. But, given function values $\delta_{x_1}(f), ..., \delta_{x_n}(f)$ of a function $f \in \mathcal{H}_{K_w}$, the interpolant can be computed efficiently with the update formula from Corollary 6.10, where we can use that $P_{X_i}(x_{i+1}) = \mathfrak{n}_{i+1}(x_{i+1})$ for $i = 1, ..., n-1$ and $X_i = \{x_1, ..., x_i\}$.

Nevertheless, we do not further concentrate on standard interpolation with weighted kernels. Numerical examples for the standard interpolation case can be found in [57, Chapter 4]. Our main intention for introducing weighted kernels is to ensure the well-definedness of integral operators and their associated discretized versions as elements of the dual space.

10.2. Radon data interpolation with weighted kernel functions

In Example 10.6, we have already seen which effect the weight function can have on the original kernel K. The properties of the weight function w can be used to enforce similar properties on the weighted kernel. For example, if the kernel K is bounded and the weight function is integrable, then K_w is integrable on the diagonal of its domain $\mathbb{R}^d \times \mathbb{R}^d$ so that the native space can be embedded into $L^1(\mathbb{R}^2)$.

Proposition 10.13. *Let $K \in \mathscr{C}(\mathbb{R}^2 \times \mathbb{R}^2)$ be a bounded positive semi-definite kernel and the weight function satisfy $w \in \mathscr{C}(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$. Then we have $K_w \in \mathscr{C}(\mathbb{R}^2 \times \mathbb{R}^2)$ and*

$$\int_{\mathbb{R}^d} K_w(x,x)^{1/2}\, dx < \infty,$$

so that $\mathcal{H}_{K_w} \subset \mathscr{C}(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$.

Proof. It is clear that $K_w \in \mathscr{C}(\mathbb{R}^2 \times \mathbb{R}^2)$ holds. Due to our assumptions, we can find $C > 0$ such that

$$K(x,x) \leq C \qquad \text{for all } x \in \mathbb{R}^d.$$

Hence, we can estimate

$$\int_{\mathbb{R}^d} K_w(x,x)^{1/2}\, dx = \int_{\mathbb{R}^d} |w(x)| \cdot K(x,x)^{1/2}\, dx \leq C^{1/2} \cdot \|w\|_{L^1(\mathbb{R}^2)} < \infty.$$

As a consequence of Proposition 3.14 part (2) and (3), we have $\mathcal{H}_{K_w} \subset \mathscr{C}(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$. □

Note that Proposition 3.14 also gives the norm estimate

$$\|f\|_{L^1(\mathbb{R}^d)} \le C^{1/2} \cdot \|w\|_{L^1(\mathbb{R}^d)} \cdot \|f\|_{K_w} \qquad \text{for all } f \in \mathcal{H}_{K_w}$$

for a suitable constant $C > 0$ under the assumptions of Lemma 10.13, so that the Radon transform

$$\mathcal{R} : (\mathcal{H}_{K_w}, \|\cdot\|_{K_w}) \to L^1(\mathbb{R} \times [0, \pi))$$

is a bounded linear operator, see Remark 3.16. But, this does not prove that the single Radon functionals from (10.1) are elements of the dual space. For this, we have to extend the assumptions on w minimally, such that $\mathcal{R}w(r, \theta)$ exists and is finite for the considered parameter pair $(r, \theta) \in \mathbb{R} \times [0, \pi)$, as this is not guaranteed by the condition $w \in L^1(\mathbb{R}^2)$. In the proof, we make use of the following lemma.

Lemma 10.14. *If K is a bounded positive semi-definite kernel on $\mathbb{R}^d$ and $w : \mathbb{R}^d \to \mathbb{R}$ is a function, there is $C > 0$ such that*

$$|f(x)| \le C^{1/2} \cdot |w(x)| \cdot \|f\|_{K_w} \qquad \textit{for all } f \in \mathcal{H}_{K_w}, x \in \mathbb{R}^d$$

and

$$\|K_w(\cdot, x) - K_w(\cdot, y)\|_{K_w} \le C^{1/2} \cdot (|w(x)| + |w(y)|) \qquad \textit{for all } x, y \in \mathbb{R}^d.$$

Proof. Let $C > 0$ satisfy $|K(x, y)| \le C$ for all $x, y \in \mathbb{R}^d$. For the first property, we estimate

$$|f(x)| = |\langle f, K_w(\cdot, x)\rangle_{K_w}| \le \|K_w(\cdot, x)\|_{K_w} \cdot \|f\|_{K_w} = |w(x)| \cdot K(x, x)^{1/2} \cdot \|f\|_{K_w} \le C^{1/2} \cdot |w(x)| \cdot \|f\|_{K_w}$$

for all $f \in \mathcal{H}_{K_w}$ and all $x \in \mathbb{R}^d$. The second property follows from

$$\begin{aligned} \|K_w(\cdot, x) - K_w(\cdot, y)\|_{K_w}^2 &= w(x)^2 \cdot K(x, x) - 2 \cdot w(x) \cdot K(x, y) \cdot w(y) + w(y)^2 \cdot K(y, y) \\ &\le |w(x)|^2 \cdot K(x, x) + 2 \cdot |w(x)| \cdot |K(x, y)| \cdot |w(y)| + |w(y)|^2 \cdot K(y, y) \\ &\le C \cdot (|w(x)| + |w(y)|)^2 \end{aligned}$$

for all $x, y \in \mathbb{R}^d$. □

Proposition 10.15. *In the setting of Proposition 10.13, where w does not have to be in $L^1(\mathbb{R}^2)$, let $(r, \theta) \in \mathbb{R} \times [0, \pi)$ satisfy $\mathcal{R}w(r, \theta) < \infty$. Then, the respective Radon functional from (10.1) is an element of the dual space, i.e. $\mathcal{R}_{r,\theta} \in \mathcal{H}_{K_w}^*$.*

Proof. Due to the properties of the Lebesgue integral, we also have

$$\mathcal{R}[|w|](r, \theta) = \int\limits_{\ell_{r,\theta}} |w(x)| \, dx < \infty.$$

With the first property from Lemma 10.14, we can find $C > 0$ such that

$$|\mathcal{R}_{r,\theta}(f)| = \left| \int\limits_{\ell_{r,\theta}} f(x) \, dx \right| \le \int\limits_{\ell_{r,\theta}} |f(x)| \, dx \le C^{1/2} \cdot \|f\|_{K_w} \cdot \int\limits_{\ell_{r,\theta}} |w(x)| \, dx = C^{1/2} \cdot \mathcal{R}[|w|](r, \theta) \cdot \|f\|_{K_w}$$

for all $f \in \mathcal{H}_{K_w}$, which proves that $\mathcal{R}_{r,\theta} \in \mathcal{H}_{K_w}^*$. □

Example 10.16. For the practicability of the reconstruction method, it is important that we can evaluate the Riesz representers and the inner products of the Radon functionals efficiently. Hence, it is desirable to develop simpler representations for the line integrals of certain weighted kernel functions, as numerical integration would slow down our reconstruction process immensely. We want to give one example here, which is the only known so far: Consider the weighted Gaussian kernel

$$K_w(x, y) = e^{-\nu_2 \cdot \|x\|_2^2} \cdot e^{-\nu_1 \cdot \|x - y\|_2^2} \cdot e^{-\nu_2 \cdot \|y\|_2^2} \qquad \text{for } x, y \in \mathbb{R}^2$$

with shape parameters $\nu_1, \nu_2 > 0$ from Example 10.6, which satisfies the assumptions of Proposition 10.15. The Riesz representers of the Radon functionals are given by

$$\mathcal{R}^y_{r,\theta} K_w(x,y) = \sqrt{\frac{\pi}{\nu_1+\nu_2}} \cdot e^{-\left[(\nu_1+\nu_2)\cdot\left(r^2+\|x\|_2^2\right)-2\nu_1 r\cdot\langle x,n_\theta\rangle_2-\frac{\nu_1^2}{\nu_1+\nu_2}\cdot\langle x,v_\theta\rangle^2\right]} \qquad \text{for all } x\in\mathbb{R}^2$$

and for all parameter pairs $(r,\theta) \in \mathbb{R}\times[0,\pi)$. Additionally, we have

$$\left\langle \mathcal{R}_{r_1,\theta_1}, \mathcal{R}_{r_2,\theta_2} \right\rangle_{K_w} = \frac{\pi}{\sqrt{q_{\nu_1,\nu_2}(\theta_1,\theta_2)}} \cdot e^{-\nu_2\cdot(2\nu_1+\nu_2)\cdot\frac{p_{\nu_1,\nu_2}(r_1,r_2,\theta_1,\theta_2)}{q_{\nu_1,\nu_2}(\theta_1,\theta_2)}}$$

for $(r_1,\theta_1),(r_2,\theta_2) \in \mathbb{R}\times[0,\pi)$, where

$$\begin{aligned} p_{\nu_1,\nu_2}(r_1,r_2,\theta_1,\theta_2) &= (\nu_1+\nu_2)\cdot\left(r_1^2+r_2^2\right) - 2\nu_1 r_1 r_2\cdot\cos(\theta_1-\theta_2) \\ q_{\nu_1,\nu_2}(\theta_1,\theta_2) &= (\nu_1+\nu_2)^2 - \nu_1^2\cdot\cos(\theta_1-\theta_2)^2. \end{aligned}$$

We do not prove these representations here and refer to [38, Section 4] for the computations. A visualization of the Riesz representers for this particular kernel is given in Figure 10.2. We can see that the functions reflect the directions of the considered lines.

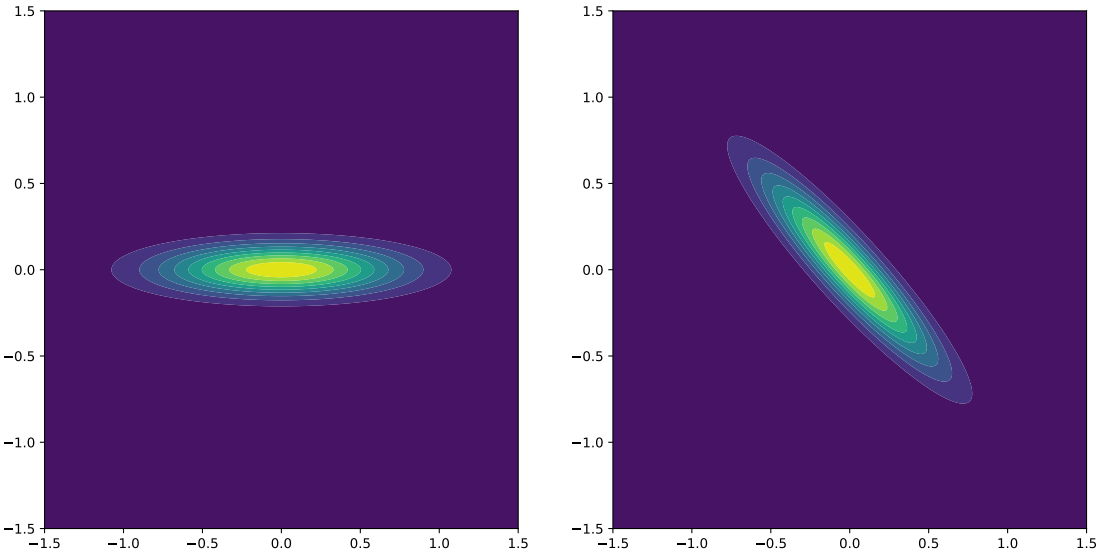

Figure 10.2.: Contour plot of the Riesz representers $\mathcal{R}^y_{0,\pi/2}K_w(\cdot,y)$ (left) and $\mathcal{R}^y_{0,\pi/4}K_w(\cdot,y)$ (right) for a weighted Gaussian kernel with shape parameters $\nu_1 = 50$ and $\nu_2 = 1$

Remark 10.17. So far, we have only used the continuity of the involved functions to ensure that each $f \in \mathcal{H}_{K_w}$ is measurable, which is also guaranteed if we assume that $K_w(\cdot,x)$ is measurable for all $x\in\mathbb{R}^d$, and to ensure that the mapping

$$\mathfrak{i}_{L^1(\mathbb{R}^2)} : \mathcal{H}_{K_w} \to L^1(\mathbb{R}^2),\ f \mapsto [f]_{L^1(\mathbb{R}^2)},$$

which maps a function from the native space to its function class in $L^1(\mathbb{R}^2)$, is injective. Due to the injectivity of $\mathfrak{i}_{L^1(\mathbb{R}^2)}$, the functions in $\mathcal{H}_{K_w}$ can also be distinguished in the L^1-sense, so that it can be seen as subspace of $L^1(\mathbb{R}^2)$. Note that this mapping does not always have to be injective. For example, consider the discontinuous kernel $K : \mathbb{R}^2\times\mathbb{R}^2 \to \mathbb{R}$

$$K(x,y) := \begin{cases} 1, & \text{if } x = y \\ 0, & \text{if } x \neq y. \end{cases}$$

The kernel K is bounded and positive definite on $\mathbb{R}^2$, but the standard basis function $K(\cdot,x)$ is equivalent to the zero function in $L^1(\mathbb{R}^2)$ for every $x\in\mathbb{R}^2$. It could be beneficial to investigate the construction of discontinuous kernels whose native space can be embedded into $L^1(\mathbb{R}^d)$ so that the theoretical results of this chapter can be extended to the discontinuous case.

Given finite Radon data $\mathcal{R}f(r_1,\theta_1),...,\mathcal{R}f(r_n,\theta_n)$ of a function $f:\mathbb{R}^2\to\mathbb{R}$, the proposed method from [38] generates an approximate reconstruction by applying the kernel-based generalized interpolation method with respect to the Radon functionals

$$\Lambda=\left\{\mathcal{R}_{r_1,\theta_1},...,\mathcal{R}_{r_n,\theta_n}\right\}\subset\mathcal{H}_{K_w}^*$$

with pairwise distinct parameter pairs

$$X=\left\{(r_1,\theta_1),...,(r_n,\theta_n)\right\}\subset\mathbb{R}\times[0,\pi).$$

Overall, the method can be classified as an algebraic reconstruction method, as we finitely discretize the Radon transform and solve the resulting generalized interpolation problem in the associated reproducing kernel Hilbert space. This means that it inherits the advantages mentioned at the end of Section 9.2, e.g. high flexibility regarding the scanning geometry. However, we want to point out two further advantages of the kernel-based approach in comparison to the usual concept of algebraic reconstruction methods:

- **Power function:** We can keep track of the power function after each update step, which gives an estimate for the pointwise interpolation error via (5.2) and is an indicator for the stability of the reconstruction method according to Section 7.1. Thus, the power function provides a tool for monitoring the reconstruction process. In particular, we can identify numerically linearly independent Radon functionals, or in other words, numerical redundancies in the given Radon data.
- **Greedy data selection algorithms:** The greedy algorithms from Chapter 7 can be applied in this case without further modifications. By applying a fixed greedy selection rule, we establish a hierarchy within the set of Radon functionals in terms of their relevance for the reconstruction process. To reduce the number of data points, we can only use a certain percentage of the initial data consisting of the most relevant Radon functionals. For the algebraic reconstruction technique from Section 9.2, it is known that a suitable reordering of the lines leads to an improved convergence speed (cf. [59], [63], [127]), as neighboring lines in the standard ordering might contain very similar information. The application of the greedy methods within the kernel-based approach can also be seen as a reordering algorithm for the data points in terms of their relevance, which is immediately available in the framework of kernel-based generalized interpolation.

Remark 10.18. In [102, Section 5.4], reconstruction methods that rely on the concept of generalized interpolation as discussed in the introduction of Chapter 4 are called *direct algebraic algorithms*. An example of such a method was already given in [28], where the Radon functionals were represented via the L^2 inner product with characteristic functions over X-ray beams, called *natural pixels*. However, it remains unclear whether the natural pixel approach can be extended by other models with higher regularity without losing the accurate representation of the Radon functionals so that it does not yield the same flexibility as the reconstruction via weighted kernel functions. In the kernel-based approach, we can simply switch between kernels with different regularities to change the approximation model.

10.2.1. Linear independence of Radon functionals

For the discussion about linear independence, we rely on the idea from Theorem 10.7 and the following lines, where we proved that, given a weight function $w:\mathbb{R}^d\to\mathbb{R}$ that has no zeros and a positive semi-definite kernel, the mapping

$$\Pi_w:\mathcal{H}_K\to\mathcal{H}_{K_w},\ f\mapsto w\cdot f$$

is an isometric isomorphism. Consider the mappings

$$\varphi:\mathcal{H}_K\to\mathcal{H}_K^*,\ f\mapsto\langle\cdot,f\rangle_K\qquad\text{and}\qquad\varphi_w:\mathcal{H}_{K_w}\to\mathcal{H}_{K_w}^*,\ f\mapsto\langle\cdot,f\rangle_{K_w}$$

and let Π_w^* denote the dual mapping of Π_w, i.e.

$$\Pi_w^*:\mathcal{H}_{K_w}^*\to\mathcal{H}_K^*,\ \lambda\mapsto\lambda\circ\Pi_w.$$

Then we obtain the commutative diagram in Figure 10.3, so that Π_w^* is an isometric isomorphism between the two dual spaces. For the interpolation with respect to Radon functionals $\Lambda = \{\mathcal{R}_{r_1,\theta_1}, ..., \mathcal{R}_{r_n,\theta_n}\}$, which lie in $\mathcal{H}_{K_w}^*$ under the assumptions of Proposition 10.15, the linear independence of Λ is equivalent to the linear independence of the functionals

$$\mathcal{R}_{r_i,\theta_i}(\Pi_w(f)) = \int\limits_{\ell_{r_i,\theta_i}} w(x) \cdot f(x)\, dx \qquad \text{for } f \in \mathcal{H}_K,\ i = 1, ..., n \tag{10.7}$$

in the dual space $\mathcal{H}_K^*$ of the original kernel's native space. Note that the kernel K is usually a well-known standard kernel so that the analysis in $\mathcal{H}_K^*$ is probably easier than the analysis in $\mathcal{H}_{K_w}^*$.

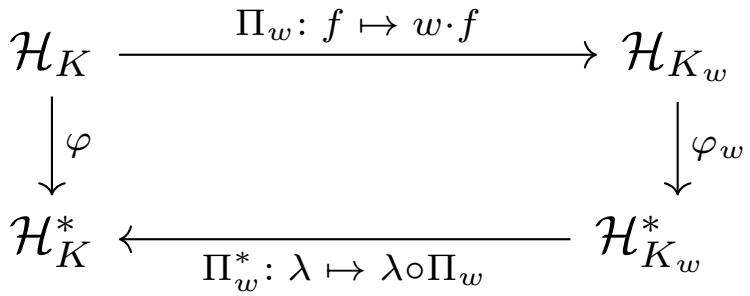

Figure 10.3.: Relation between the native spaces and their dual spaces

Here, we want to make use of the ideas from Section 4.1 for translation-invariant kernels and derive an integral representation of the quadratic form that is induced by the interpolation matrix $A_{K_w,\Lambda}$. To this end, we define the weighted Radon functionals as

$$\mathcal{R}_{r,\theta}^{(w)}(f) := \int\limits_{\ell_{r,\theta}} w(x) \cdot f(x)\, dx \qquad \text{for } f \in \mathcal{H}_K$$

for a given weight function and parameters $(r, \theta) \in \mathbb{R} \times [0, \pi)$ according to (10.7), and approximate these functionals again with suitable sequences in $S_K^{(*)}$.

Proposition 10.19. *Let $K = \Phi(\cdot - \cdot)$ be a translation-invariant positive definite kernel function with $\Phi \in \mathscr{C}(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$ and let $w \in \mathscr{C}(\mathbb{R}^2)$ such that $w(x) > 0$ for $x \in \mathbb{R}^2$ and its Radon transform $\mathcal{R}w$ is finite everywhere on $\mathbb{R} \times [0, \pi)$. Given a set of Radon functionals $\Lambda = \{\mathcal{R}_{r_1,\theta_1}, ..., \mathcal{R}_{r_n,\theta_n}\}$, we have the identity*

$$c^T \cdot A_{K_w,\Lambda} \cdot c = (2\pi)^{d/2} \cdot \int\limits_{\mathbb{R}^2} \Big| \sum_{k=1}^{n} c_k \cdot \int\limits_{\ell_{r_k,\theta_k}} w(x) \cdot e^{-i\cdot\langle x,\omega\rangle}\, dx \Big|^2 \cdot \mathcal{F}\Phi(\omega)\, d\omega \qquad \textit{for all } c \in \mathbb{R}^n. \tag{10.8}$$

Proof. We divide the proof into three steps:

(i) Given a parameter pair $(r, \theta) \in \mathbb{R} \times [0, \pi)$, our previous discussion shows that $\mathcal{R}_{r,\theta}^{(w)} \in \mathcal{H}_K^*$. In order to approximate the weighted Radon functional, we construct a sequence of functionals in the following way: For $n \in \mathbb{N}$, we can find $R_n > 0$ such that

$$\int\limits_{\mathbb{R}\setminus[-R_n,R_n]} w(x_s)\, ds < \frac{1}{n},$$

where we use the notation

$$x_s := r \cdot n_\theta + s \cdot v_\theta \qquad \text{for } s \in \mathbb{R}.$$

We can even assume that $R_n \geq 1$ for all $n \in \mathbb{N}$ and set

$$\varepsilon_n = \frac{1}{nR_n} \qquad \text{and} \qquad \tilde{R}_n = \sqrt{R_n^2 + r^2}.$$

Since w and the mapping $x \mapsto K(\cdot, x)$ are uniformly continuous on the closed ball $\overline{B_{\tilde{R}_n}(0)}$, there is $0 < \delta_n < 1/n$ such that

$$|w(x_s) - w(x_{\tilde{s}})| < \varepsilon_n \quad \text{and} \quad \|K(\cdot, x_s) - K(\cdot, x_{\tilde{s}})\|_K < \varepsilon_n \qquad \text{for } |s - \tilde{s}| < \delta_n.$$

Now choose $M_n \in \mathbb{N}$ sufficiently large such that

$$t_n := \frac{R_n}{M_n} < \delta_n \qquad \text{and set} \qquad s_j = (k + 1/2) \cdot t_n \quad \text{for } j \in \mathbb{Z}.$$

With this, we define a sequence of functionals via

$$\lambda_n = \sum_{j=-M_n}^{M_n-1} w(x_{s_j}) \cdot t_n \cdot \delta_{x_{s_j}} \in \mathcal{S}_K^{(*)} \qquad \text{for } n \in \mathbb{N}.$$

For any $f \in \mathcal{H}_K$, we can estimate

$$\left|\mathcal{R}_{r,\theta}^{(w)}(f) - \lambda_n(f)\right| \leq \int\limits_{\mathbb{R}\setminus[-R_n,R_n]} |w(x_s) \cdot f(x_s)| \, ds + \sum_{j=-M_n}^{M_n-1} \int\limits_{x_{s_j}-t_n/2}^{x_{s_j}+t_n/2} |w(x_s) \cdot f(x_s) - w(x_{s_j}) \cdot f(x_{s_j})| \, ds.$$

For the first summand, we can use the boundedness of $K = \Phi(\cdot - \cdot)$ (cf. Corollary 2.23) and the reproduction property in $\mathcal{H}_K$ to get

$$\begin{aligned}
\int\limits_{\mathbb{R}\setminus[-R_n,R_n]} |w(x_s) \cdot f(x_s)| \, ds &= \int\limits_{\mathbb{R}\setminus[-R_n,R_n]} w(x_s) \cdot |\langle f, K(\cdot, x_s)\rangle_K| \, ds \\
&\leq \|f\|_K \cdot \Phi(0)^{1/2} \cdot \int\limits_{\mathbb{R}\setminus[-R_n,R_n]} w(x_s) \, ds \\
&< \Phi(0)^{1/2} \cdot 1/n \cdot \|f\|_K.
\end{aligned}$$

For the second summand, we have

$$\begin{aligned}
\sum_{j=-M_n}^{M_n-1} \int\limits_{x_{s_j}-t_n/2}^{x_{s_j}+t_n/2} |w(x) \cdot f(x) - w(x_{s_j}) \cdot f(x_{s_j})| \, ds &\leq \sum_{j=-M_n}^{M_n-1} \int\limits_{x_{s_j}-t_n/2}^{x_{s_j}+t_n/2} |w(x_s) \cdot f(x_s) - w(x_s) \cdot f(x_{s_j})| \, ds \\
&\quad + \sum_{j=-M_n}^{M_n-1} \int\limits_{x_{s_j}-t_n/2}^{x_{s_j}+t_n/2} |w(x_s) \cdot f(x_{s_j}) - w(x_{s_j}) \cdot f(x_{s_j})| \, ds \\
&\leq \|f\|_K \cdot \varepsilon_n \cdot \int\limits_{-R_n}^{R_n} w(x_s) \, ds + \|f\|_K \cdot 2R_n \Phi(0)^{1/2} \cdot \varepsilon_n \\
&< \left(\mathcal{R}w(r,\theta) + 2\Phi(0)^{1/2}\right) \cdot 1/n \cdot \|f\|_K,
\end{aligned}$$

where we used that $R_n \geq 1$. In total, we have

$$\left\|\mathcal{R}_{r,\theta}^{(w)} - \lambda_n\right\|_K \leq \left(\mathcal{R}w(r,\theta) + 3\Phi(0)^{1/2}\right) \cdot 1/n \xrightarrow{n\to\infty} 0.$$

Regarding the Fourier-Laplace transform of these functionals (cf. Definition A.22), similar estimates show that the pointwise convergence

$$\mathcal{F}_{\mathscr{E}'}\lambda_n(\omega) \xrightarrow{n\to\infty} \int\limits_{\ell_{r,\theta}} w(x) \cdot e^{-i\cdot\langle x,\omega\rangle_2} \, dx \qquad \text{for all } \omega \in \mathbb{R}^2 \tag{10.9}$$

holds, where we make use of the estimate

$$\left|e^{-i\cdot\langle x,\omega\rangle_2} - e^{-i\cdot\langle y,\omega\rangle_2}\right| \leq |\langle x,\omega\rangle_2 - \langle y,\omega\rangle_2| \leq \|x - y\|_2 \cdot \|\omega\|_2 \qquad \text{for all } x, y, \omega \in \mathbb{R}^2$$

and the convergence $\delta_n \to 0$ for $n \to \infty$ by construction. Additionally, we have the uniform bound

$$\begin{aligned}|\mathcal{F}_{\mathscr{E}'}\lambda_n(\omega)| &\leq \sum_{j=-M_n}^{M_n-1} \int_{x_{s_j}-t_n/2}^{x_{s_j}+t_n/2} w(x_{s_j})\, ds \\ &\leq \sum_{j=-M_n}^{M_n-1} \int_{x_{s_j}-t_n/2}^{x_{s_j}+t_n/2} w(x_s)\, ds + \sum_{j=-M_n}^{M_n-1} \int_{x_{s_j}-t_n/2}^{x_{s_j}+t_n/2} |w(x_s) - w(x_{s_j})|\, ds \\ &\leq \mathcal{R}w(r,\theta) + 2R_n \cdot \varepsilon_n,\end{aligned}$$

which results in

$$|\mathcal{F}_{\mathscr{E}'}\lambda_n(\omega)| \leq \mathcal{R}w(r,\theta) + 2 \qquad \text{for all } n \in \mathbb{N}, \omega \in \mathbb{R}^2. \tag{10.10}$$

(ii) For two weighted Radon functionals $\lambda = \mathcal{R}^{(w)}_{r,\theta}, \mu = \mathcal{R}^{(w)}_{\tilde{r},\tilde{\theta}} \in \mathcal{H}^*_K$, we follow part (i) to construct the sequences

$$(\lambda_n)_{n\in\mathbb{N}}, \quad (\mu_n)_{n\in\mathbb{N}}$$

in $\mathcal{S}^{(*)}_K$ which converge to λ, μ. Due to the properties (10.9) and (10.10), we can imitate the steps in the proof of Lemma 4.4 to get

$$\left\langle \mathcal{R}^{(w)}_{r,\theta}, \mathcal{R}^{(w)}_{\tilde{r},\tilde{\theta}} \right\rangle_K = (2\pi)^{d/2} \cdot \int_{\mathbb{R}^2} \left(\overline{\int_{\ell_{r,\theta}} w(x) \cdot e^{-i\cdot\langle x,\omega\rangle_2}\, dx} \right) \cdot \left(\int_{\ell_{\tilde{r},\tilde{\theta}}} w(x) \cdot e^{-i\cdot\langle x,\omega\rangle_2}\, dx \right) \cdot \mathcal{F}\Phi(\omega)\, d\omega.$$

(iii) Set $\Lambda^{(w)} = \left\{ \mathcal{R}^{(w)}_{r_1,\theta_1}, ..., \mathcal{R}^{(w)}_{r_n,\theta_n} \right\} \subset \mathcal{H}^*_K$. By multilinear continuation of part (ii), we get

$$c^T \cdot A_{K_w,\Lambda} \cdot c = c^T \cdot A_{K,\Lambda^{(w)}} \cdot c = (2\pi)^{d/2} \cdot \int_{\mathbb{R}^2} \left| \sum_{k=1}^n c_k \cdot \int_{\ell_{r_k,\theta_k}} w(x) \cdot e^{-i\cdot\langle x,\omega\rangle}\, dx \right|^2 \cdot \mathcal{F}\Phi(\omega)\, d\omega$$

for each coefficient vector $c \in \mathbb{R}^n$, where we used that Π^*_w is an isometric isomorphism and

$$\mathcal{R}^{(w)}_{r_k,\theta_k} = \Pi^*_w(\mathcal{R}_{r_k,\theta_k}) \qquad \text{for } k = 1, ..., n.$$

□

Remark 10.20. If w is continuous and does not have any zeros, the *intermediate value theorem* implies that w is either strictly negative or strictly positive on the domain $\mathbb{R}^2$. Hence, the requirement that w is strictly positive does not lead to a loss of generality. The case that w is strictly negative can be handled in the same way.

Although we get the same integral representation for the Radon functionals, we cannot argue as in Section 4.1, since the weighted Radon functionals are not compactly supported for strictly positive weight functions $w \in \mathscr{C}(\mathbb{R}^2)$. Hence, we cannot apply the Fourier-Laplace transform on the whole complex space $\mathbb{C}^2$, and it is not guaranteed that the expressions inside the integral representations are holomorphic functions. Until now, it remains an open problem to find a suitable bridge between our integral representation and the theory of distributions. In the space of (tempered) distributions, it is again relatively easy to show that the finite set of Radon functionals is linearly independent for pairwise different parameter pairs since there is only a finite number of intersection points between the lines.

However, we recall that the power function can be evaluated numerically throughout the reconstruction process and is an indicator of linear dependence (cf. Remark 5.5). This means that we do not need a theoretical result that guarantees linear independence in practical cases where we keep track of the power function, which is highly recommended to monitor the numerical stability (cf. Section 7.1). According to Theorem 7.3, we already run into numerical trouble for small power function values, which makes a plain theoretical result on the linear independence even less relevant.

10.2.2. Convergence of the reconstruction method

For the convergence analysis, we restrict to the case that our data points are successively selected via the discussed greedy algorithm from Chapter 7. In order to apply the respective convergence results, we have to find suitable totally bounded subsets of Radon functionals. We follow the idea of Remark 7.7.

By construction of the Radon transform, the mapping

$$\varrho : \mathbb{R} \times [0, \pi) \to \mathcal{H}_{K_w}^*, \ (r, \theta) \mapsto \mathcal{R}_{r,\theta} \tag{10.11}$$

is a parametrization of the set of Radon functionals (cf. Section 5.3), if the assumptions of Proposition 10.15 are satisfied. Moreover, if we restrict the radius parameter $r \in \mathbb{R}$ to a compact interval, we can show that the resulting restriction of ϱ is uniformly continuous under suitable conditions on the weighted kernel.

Proposition 10.21. *Let $K \in \mathscr{C}(\mathbb{R}^2 \times \mathbb{R}^2)$ be a bounded positive semi-definite kernel and let $w \in \mathscr{C}(\mathbb{R}^2)$ satisfy the following condition: There is a function $g \in L^1(\mathbb{R})$, such that*

$$|w(r \cdot n_\theta + s \cdot v_\theta)| \leq g(s) \qquad \text{for all } (r, \theta) \in \mathbb{R} \times [0, \pi), s \in \mathbb{R}. \tag{10.12}$$

Then, for any $R > 0$, the restricted mapping

$$\varrho_R : [-R, R] \times [0, \pi) \to \mathcal{H}_{K_w}^*, \ (r, \theta) \mapsto \mathcal{R}_{r,\theta} \tag{10.13}$$

is well-defined and uniformly continuous.

Proof. According to Proposition 10.15, it is sufficient to show that $\mathcal{R}w(r, \theta) < \infty$ for all $(r, \theta) \in \mathbb{R} \times [0, \pi)$ in order to guarantee to well-definedness of ϱ_R for all $R > 0$. But this follows from (10.12), since

$$\mathcal{R}[|w|](r, \theta) = \int_{\mathbb{R}} |w(r \cdot n_\theta + s \cdot v_\theta)| \, ds \leq \int_{\mathbb{R}} g(s) \, ds < \infty \qquad \text{for all } (r, \theta) \in \mathbb{R} \times [0, \pi).$$

For the uniform continuity, let $R > 0$ and $\varepsilon > 0$. Our estimate is divided into several steps:

(i) With Lemma 10.14, we can find $C > 0$ such that

$$\|K_w(\cdot, x) - K_w(\cdot, y)\|_{K_w} \leq C^{1/2} \cdot (|w(x)| + |w(y)|) \qquad \text{for all } x, y \in \mathbb{R}^d.$$

(ii) Since $g \in L^1(\mathbb{R})$, we can find $\tilde{R} > 0$ such that

$$\int_{\mathbb{R} \setminus [-\tilde{R}, \tilde{R}]} g(t) \, dt < \frac{\varepsilon}{4C^{1/2}}.$$

(iii) Set $R' = \sqrt{R^2 + \tilde{R}^2}$. Due to the continuity of K_w, the mapping

$$x \mapsto K_w(\cdot, x)$$

is continuous on $\mathbb{R}^d$, and therefore uniformly continuous on the closed ball $\overline{B_{R'}(0)}$. Hence, we can find $\tilde{\delta} > 0$ such that

$$\|K_w(\cdot, x) - K_w(\cdot, y)\|_{K_w} < \frac{\varepsilon}{4\tilde{R}}$$

for all $x, y \in \overline{B_{R'}(0)}$ with $\|x - y\|_2 < \tilde{\delta}$.

(iv) Lastly, we can find $\delta > 0$ such that

$$\sup_{s \in [-\tilde{R}, \tilde{R}]} \|r_1 \cdot n_{\theta_1} + s \cdot v_{\theta_1} - (r_2 \cdot n_{\theta_2} + s \cdot v_{\theta_2}) \|_2 < \tilde{\delta}$$

for all $(r_1, \theta_1), (r_2, \theta_2) \in [-R, R] \times [0, \pi)$ with

$$\left\| (r_1, \theta_1)^T - (r_2, \theta_2)^T \right\|_2 < \delta.$$

This follows from the triangle inequality and the Lipschitz continuity of the mappings

$$\theta \mapsto n_\theta, \quad \theta \mapsto v_\theta.$$

Given $f \in \mathcal{H}_{K_w}$ and $(r_1, \theta_1), (r_2, \theta_2) \in [-R, R] \times [0, \pi)$, we can write

$$\begin{aligned}
|\mathcal{R}_{r_1,\theta_1}(f) - \mathcal{R}_{r_2,\theta_2}(f)| &= \left| \int_{\mathbb{R}} f(r_1 \cdot n_{\theta_1} + s \cdot v_{\theta_1})\, ds - \int_{\mathbb{R}} f(r_2 \cdot n_{\theta_2} + s \cdot v_{\theta_2})\, ds \right| \\
&\leq \int_{\mathbb{R}} |\langle f, K(\cdot, r_1 \cdot n_{\theta_1} + s \cdot v_{\theta_1}) - K(\cdot, r_2 \cdot n_{\theta_2} + s \cdot v_{\theta_2}) \rangle_{K_w}|\, ds \\
&\leq \|f\|_{K_w} \cdot \int_{\mathbb{R}} \|K_w(\cdot, r_1 \cdot n_{\theta_1} + s \cdot v_{\theta_1}) - K_w(\cdot, r_2 \cdot n_{\theta_2} + s \cdot v_{\theta_2})\|_{K_w}\, ds.
\end{aligned}$$

We split the integration domain into $\mathbb{R} = \mathbb{R} \setminus [-\tilde{R}, \tilde{R}] \cup [-\tilde{R}, \tilde{R}]$. The properties (i) and (ii) lead to

$$\begin{aligned}
&\int_{\mathbb{R}\setminus[-\tilde{R},\tilde{R}]} \|K_w(\cdot, r_1 \cdot n_{\theta_1} + s \cdot v_{\theta_1}) - K_w(\cdot, r_2 \cdot n_{\theta_2} + s \cdot v_{\theta_2})\|_{K_w}\, ds \\
&\leq C^{1/2} \cdot \int_{\mathbb{R}\setminus[-\tilde{R},\tilde{R}]} |w(r_1 \cdot n_{\theta_1} + s \cdot v_{\theta_1})| + |w(r_2 \cdot n_{\theta_2} + s \cdot v_{\theta_2})|\, ds \\
&\leq 2C^{1/2} \cdot \int_{\mathbb{R}\setminus[-\tilde{R},\tilde{R}]} g(s)\, ds \\
&< \frac{\varepsilon}{2},
\end{aligned}$$

and (iii) in combination with (iv) implies

$$\int_{-\tilde{R}}^{\tilde{R}} \|K_w(\cdot, r_1 \cdot n_{\theta_1} + s \cdot v_{\theta_1}) - K_w(\cdot, r_2 \cdot n_{\theta_2} + s \cdot v_{\theta_2})\|_{K_w}\, ds < 2\tilde{R} \cdot \frac{\varepsilon}{4\tilde{R}} = \frac{\varepsilon}{2}$$

in the case that

$$\left\| (r_1, \theta_1)^T - (r_2, \theta_2)^T \right\|_2 < \delta.$$

In total, we have

$$|\mathcal{R}_{r_1,\theta_1}(f) - \mathcal{R}r_2, \theta_2(f)| < \varepsilon \cdot \|f\|_{K_w},$$

and therefore

$$\|\varrho_R(r_1, \theta_1) - \varrho_R(r_2, \theta_2)\|_{K_w} = \|\mathcal{R}_{r_1,\theta_1} - \mathcal{R}_{r_2,\theta_2}\|_{K_w} \leq \varepsilon.$$

Thus, ϱ_R is uniformly continuous on $[-R, R] \times [0, \pi)$. □

The requirement (10.12) seems to be very restrictive at first sight, but it holds for commonly used radial symmetric functions like the Gaussian functions.

Corollary 10.22. *Let $w : \mathbb{R}^2 \to \mathbb{R}$ be radially symmetric, i.e. there is a function $\tilde{w} : [0, \infty) \to \mathbb{R}$ such that*

$$w(x) = \tilde{w}(\|x\|_2) \qquad \text{for all } x \in \mathbb{R}^2.$$

If $\tilde{w}$ is non-negative, monotonically decreasing and satisfies $\tilde{w} \in \mathscr{C}([0, \infty)) \cap L^1([0, \infty))$, then w satisfies the conditions of Proposition 10.21. In particular, for any bounded kernel $K \in \mathscr{C}(\mathbb{R}^2 \times \mathbb{R}^2)$ and $R > 0$, the mapping ϱ_R from (10.13) is uniformly continuous.

Proof. Define the function

$$g(s) := \tilde{w}(|s|) \qquad \text{for } s \in \mathbb{R}.$$

Then, we have

$$\int_{\mathbb{R}} g(s)\, ds = 2 \cdot \int_0^\infty \tilde{w}(s)\, ds < \infty$$

and

$$|w(r \cdot n_\theta + s \cdot v_\theta)| = \tilde{w}(\|r \cdot n_\theta + s \cdot v_\theta\|_2) = \tilde{w}\left(\sqrt{r^2 + s^2}\right) \leq \tilde{w}\left(\sqrt{s^2}\right) = g(s)$$

for all $(r, \theta) \in \mathbb{R} \times [0, \pi)\,, s \in \mathbb{R}$. □

Under the assumptions of Proposition 10.21, we can conclude that the restricted set

$$\Gamma_R := \operatorname{im}(\varrho_R) = \left\{ \mathcal{R}_{r,\theta} \;\middle|\; (r, \theta) \in [-R, R] \times [0, \pi) \right\}$$

of Radon functionals is totally bounded for any $R > 0$. Consequently, the greedy data selection algorithms from Chapter 7 lead to a convergent interpolation method, see Theorem 7.11, 7.15 and 7.19.

Theorem 10.23. *In the setting of Proposition 10.21, let $R > 0$ and $\Gamma = \Gamma_R$. If $(\Lambda_n)_{n \in \mathbb{N}}$ is chosen via the geometric greedy algorithm (7.4), the parameter space geometric greedy algorithm (7.6) or the β-greedy algorithm for fixed $\beta \in [0, \infty]$ (7.11), then we have*

$$\|f - I_{K_w, \Lambda_n}(f)\|_{K_w} \xrightarrow{n \to \infty} 0 \qquad \text{for all } f \in \mathcal{H}_{K_w, \Gamma_R}.$$

Corollary 10.24. *In the setting of Theorem 10.23, we can make the following additional conclusions:*

(1) If $w \in \mathscr{C}(\mathbb{R}^2)$ is bounded, we get convergence with respect to the supremum norm, i.e.

$$\|f - I_{K_w, \Lambda_n}(f)\|_\infty \xrightarrow{n \to \infty} 0 \qquad \text{for all } f \in \mathcal{H}_{K_w, \Gamma_R}.$$

(2) If $w \in \mathscr{C}(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$ is bounded, we get convergence with respect to the L^p-norm for every $p \in [1, \infty]$, i.e.

$$\|f - I_{K_w, \Lambda_n}(f)\|_{L^p(\mathbb{R}^2)} \xrightarrow{n \to \infty} 0 \qquad \text{for all } f \in \mathcal{H}_{K_w, \Gamma_R}.$$

(3) Under the assumptions of Proposition 10.9, we get convergence with respect to the Sobolev norm for any smoothness level $0 \leq \sigma \leq \min(a_1, a_2)$, i.e.

$$\|f - I_{K_w, \Lambda_n}(f)\|_{\mathcal{H}^\sigma(\mathbb{R}^2)} \xrightarrow{n \to \infty} 0 \qquad \text{for all } f \in \mathcal{H}_{K_w, \Gamma_R}.$$

Proof. The statements follow from Remark 3.7 and Proposition 3.14, 3.15, 10.9 & 10.13. □

Hence, under suitable assumptions on the kernel and the weight function, we get convergence towards the target function in terms of more common error measurements. To close out the discussion on convergence, we want to state two further ideas for future work.

Remark 10.25. As we have already mentioned in Remark 7.20, convergence rates for the β-greedy algorithms in the case of *linear elliptic differential operators of second order* and *Sobolev kernels* have been derived in [146]. In this preprint, the authors use the concept of *Kolmogorov n-widths*, which in our special kernel setting are defined as

$$d_n(\Gamma) := \inf_{\substack{S \subset \mathcal{H}^*_{K,\Gamma} \\ \dim(S) = n}} \sup_{\lambda \in \Gamma} \|\lambda - \mathcal{P}_S(\lambda)\|_K \qquad \text{for } n \in \mathbb{N}$$

for a given kernel K and a superset $\Gamma \subset \mathcal{H}^*_K$, where $\mathcal{P}_S$ is the orthogonal projection onto S. A general treatment of Kolmogorov n-widths can be found in [106]. According to [42] and [146], the decay rate of $d_n(\Gamma)$ translates to the power function, and therefore to the pointwise interpolation error. Hence, to adopt this approach in the context of Radon functionals, we would need to determine decay rates for $d_n(\Gamma_R)$ in dependence on the regularity of the weighted kernel K_w and fixed $R > 0$. Recall from Corollary 10.9 that, under suitable regularity assumptions on the kernel K and the weight function w, we can guarantee that $\mathcal{H}_{K_w}$ is contained in a Sobolev space, and make use of their well-known theory.

Remark 10.26. In our previous analysis, we have only proven the convergence of the interpolation method for certain subspaces of the native space. Given a fixed maximal radius $R > 0$ and a function $f \in \mathcal{H}_{K_w} \setminus \mathcal{H}_{K_w,\Gamma_R}$, we can consider the decomposition

$$\begin{aligned} f - I_{K,\Lambda_n}(f) &= f - \mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}(f) + \mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}(f) - I_{K,\Lambda_n}(f) \\ &= f - \mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}(f) + \mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}(f) - I_{K,\Lambda_n}\left(\mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}(f)\right) \end{aligned}$$

of the interpolation error, where $\mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}$ is the orthogonal projection onto the closed subspace $\mathcal{H}_{K_w,\Gamma_R}$ and $\Lambda_n \subset \Gamma_R$ is the current finite set of Radon functionals for the reconstruction of the target function. Note that we have used the identity

$$I_{K,\Lambda_n}\left(\mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}(f)\right) = \mathcal{P}_{S_{K,\Lambda_n}}\left(\mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}(f)\right) = \mathcal{P}_{S_{K,\Lambda_n}}(f) = I_{K,\Lambda_n}(f)$$

in the second line, which holds due to $S_{K,\Lambda_n} \subset \mathcal{H}_{H_w,\Gamma_R}$. The approximation error

$$\left\| \mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}(f) - I_{K,\Lambda_n}\left(\mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}(f)\right) \right\|_{K_w}$$

for the component in $\mathcal{H}_{K_w,\Gamma_R}$ has already been treated in Theorem 10.23. In order to get convergence results for the overall interpolation error, an investigation of the complementary operator

$$I - \mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}$$

in dependence of R could be beneficial, where I is the identity on $\mathcal{H}_{K_w}$. Recall that for any $f \in \mathcal{H}_{K_w}$, the property

$$f - \mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}(f) \in \mathcal{H}_{K_w,\Gamma_R}^{\perp}, \qquad \text{and therefore} \qquad \lambda\left(f - \mathcal{P}_{\mathcal{H}_{K_w,\Gamma_R}}(f)\right) = 0 \quad \text{for all } \lambda \in \Gamma_R$$

holds due to the generalized reproduction property from Theorem 3.12. Additionally, by increasing the parameter R, one has to reconsider the selection of the data sets Λ_n, as the superset Γ_R gets larger as well. We leave these ideas here for future work.

10.3. Problem-adapted point selection

So far, we have not taken into account the special properties of the Radon transform or the design of the scanning procedure into the greedy selection algorithms. In the following, we propose two adaptions that might improve the numerical performance in practical cases.

Periodic geometric greedy selection

Instead of restricting to parameter pairs from $\mathbb{R} \times [0, \pi)$, we can also extend the definition of the Radon transform of a target function $f \in \mathcal{H}_{K_w}$ on the whole plane $\mathbb{R}^2$, i.e.

$$\mathcal{R}f(r,\theta) = \int_{\ell_{r,\theta}} f(x)\, dx \qquad \text{for } (r,\theta) \in \mathbb{R}^2.$$

Due to the properties of the sine and cosine functions, we have the relation

$$\ell_{r,\theta} = \ell_{(-1)^j \cdot r, \theta + j\pi} \qquad \text{for all } (r,\theta) \in \mathbb{R}^2, j \in \mathbb{Z}, \tag{10.14}$$

which translates to the Radon transform. This identification, which can be associated with a Möbius strip of infinite width, raises some concern regarding the geometric greedy selection from (7.4), where we have relied on the standard Euclidean distance between two parameter pairs from $\mathbb{R} \times [0, \pi)$ in the previous section. As an example, consider the parameter pairs

$$(0, \varepsilon)\,,\ (0, \pi - \varepsilon) \in \mathbb{R} \times [0, \pi)$$

for small $\varepsilon > 0$. The standard Euclidean distance between these points is given by $\pi - 2\varepsilon$ and is therefore sufficiently large. However, the second line is also represented by $\ell_{0,-\varepsilon}$, yielding the small distance 2ε. Thus, naively using the standard norm might not reflect the actual distance between Radon functionals in the dual space. We give two examples of alternative metrics that take (10.14) into account.

(i) Consider the bijective, affine-linear mappings

$$\psi_j : \mathbb{R}^2 \to \mathbb{R}^2,\ (r,\theta) \mapsto \big((-1)^j \cdot r,\ \theta + j\pi\big) \qquad \text{for } j \in \mathbb{Z}$$

that preserve Euclidean distances between parameter pairs and satisfy $\psi_j \circ \psi_k = \psi_{j+k}$ for all $j,k \in \mathbb{Z}$. With the help of these functions, we define the equivalence relation

$$(r,\theta) \sim (\tilde{r},\tilde{\theta}) \quad :\Longleftrightarrow \quad \exists j \in \mathbb{Z} : \psi_j(r,\theta) = (\tilde{r},\tilde{\theta})$$

and the metric

$$d_{\mathbb{R}^2/\sim}\Big([(r,\theta)],[(\tilde{r},\tilde{\theta})]\Big) := \min_{j\in\mathbb{Z}} \|(r,\theta) - \psi_j(\tilde{r},\tilde{\theta})\|_2 \qquad \text{for } [(r,\theta)],[(\tilde{r},\tilde{\theta})] \in \mathbb{R}^2/\sim. \tag{10.15}$$

Note that each equivalence class has exactly one representative in $\mathbb{R}\times[0,\pi)$, so that (10.15) induces a metric on this restricted space. For the implementation, we make use of

$$\min_{j\in\mathbb{Z}} \|(r,\theta) - \psi_j(\tilde{r},\tilde{\theta})\|_2 = \min_{j\in\{-1,0,1\}} \|(r,\theta) - \psi_j(\tilde{r},\tilde{\theta})\|_2 \qquad \text{for } (r,\theta),(\tilde{r},\tilde{\theta}) \in \mathbb{R}\times[0,\pi)\,.$$

(ii) Recall from Section 9.3 that we can embed the parameter space $\mathbb{R}\times[0,\pi)$ into the projective space $\mathbb{P}^2 \cong \mathbb{S}^2/\sim$ via the mapping (9.19). In [17, Section 3], the respective authors consider the metric

$$d_{\mathbb{P}^2}([x],[y]) = \min\big(\cos^{-1}(\langle x,y\rangle_2),\ \cos^{-1}(\langle x,-y\rangle_2)\big) \qquad \text{for } x,y \in \mathbb{S}^2. \tag{10.16}$$

that reflects the symmetry of the Radon transform since antipodal points on the sphere represent the same line in the plane. In combination with (9.19), this yields a metric on $\mathbb{R}\times[0,\pi)$. However, given a maximum radius $R>0$, it is recommended in [17, Subsection 7.1] to use the scaled version

$$\psi_s : [-R,R]\times[0,\pi) \mapsto \mathbb{P}^2,\ (r,\theta) \mapsto \left[\frac{1}{\sqrt{1+(s\cdot r)^2}} \cdot (\cos(\theta),\sin(\theta),s\cdot r)^T\right] \tag{10.17}$$

for the identification in $\mathbb{P}^2$, where $s>0$ is a scaling parameter that should be chosen as a multiple of $1/R$, and then apply the metric (10.16). The idea of this scaling is to avoid distortion of average distances between lines within the ball $B_R(0)$.

Sequential scanning procedure

The measurement of the Radon data is usually a sequential process, which means that not all measurements are taken at the same time. An example was already given by the parallel beam geometry in Section 9.1, where the scanner rotates around an object with fixed angular separation and at each position emits a fixed amount of X-rays. To generalize this situation, let us assume that the set of Radon functionals Γ is split into the disjoint batches $\Gamma_j \subset \Gamma$, $j = 1,...,N$, and each of these batches becomes available at a different time when the scanner has performed measurements at the respective position. Assuming that we have measured the data for the batches Γ_j, $j \geq j_0$, where $1 \leq j_0 \leq N$, we can already start the greedy point selection on the subset

$$\Gamma^{(j_0)} = \bigcup_{j=1}^{j_0} \Gamma_j.$$

Throughout the process, we have to update the power function values and evaluations at the interpolant for the functionals from $\Gamma \setminus \Gamma^{(j_0)}$ as well. Once the measurements for batch Γ_{j_0+1} are available, we extend the greedy selection of the next functionals to the set $\Gamma^{(j_0)} \cup \Gamma_{j_0+1}$, where we can use the pre-computed information to efficiently evaluate the required error measurements for the selection rule.

In comparison to the previously discussed greedy algorithms, the reconstruction is available earlier, as we do not have to wait until the end of the scanning process to start with our computations. But, since we do not have access to the whole data set at all times, we expect that this approach leads to an inferior data point selection. It needs to be investigated whether there is a significant difference in the reconstruction quality between the sequential and non-sequential approaches.

10.4. Non-symmetric approach with standard kernels

Although Proposition 10.1 implies that the generalized interpolation approach with respect to Radon functionals is not compatible with most standard kernels, we can still develop an ART-like reconstruction method based on translation-invariant kernels. To this end, we consider a finite set $X = \{x_1, ..., x_n\} \subset \mathbb{R}^2$ and the respective set of basis functions

$$b_j := K(\cdot, x_j) = \Phi(\cdot - x_j) \qquad j = 1, ..., n$$

for a translation-invariant kernel $K(\cdot, \cdot) = \Phi(\cdot - \cdot)$. Due to the shifting property of the Radon transform (see, e.g., [18, Proposition 5.21]), the Radon transform of the basis functions can be computed via

$$\mathcal{R}b_j(r, \theta) = \mathcal{R}\Phi(r - \langle x_j, n_\theta \rangle_2, \theta) \qquad \text{for } (r, \theta) \in \mathbb{R} \times [0, \pi),\ j = 1, ..., n, \tag{10.18}$$

assuming that $\mathcal{R}\Phi$ exists everywhere. We can then proceed as in Section 9.2 to derive a reconstruction of the form

$$s = \sum_{j=1}^{n} c_j \cdot \Phi(\cdot - x_j)$$

from the given Radon data. This non-symmetric kernel-based approach follows the same idea as the blob function approach mentioned in Remark 9.11, as we can switch between kernels with different smoothness properties. But, we need to point out here that we do not have access to the tools of the generalized interpolation approach within this non-symmetric method. This makes the choice of the center points $X = \{x_1, ..., x_n\}$, and therefore the choice of basis functions, more difficult as we cannot rely on the discussed greedy algorithms. Moreover, we cannot use the power function to monitor the reconstruction process.

Example 10.27. As an example of this approach, we want to state the non-weighted version of Example 10.16. To this end, consider the Gaussian function

$$\Phi(x) = e^{-\nu \cdot \|x\|_2^2} \qquad \text{for } x \in \mathbb{R}^2$$

with shape parameter $\nu > 0$ that generates a positive definite Gaussian kernel function. According to Example 8.5 (i) and (10.18), the Radon transforms of the resulting basis functions are given by

$$\mathcal{R}\left[\Phi(\cdot - x)\right](r, \theta) = \sqrt{\frac{\pi}{\nu}} \cdot e^{-\nu \cdot (r - \langle x, n_\theta \rangle_2)^2} \qquad \text{for all } (r, \theta) \in \mathbb{R} \times [0, \pi), x \in \mathbb{R}^2.$$

11. Numerical Examples

In this chapter, we want to evaluate the numerical performance of the kernel-based reconstruction method that was analyzed in Section 10.2. To this end, we ran several numerical tests in *Python*[1]. The implementation of the considered algorithms can be found in the following *GitHub*[2] *Repository*:

`https://github.com/krischi10/KernelCT`

Regarding the implementation of the kernel-based method, we want to remark that the repository [113] served as a valuable inspiration. In Section 11.1, we list the test objects that were used in our numerical tests. These are *mathematical phantoms* whose Radon data can be computed efficiently for all radial and angular parameters. For the considered objects, we compare different greedy selection algorithms from Chapter 7 and Section 10.3 regarding their stability and approximation quality in Section 11.2. As proposed in Subsection 7.2.5, we test the algorithms in a data thinning setting. In Section 11.3, we compare the reconstruction via weighted kernels with the FBP method from Section 9.1. Our main focus is the number of data points required to obtain a reasonable reconstruction of the given objects. Lastly, we demonstrate the effect of regularization tools (cf. Section 7.3) on the reconstruction quality in Section 11.4.

11.1. Mathematical phantoms

In our numerical tests, the target objects were given by so-called *mathematical phantoms*, i.e. mathematical functions that represent simplified versions of real-world target objects, and whose point evaluations and line integral values can be computed efficiently. Thus, one can simulate the required measurements for any scenario and evaluate the reconstruction error. We describe the two considered phantoms in the following.

Shepp-Logan phantom

In Example 8.5 (ii), we have already described the Radon transform computation for the characteristic function of a shifted and rotated elliptical region. The *Shepp-Logan phantom* is a linear combination of such functions, each of them corresponding to an elliptical region with individual parameters. It was introduced in [124] and represents a highly simplified cross-section of a human head (cf. Figure 11.1). Due to the linearity of the Radon transform, we can evaluate its Radon transform, also referred to as its *sinogram*, very efficiently. For our tests, we used the high-contrast version of this phantom from [18, Subsection 5.2.2]. The parameters of the ellipses and the coefficients of the linear combination are listed in [18, page 118, Table 5.1].

Smooth phantom

Similar to the Shepp-Logan phantom, we can combine modifications of the bump-shaped functions f_ν from Example 8.5 (iv) for a fixed smoothness parameter $\nu \geq 0$. Since $f_\nu \in H^\sigma(\mathbb{R}^2)$ for all $\sigma < \nu + 1/2$, we can generate phantoms of arbitrary smoothness levels within this approach and potentially observe higher convergence rates in the reconstruction process (cf. [18, Section 5.3], [111, Section 6]). Setting $\nu = 0$ yields the same regularity as the Shepp-Logan phantom. In total, the *smooth phantom* is a linear combination of three rotated, shifted and stretched versions of f_ν, whose parameters and Radon transforms are provided in [18, Subsection 5.2.3]. We chose $\nu = 3$ in our numerical tests (cf. Figure 11.2).

[1] `https://www.python.org`
[2] `https://github.com`

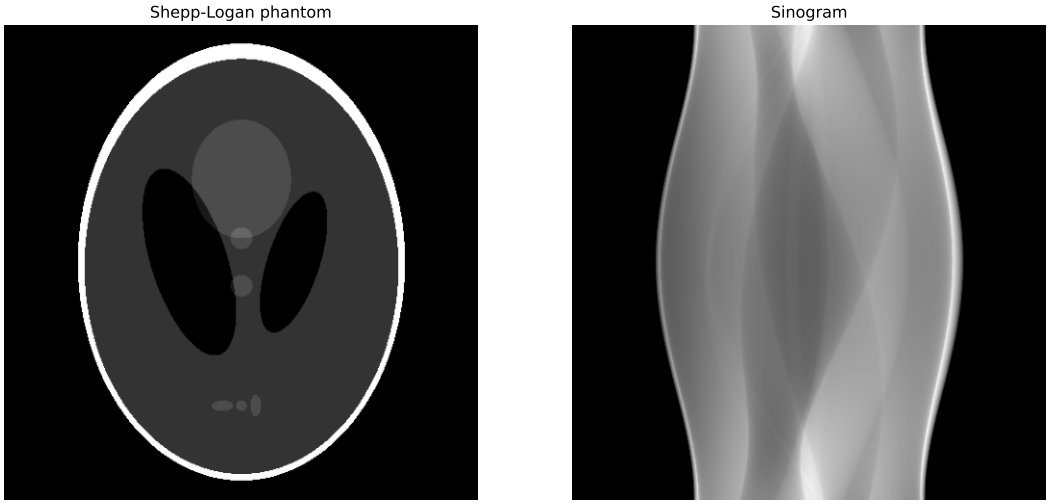

Figure 11.1.: Visualization of Shepp-Logan phantom (left) and its sinogram (right)

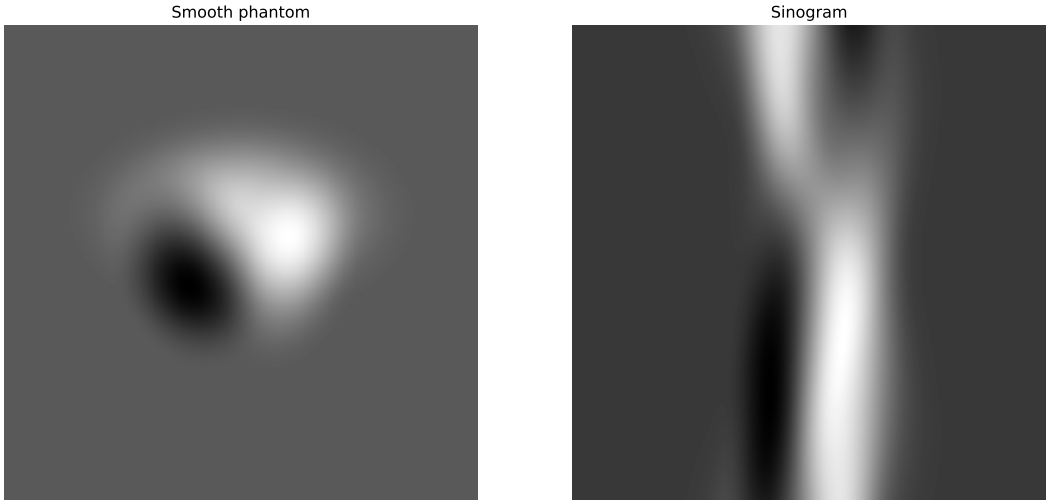

Figure 11.2.: Visualization of smooth phantom (left) and its sinogram (right) for $\nu = 3$

11.2. Comparison of greedy methods

As the first numerical investigation of the reconstruction method via kernel-based generalized interpolation, we compared different greedy selection algorithms in terms of accuracy and stability. To this end, we randomly generated 10^4 parameter pairs from $[-1, 1] \times [0, \pi)$, denoted by Ω_{rand}, which should result in a scattered parameter set. Consequently, the selection space of the greedy algorithms was given by the respective set of functionals

$$\Gamma = \varrho_1(\Omega_{\text{rand}}),$$

where ϱ_1 is the parametrizing mapping from (10.13) with $R = 1$. Note that the random number generator of the package *NumPy*[3], and therefore the precise outcome of the experiment, depends on the input seed. To ensure the reproducibility of our results, we state the selected seed for each of the described test runs. In this section, we only provide the results of one specific seed for each experiment. However, we verified the general trends within the results by varying the input seed.

In total, the greedy algorithms selected up to 2500 functionals. These functionals and their correspond-

[3] `https://numpy.org/doc/stable/reference/random/generator.html`, last checked: Tuesday 10th December, 2024

ing Radon values were then used to compute a reconstruction on the 512×512 pixel center grid

$$I_{\text{pixel}} = \left\{ -1 + \frac{1}{512} + k \cdot \left(\frac{2}{511} - \frac{2}{512 \cdot 511} \right) \;\middle|\; k \in \{0, ..., 511\} \right\}^2 \subset [-1, 1]^2 \tag{11.1}$$

via generalized interpolation. We measured the reconstruction accuracy with the *root mean squared error (RMSE)* defined as

$$\text{RMSE}(f, s) := \left(\frac{1}{512^2} \cdot \sum_{x \in I_{\text{pixel}}} (f(x) - s(x))^2 \right)^{1/2} \tag{11.2}$$

for two images $f, s : I_{\text{pixel}} \to \mathbb{R}$. Moreover, we measured the interpolation accuracy in the domain of the Radon transform via the error term

$$\left(\frac{1}{100^2} \cdot \sum_{(r,\theta) \in I_{\text{polar}}} (\mathcal{R}_{r,\theta}(f) - \mathcal{R}_{r,\theta}(s))^2 \right)^{1/2} \tag{11.3}$$

of $f, s : \mathbb{R}^2 \to \mathbb{R}$ on the 100×100 validation grid

$$I_{\text{polar}} = \left\{ -1 + j \cdot \frac{2}{99} \;\middle|\; j \in \{0, ..., 99\} \right\} \times \left\{ k \cdot \frac{\pi}{100} \;\middle|\; k \in \{0, ..., 99\} \right\} \subset [-1, 1] \times [0, \pi) \,.$$

Of course, f and s are set to be the original phantom and the generalized interpolant when computing the error measurements. In the description of the results, we refer to (11.2) as the *reconstruction error* and call (11.3) the *validation error*. For the evaluation of the stability, we computed the spectral condition numbers of the resulting interpolation matrices (cf. (7.1)).

As we have pointed out, the weighted Gaussian kernel from Example 10.16 is the only known kernel that yields an efficient computation of the Riesz representers and the interpolation matrix entries. Thus, we had to restrict our tests to this kernel model. Similar to the standard interpolation case, the involved shape parameters influence the accuracy and stability of the reconstruction method and an optimal choice depends on the target functions and the point selection strategies. For the test runs concerning the Shepp-Logan phantom, we set the shape parameters of the weighted Gaussian kernel to

$$\nu_1^{(SL)} = 2000, \qquad \nu_2^{(SL)} = 2. \tag{11.4}$$

The shape parameters regarding the smooth phantom were set to

$$\nu_1^{(SP)} = 700, \qquad \nu_2^{(SP)} = 3. \tag{11.5}$$

To avoid bias towards one of the greedy selection algorithms, we determined these parameters via *trial and error* on parallel beam geometry data (cf. (9.5)). Note that these parameters can still be optimized for one specific greedy algorithm, an example is discussed in Section 11.3. For better visibility, we group the test results by the different types of greedy algorithms.

Remark 11.1. Due to the inefficiency of the Newton basis (cf. 6.12), we did not use the point evaluations of the Newton basis elements for the point evaluation of the interpolant as suggested in Algorithm 1, lines 14 & 15. Instead, we made a change of basis and evaluated the interpolant in terms of the standard basis to compute the pixel values of the reconstruction image.

11.2.1. β-greedy algorithms

Regarding the β-greedy algorithms from Subsection 7.2.3, we tested the following versions and labels:

- `P-greedy`: This represents the choice $\beta = 0$, leading to the selection rule (7.3).
- `psr-greedy`: The choice $\beta = 1/2$ results in the psr-greedy algorithm as discussed in (7.10).

- `f-greedy`: The f-greedy selection rule (cf. (7.7)) represents the case $\beta = 1$.
- `beta2`: For the sake of completeness, we also tested the case $\beta = 2$, which is not associated with one of the algorithms from Subsection 7.2.2.
- `f/P-greedy`: The f/P-greedy selection rule from (7.9) can be interpreted as the limit case $\beta = \infty$ as discussed in the introduction of Subsection 7.2.3.

Shepp-Logan phantom

For the Shepp-Logan phantom, we set the seed of the random number generator to

$$\text{seed} = 2155654 \tag{11.6}$$

and used the weighted Gaussian kernel with the shape parameters from (11.4). The reconstructions after the final iteration in the case of the aforementioned β-greedy algorithms are visualized in Figure 11.3. While `psr-greedy` and `f-greedy` achieved the best reconstruction images, `beta2` and `f/P-greedy` were not able to reconstruct the smaller details of the phantom and `P-greedy` was not even able to reconstruct the outer oval shape of the phantom. In general, we can see many line artifacts in all reconstruction images, which diminish the visibility of the phantom at least to some degree.

The visual perception is also supported by Figure 11.4, where we plotted the reconstruction error, validation error and spectral conditions numbers in dependence on the number of chosen data points. `psr-greedy` and `f-greedy` achieved the lowest reconstruction and validation errors, especially in the early stage of the iterative selection. In the later stage, `psr-greedy` achieved the best error measurements. Regarding the numerical stability, it can be seen that the condition number significantly increases when increasing the parameter β of the greedy algorithm. The final measurements after selecting 2500 Radon functionals are provided in Table 11.1.

Method	Reconstruction error	Validation error	Condition number
`P-greedy`	1.74e-01	6.14e-02	3.00e+03
`psr-greedy`	7.63e-02	1.26e-02	7.20e+04
`f-greedy`	8.46e-02	1.54e-02	1.05e+06
`beta2`	8.53e-02	1.61e-02	2.04e+06
`f/P-greedy`	8.93e-02	2.21e-02	5.58e+07

Table 11.1.: Final measurements for Shepp-Logan phantom reconstruction via β-greedy methods

Smooth phantom

For the smooth phantom, we set the random number seed to

$$\text{seed} = 4578448 \tag{11.7}$$

and used the shape parameters from (11.5). In Figure 11.5, we can see that all considered β-greedy algorithms resulted in accurate reconstructions after the final iteration. The final measurements in Table 11.2 verify that the average reconstruction quality is much better than the average reconstruction quality in the test runs concerning the Shepp-Logan phantom. This indicates that the smooth phantom better matches the smooth weighted Gaussian model, leading to a faster decay of the reconstruction error. Similar to the Shepp-Logan reconstruction, Figure 11.6 indicates that a larger choice of β leads to an increase in the condition number and therefore decreases the numerical stability, where `f/P-greedy` has by far the worst condition number. However, `f-greedy`, `beta2` and `f/P-greedy` achieved better reconstruction and validation errors than `psr-greedy`. `P-greedy` again had by far the worst errors.

The previous two tests indicate that, for sufficiently smooth target functions, the parameter β can be fine-tuned to achieve an optimal tradeoff between approximation quality and numerical stability. In both tests, `psr-greedy` provided reasonable performances in accuracy and stability. `P-greedy` improved the numerical stability, but it came with a relatively low accuracy due to its independence of the target function. On the other side of the β-scale, `f/P-greedy` yielded relatively large condition numbers, but

Method	Reconstruction error	Validation error	Condition number
`P-greedy`	3.10e-04	1.35e-04	1.13e+05
`psr-greedy`	9.05e-05	2.42e-05	2.92e+05
`f-greedy`	6.43e-05	1.56e-05	5.20e+05
`beta2`	5.16e-05	1.25e-05	5.85e+06
`f/P-greedy`	6.97e-05	1.77e-05	8.26e+09

Table 11.2.: Final measurements for smooth phantom reconstruction via β-greedy methods

it also yielded low errors for the smooth phantom. We wish to remark that the stability issues have been addressed in the paper [144], but it needs to be investigated whether the stabilized versions of the β-greedy algorithms yield better reconstructions.

11.2.2. Geometric greedy algorithms

In addition, we tested the following geometric greedy algorithms:

- `P-greedy`: This is again the selection rule from (7.3). Although we already considered this algorithm in the previous two tests, it was also added to the geometric greedy tests to serve as a reference for better comparison.
- `geoDual`: The geometric greedy selection in terms of distances in the dual space is described by the selection rule (7.4). Recall that we can evaluate distances in the dual space via

$$\|\lambda_1 - \lambda_2\|_K^2 = \langle \lambda_1, \lambda_1 \rangle_K - 2 \cdot \langle \lambda_1, \lambda_2 \rangle_K + \langle \lambda_2, \lambda_2 \rangle_K \qquad \text{for } \lambda_1, \lambda_2 \in \mathcal{H}_{K_w}.$$

  For the evaluation of the inner product, we can use the formula from Example 10.16.
- `geoParam`: For the geometric greedy selection in the parameter space, which is given by the selection rule (7.6), we use the parametrizing map

$$\varrho_1 : \Omega_{\text{rand}} \to \Gamma,\ (r, \theta) \mapsto \lambda_{r,\theta}$$

  and equip the parameter space $\Omega_{\text{rand}} \subset [-1, 1] \times [0, \pi)$ with the standard Euclidean metric.
- `geoPeriodic`: We follow the approach of `geoParam`, but instead of using the standard Euclidean metric in Ω_{rand}, we consider the metric described in (10.15), which takes the periodicity of the Radon functionals into account.
- `geoSphere`: Again, we perform the thinning in the parameter space, but we use the metric (10.16) in combination with the embedding ψ_s into the projective space $\mathbb{P}^2$, see (10.17). In our tests, we set $s = 1.2$, which was determined by trial and error.

Shepp-Logan phantom

To ensure comparability, we used the same parameters as in the tests concerning the β-greedy algorithms, i.e. the shape parameters from (11.4) and the random number seed from (11.6). The reconstructions of the Shepp-Logan phantom after the final iteration are visualized in Figure 11.7. While `P-greedy` and `geoDual` were not even able to reconstruct the outer oval shape of the phantom, the methods `geoParam`, `geoPeriodic` and `geoSphere` resulted in reasonable reconstructions. Again, all images contain visible line artifacts. The error plots in Figure 11.8 verify the visual impression that `P-greedy` and `geoDual` provided significantly inferior reconstruction images. Regarding the numerical stability, the plot of the condition numbers in the same figure shows that `P-greedy` yielded the best condition numbers throughout the test run. In comparison to the target-dependent β-greedy algorithms, the geometric greedy algorithms yielded better condition numbers while achieving similar accuracy in the best cases, see Table 11.1 & 11.3.

Method	Reconstruction error	Validation error	Condition number
P-greedy	1.74e-01	6.14e-02	3.00e+03
geoDual	2.02e-01	8.47e-02	8.64e+03
geoParam	8.93e-02	1.59e-02	1.51e+04
geoPeriodic	9.09e-02	1.67e-02	1.55e+04
geoSphere	1.03e-01	2.10e-02	1.03e+04

Table 11.3.: Final measurements for Shepp-Logan phantom reconstruction via geometric greedy methods

Smooth phantom

Again, we used the parameters from the respective β-greedy test run, i.e. the shape parameters from (11.5) and the random number seed from (11.7), to ensure better comparability. Figure 11.9 shows that all methods were able to obtain accurate reconstructions of the smooth phantom, which is also supported by the final error measurements in Table 11.4. Similar to the test runs concerning the β-greedy algorithm, we can see that the average reconstruction quality is much better in the case of the smooth phantom than in the case of the Shepp-Logan phantom. The error measurements and condition numbers depending on the number of selected data points are plotted in Figure 11.10. As in the previous tests, `P-greedy` yielded the lowest condition numbers. Moreover, `P-greedy` and `geoSphere` yielded the best reconstruction errors. Comparing the results of Table 11.2 & 11.4, we see that the geometric greedy algorithms did not necessarily provide more stability, but yielded significantly worse reconstruction errors.

The two conducted tests concerning the geometric greedy algorithms do not show a clear tendency towards one of the selection strategies. `geoDual` yielded poor accuracy while not having the best condition numbers. `geoParam` and `geoPeriodic` yielded nearly the same accuracy results with `geoPeriodic` yielding lower condition numbers in the early stages of the iteration. Hence, the associated modification of the standard metric did not have a huge effect. `geoSphere` achieved a reasonable tradeoff between accuracy and stability in both test runs, but it was outperformed by `P-greedy` in the case of the smooth phantom.

Although the target-dependent β-greedy algorithms yielded significantly higher accuracy in the case of the smooth phantom, there remain major computational advantages in favor of the geometric algorithms. First, due to their independence of the target function, we can reuse the thinned data set for other target functions and thus save computation time in further tests. Second, the thinning in the parameter space is usually less costly as it does not require the computation of the inner product values in the dual space and the Newton basis. To end this section, we remark that our test results verify the applicability of the kernel-based reconstruction method to scattered Radon data.

Method	Reconstruction error	Validation error	Condition number
P-greedy	3.10e-04	1.35e-04	1.13e+05
geoDual	3.62e-03	2.29e-03	5.87e+06
geoParam	1.17e-03	3.19e-04	2.47e+06
geoPeriodic	1.22e-03	3.32e-04	1.98e+06
geoSphere	4.97e-04	1.26e-04	8.55e+05

Table 11.4.: Final measurements for smooth phantom reconstruction via geometric greedy methods

11.3. Comparison with FBP method

For the FBP method from Section 9.1, we used parallel beam geometry data as described in (9.5). Recall that the sampling parameters of this method are coupled via (9.9). Hence, for given angular sampling parameter $N \in \mathbb{N}$, the remaining parameters are given by

$$M = \lfloor N/\pi \rfloor \qquad L = N \qquad d = \frac{\pi}{L},$$

since the support of the considered phantoms is contained in $B_1(0)$. In total, this yields $N \cdot (2 \cdot \lfloor N/\pi \rfloor + 1)$ Radon lines. The parameters of the Radon lines were taken from $[-1, 1] \times [0, \pi)$ again. In our numerical

tests, we varied the number of angular samples N in low-dose settings, i.e. we let

$$N = 10,\ 20,\ ...,\ 120,$$

and kept track of the corresponding reconstruction error given by the RMSE (11.2).

Shepp-Logan phantom

For the reconstruction of the Shepp-Logan phantom, we equipped the `FBP method` with the Shepp-Logan filter due to the superior visual performance in comparison to the Ram-Lak filter (cf. Example 9.2). Regarding the kernel-based reconstruction, which we also refer to as `Kernel method`, we used the weighted Gaussian kernel with parameters from (11.4). Note that we did not use the Newton basis and thinning algorithms. Instead, we simply solved the interpolation equations via the linear system (4.3) and evaluated the interpolant via the Riesz representers of the functional.

The reconstructions for $N = 40, 80, 120$ are provided in Figure 11.11. In our eyes, the FBP method yielded slightly better visual results. However, the kernel-based method generated better reconstructions in terms of the RMSE, see Figure 11.12.

Smooth phantom

In the case of the smooth phantom, we equipped the FBP method with the Ram-Lak filter, which yielded better reconstruction errors than the Shepp-Logan filter. For the weighted Gaussian kernel, we used the shape parameters from (11.5) and proceeded as before. The reconstructions for $N = 40, 80, 120$ are given in Figure 11.13. For $N = 40$, which results in 1000 Radon lines, the FBP method provided a significantly better visual result than the kernel-based method. This is also supported by the plot of the reconstruction errors in Figure 11.14. For larger N, the kernel-based method achieved better reconstruction results than the FBP method in terms of the RMSE.

Data thinning on smooth phantom

To end this section, we want to demonstrate the effectiveness of proper data thinning in the context of computerized tomography. In Figure 11.14, we have seen that the best reconstruction of the smooth phantom via the FBP method was achieved at $N = 110$, which results in 7810 Radon lines. We will refer to the corresponding reconstruction error as `FBP reference` here. Following the procedure in Section 11.2, we applied the `psr-greedy` algorithm to the respective set of 7810 Radon functionals and tracked the reconstruction error depending on the number of the selected data points. As we have mentioned in the introduction of Section 11.2, optimal shape parameters for the weighted Gaussian kernel depend on the target function and the considered greedy algorithm. For the `psr-greedy` algorithm, we determined by trial and error that

$$\nu_1 = 300 \qquad \nu_2 = 5.5 \tag{11.8}$$

is a good choice, resulting in superior reconstructions compared to the previous shape parameters (11.5). As a further reference, we repeated the kernel-based reconstruction on parallel beam geometry data for the weighted Gaussian kernel with shape parameters from (11.8), which we refer to as `Kernel PBG`. To this end, we followed the procedure of the previous two tests of this section, but this time for the angular sample sizes

$$N = 5,\ 10,\ ...,\ 50.$$

The comparison results are visualized in Figure 11.15. Note that the error curve of the `psr-greedy` algorithm already crosses at approximately 700 selected Radon functionals, which is only a tenth of the initial data set. Moreover, we see that parallel beam geometry data is in general not optimal for the kernel-based approach, as the `psr-greedy` algorithm generated significantly better reconstructions in terms of the RMSE. This observation underlines the advantage of flexibility regarding the given data points, which is provided by weighted kernel functions.

In total, our numerical comparisons show that the kernel-based reconstruction method can compete with the FBP method, or even result in superior reconstructions in certain low-dose scenarios. However, we need to mention that the FBP method is far more efficient than the current implementation of the kernel-based method.

11.4. Regularization

In Section 11.2, the reconstructions of the Shepp-Logan phantom contained several line artifacts, which diminished the visual reconstruction quality. We want to investigate whether the regularization tools from Section 7.3 are able to reduce the number or the intensity of these artifacts. Therefore, we considered the thinned data set $\Lambda = \Lambda_{2500}$ that was selected by the `psr-greedy` algorithm in Subsection 11.2.1 for the Shepp-Logan phantom and applied regularization methods based on different penalizing functionals. Besides the reconstruction error (11.2), we also computed the *Structural Similarity (SSIM) Index*, which is implemented in the `skimage.metrics` submodule[4] in Python. The SSIM Index was introduced in [139] to improve the assessment of perceptual image quality. In total, we tested the following methods:

- `Interpolation`: This is just the reconstruction from Subsection 11.2.1 via generalized interpolation.
- `NormReg`: Given a regularization parameter $\gamma > 0$, we set $\Gamma = \Lambda$ and solve the minimization problem (7.16) that penalizes the native space norm.
- `TVReg`: For $\gamma > 0$, we solve the problem (7.17), where we again set $\Gamma = \Lambda$ and choose the set Ξ of regularization functionals as the partial differential operators

$$\Xi_{TV} = \left\{\delta_x \circ \partial_{x_1} \mid x \in I_{\text{pixel}}\right\} \cup \left\{\delta_x \circ \partial_{x_2} \mid x \in I_{\text{pixel}}\right\}$$

  at the grid points of I_{pixel} from (11.1). Since the weighted Gaussian kernel is smooth, we have the inclusion $\Xi_{TV} \subset \mathcal{H}^*_{K_w}$ (cf. Proposition 3.14 part (4)). This regularization method is well-known as *total variation (TV) regularization*.

The reconstructions for the different methods are provided in Figure 11.16. For `NormReg` and `TVReg`, we provide the results for two different regularization parameters to demonstrate that the parameter has to be chosen carefully. As in the tests before, the parameters were determined by trial and error. We can see that, for suitable choices of γ, both regularization methods were able to reduce the intensities of the line artifacts in comparison to `Interpolation`. However, we could not get rid of the line artifacts. For large regularization parameters, `NormReg` tends to the zero function and `TVReg` seems to result in a white blob.

Method	Regularization parameter γ	Reconstruction error	SSIM
`Interpolation`		7.63e-02	5.92e-01
`NormReg`	1.00e-06	8.87e-02	6.61e-01
`NormReg`	1.00e+00	2.47e-01	2.32e-01
`TVReg`	1.00e-10	9.61e-02	6.00e-01
`TVReg`	1.00e+00	2.47e-01	6.53e-02

Table 11.5.: Error measurements for Shepp-Logan phantom reconstruction via different regularization methods

The error measurements for the considered methods are provided in Table 11.5. Both regularization methods resulted in a worse reconstruction error, but `TVReg` achieved a similar SSIM index value as `Interpolation` while `NormReg` yielded the best SSIM index value. Note that the SSIM score should be as close to 1 as possible. Overall, we were able to achieve visual improvements, but we did not end up with a clear image of the Shepp-Logan phantom. This underlines the impression from Section 11.2 that the smooth weighted Gaussian model is not the right choice for this type of phantom.

[4] `https://scikit-image.org/docs/stable/api/skimage.metrics.html`, last checked: Tuesday 10th December, 2024

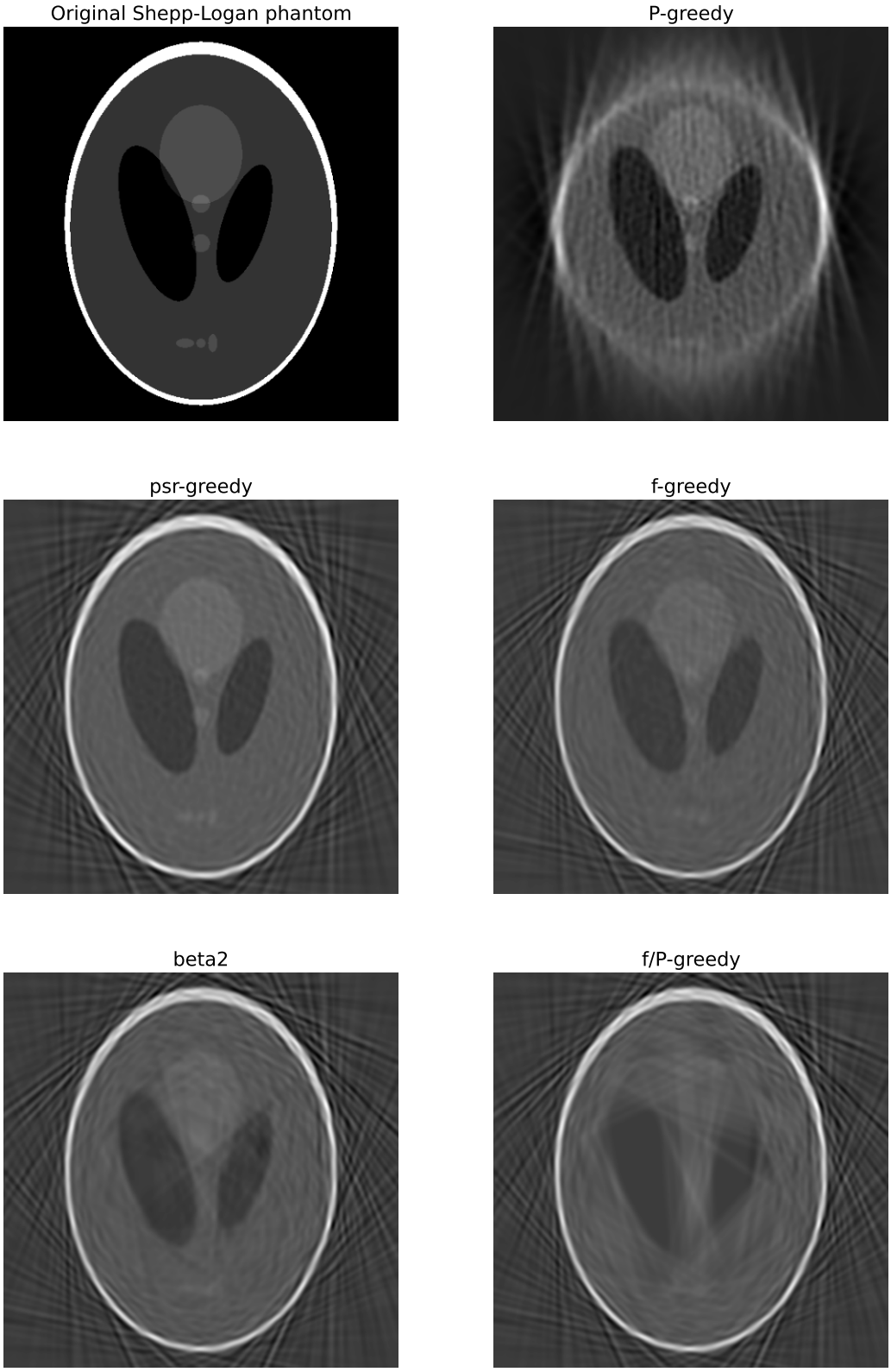

Figure 11.3.: Reconstructions of the Shepp-Logan phantom for different β-greedy algorithms

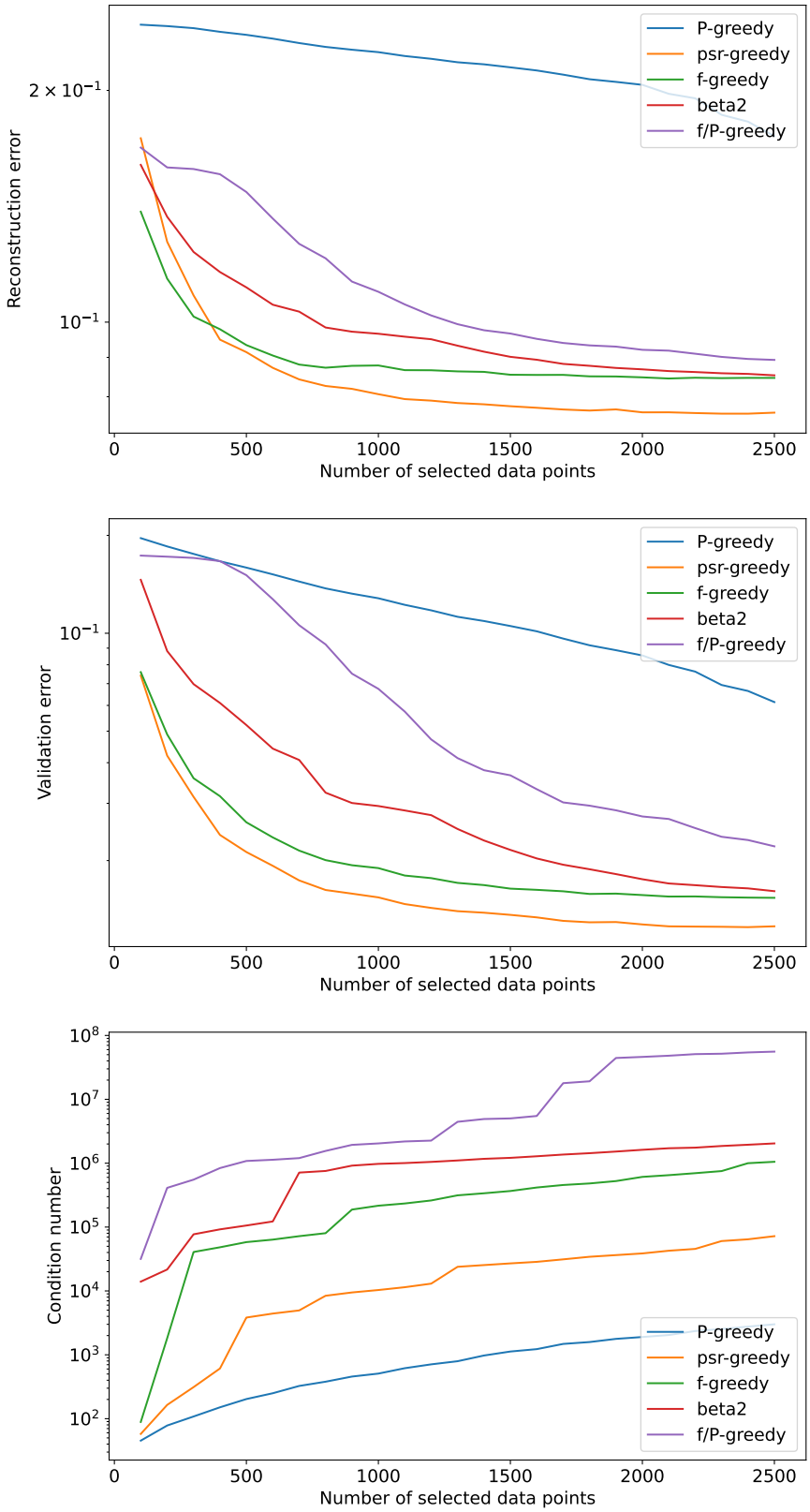

Figure 11.4.: Error and stability measurements for Shepp-Logan phantom reconstruction via β-greedy methods

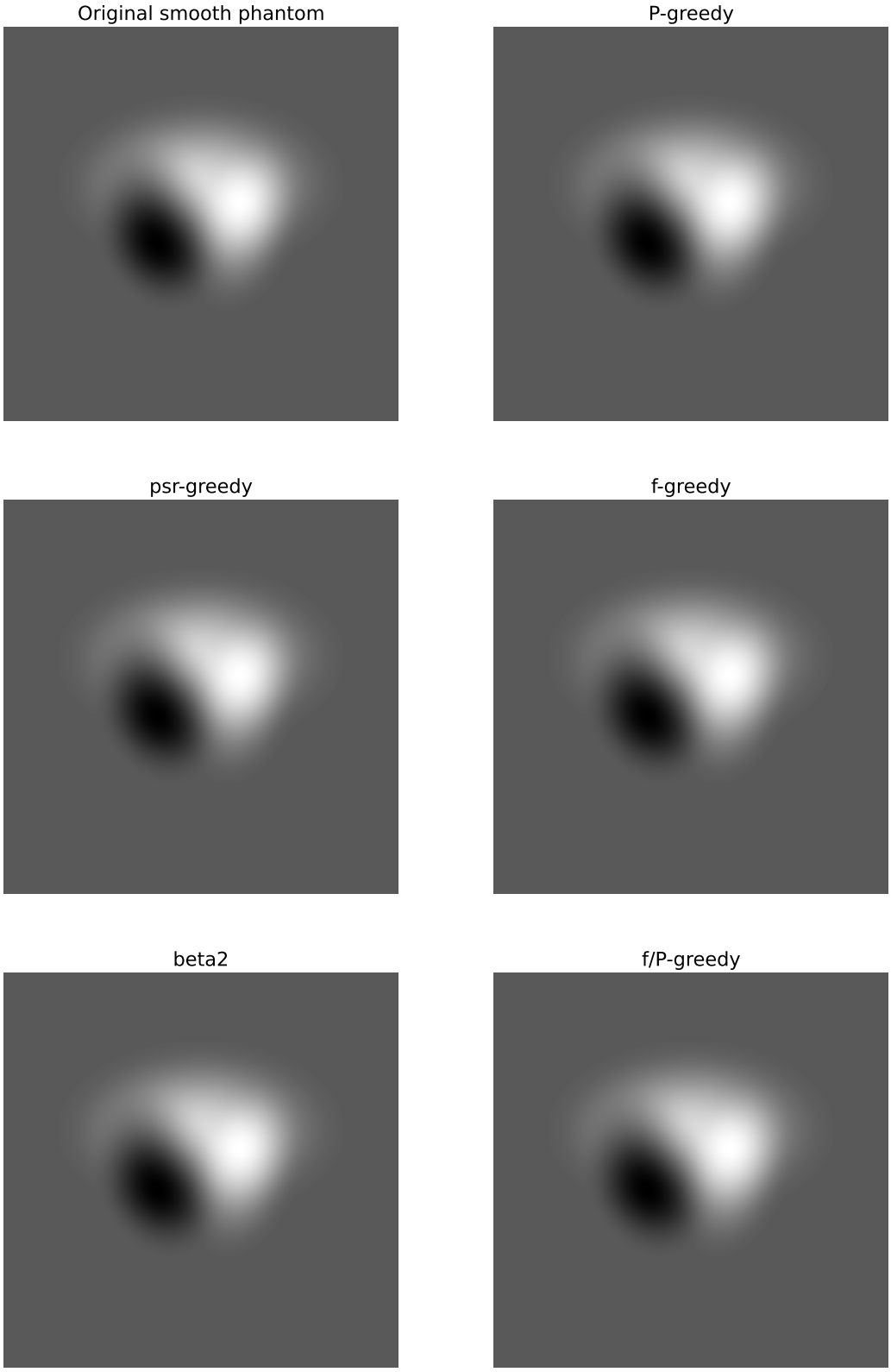

Figure 11.5.: Reconstructions of the smooth phantom for different β-greedy algorithms

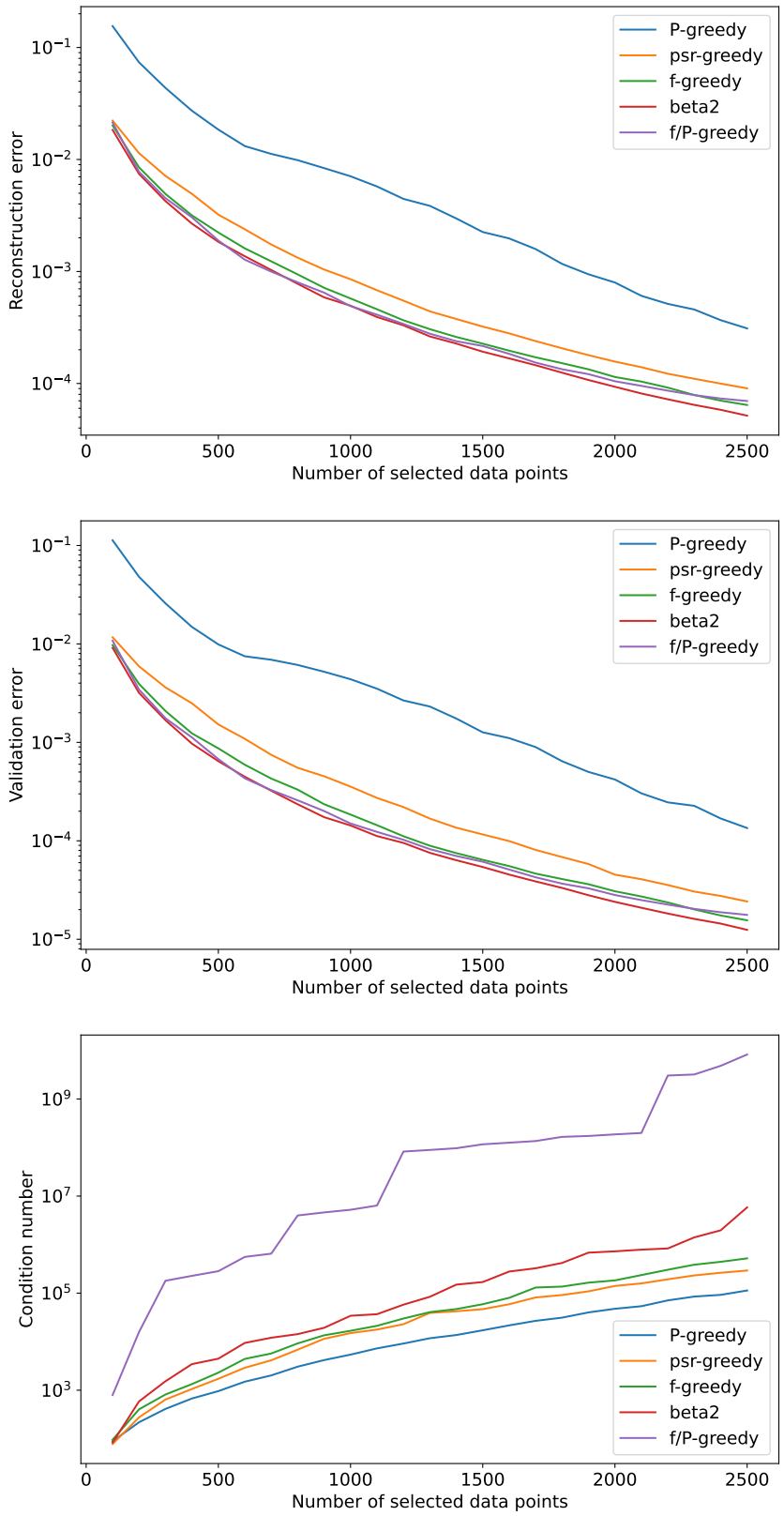

Figure 11.6.: Error and stability measurements for smooth phantom reconstruction via β-greedy methods

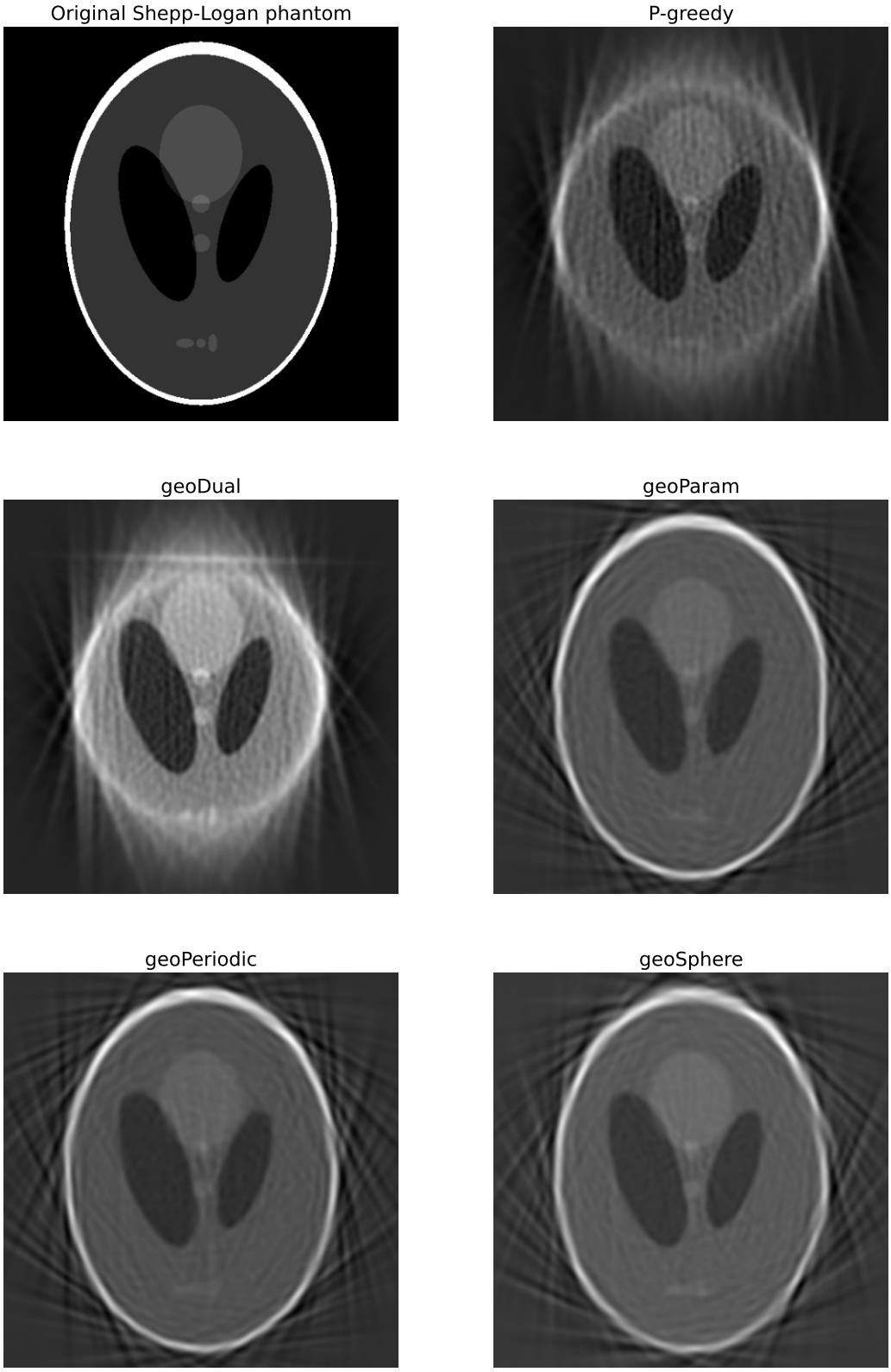

Figure 11.7.: Reconstructions of the Shepp-Logan phantom for different geometric greedy algorithms

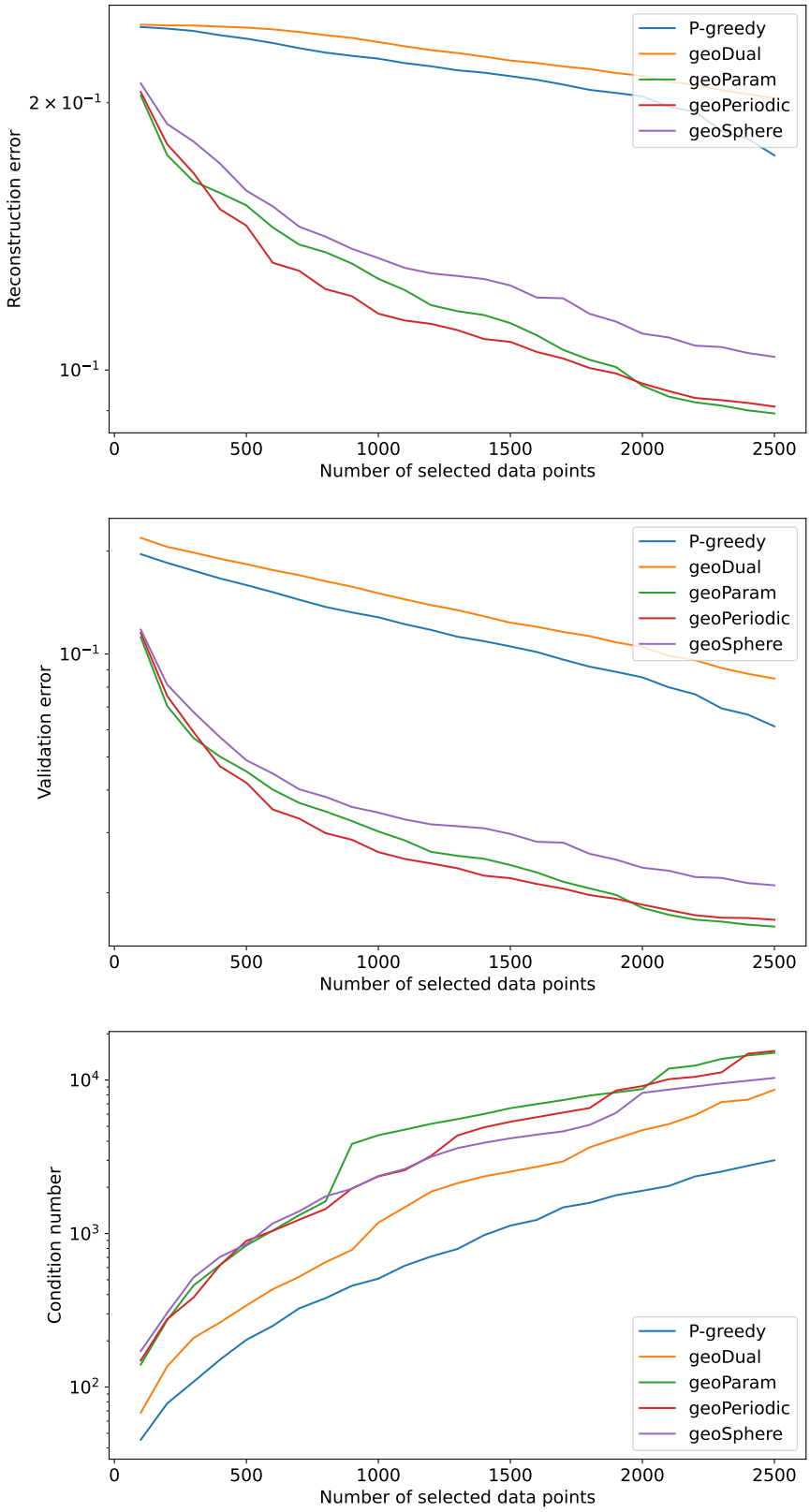

Figure 11.8.: Error and stability measurements for Shepp-Logan phantom reconstruction via geometric greedy methods

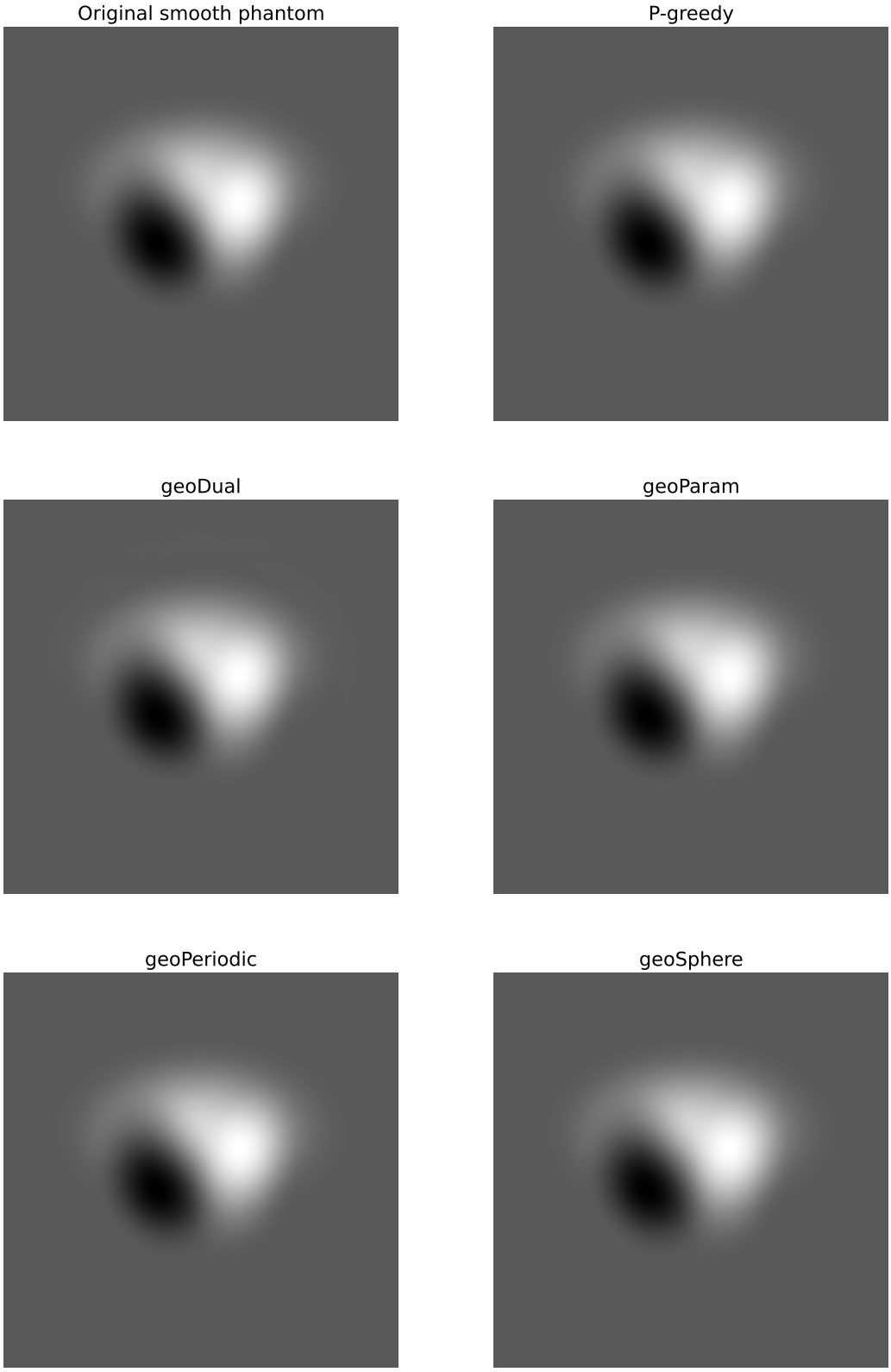

Figure 11.9.: Reconstructions of the smooth phantom for different geometric greedy algorithms

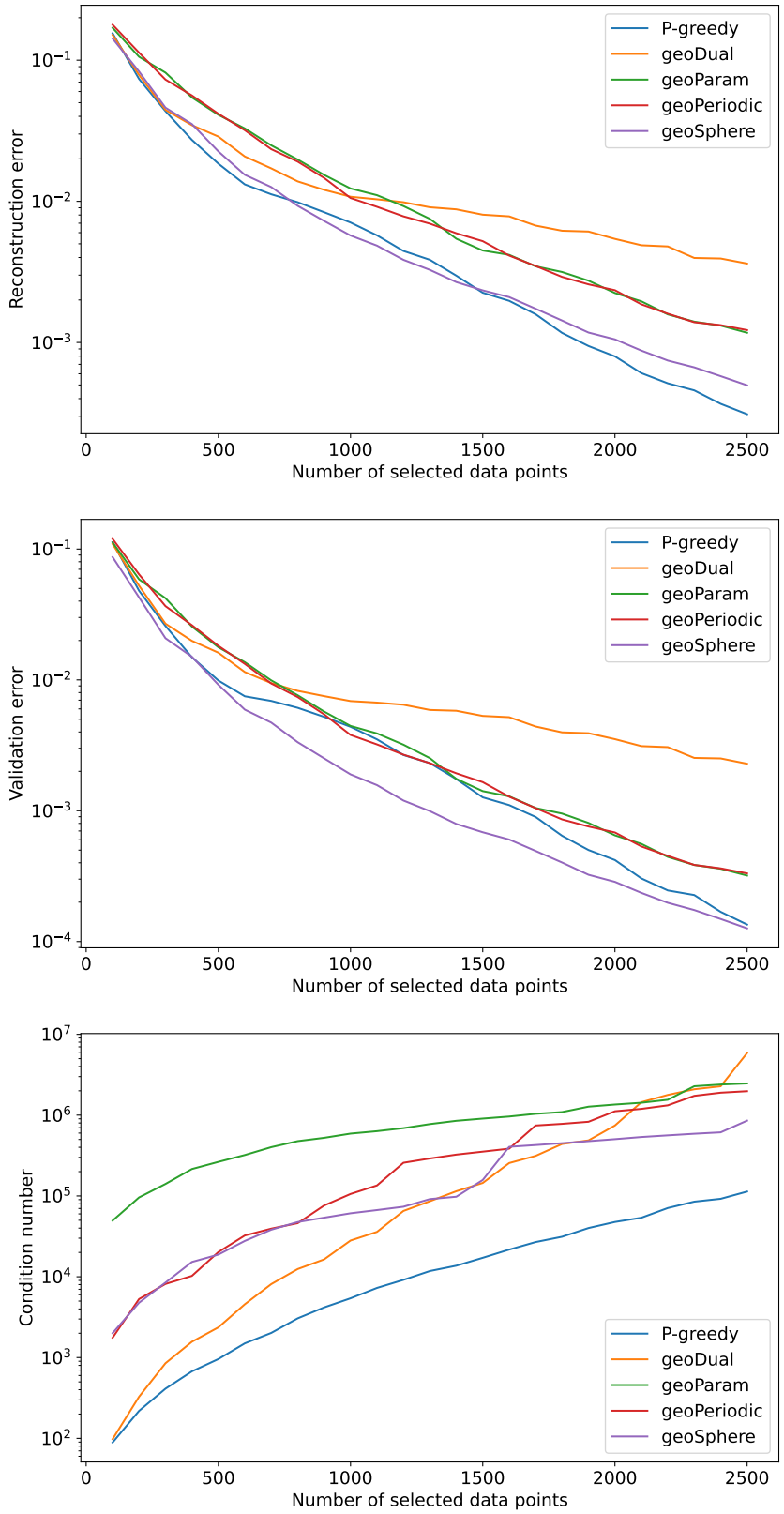

Figure 11.10.: Error and stability measurements for smooth phantom reconstruction via geometric greedy methods

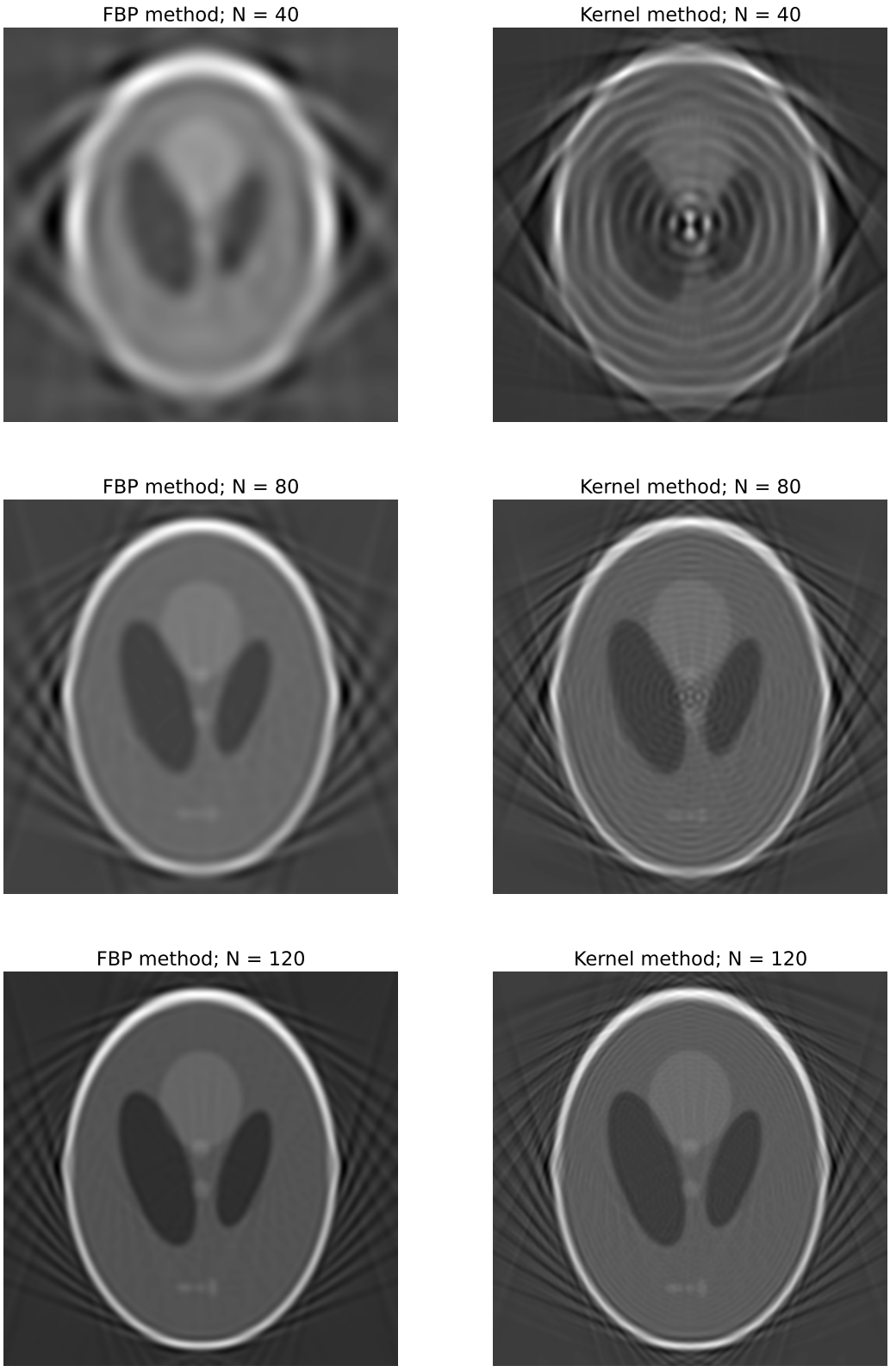

Figure 11.11.: Reconstructions of Shepp-Logan phantom via FBP and kernel method for different angular sampling rates

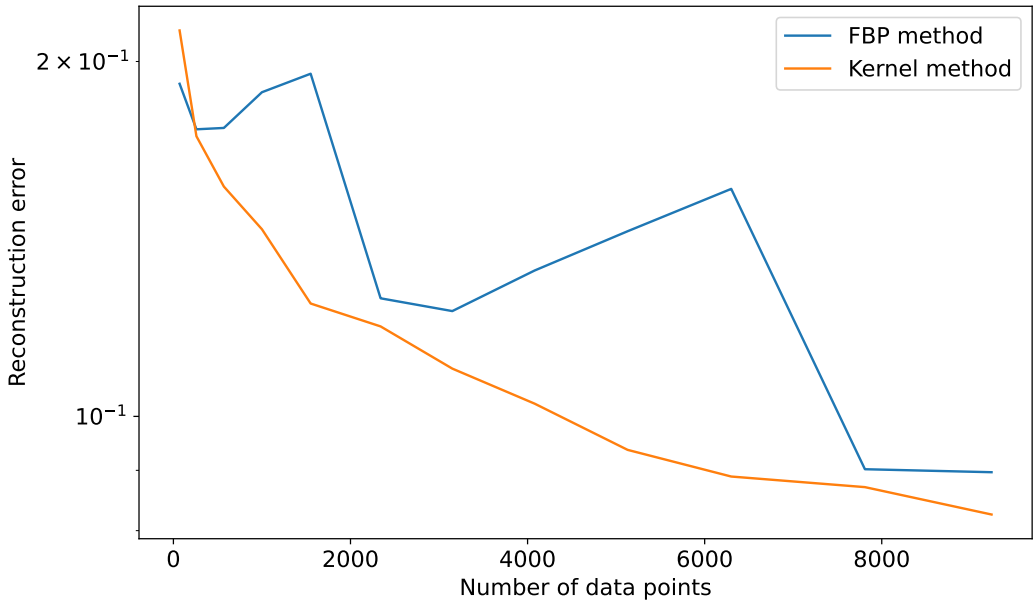

Figure 11.12.: Error decay for Shepp-Logan phantom reconstruction via FBP and kernel method

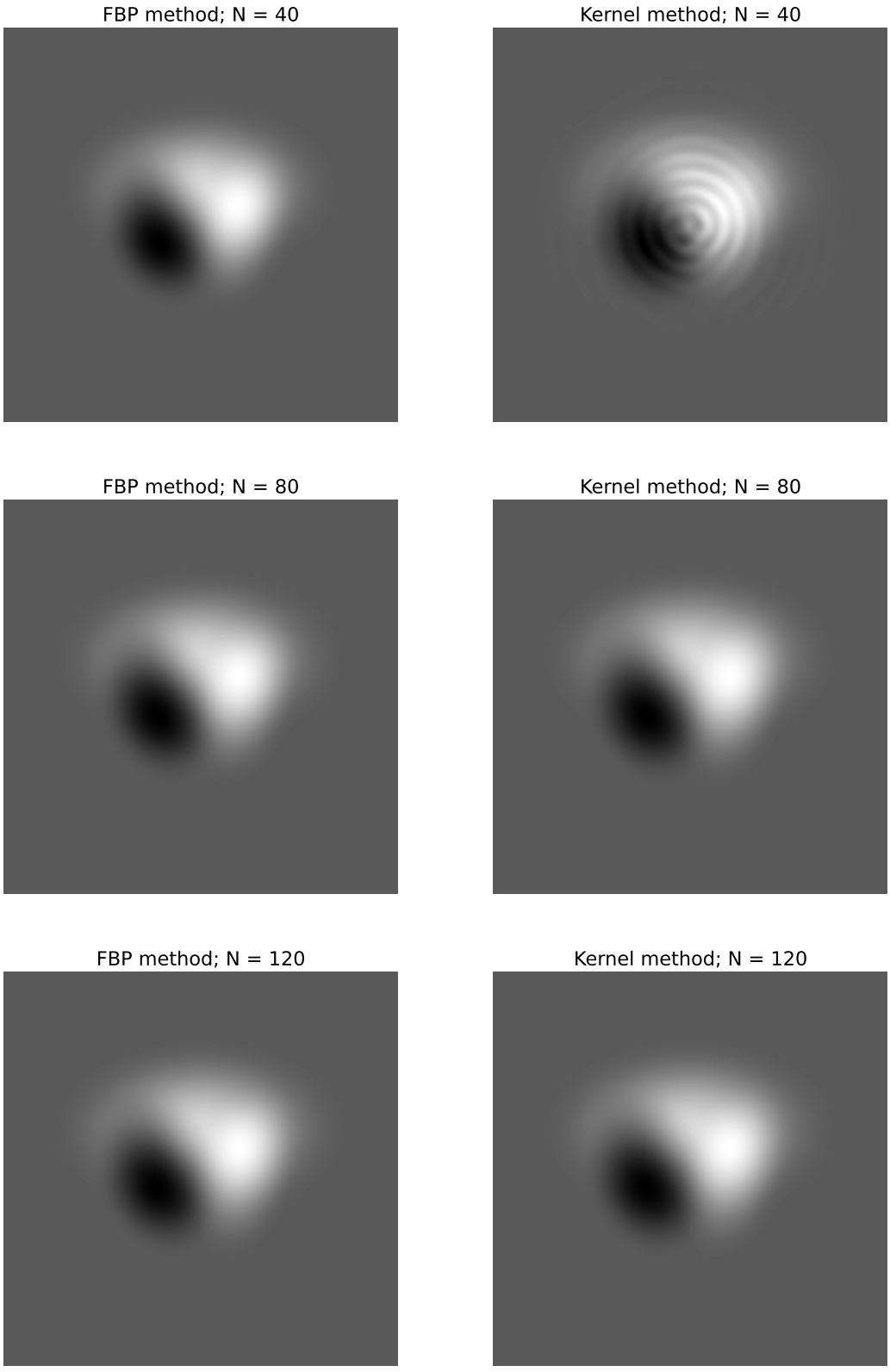

Figure 11.13.: Reconstructions of smooth phantom via FBP and kernel method for different angular sampling rates

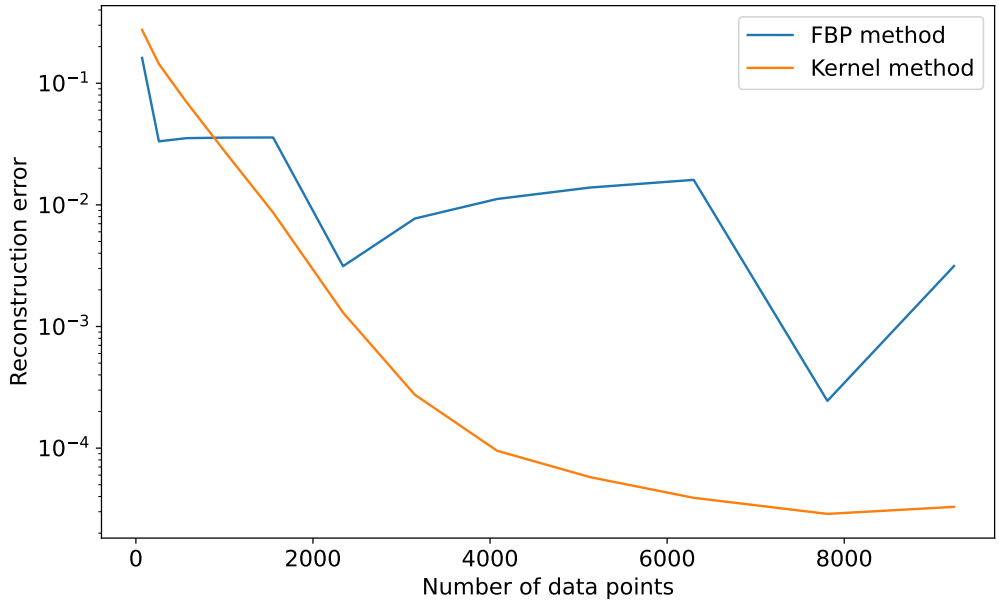

Figure 11.14.: Error decay for smooth phantom reconstruction via FBP and kernel method

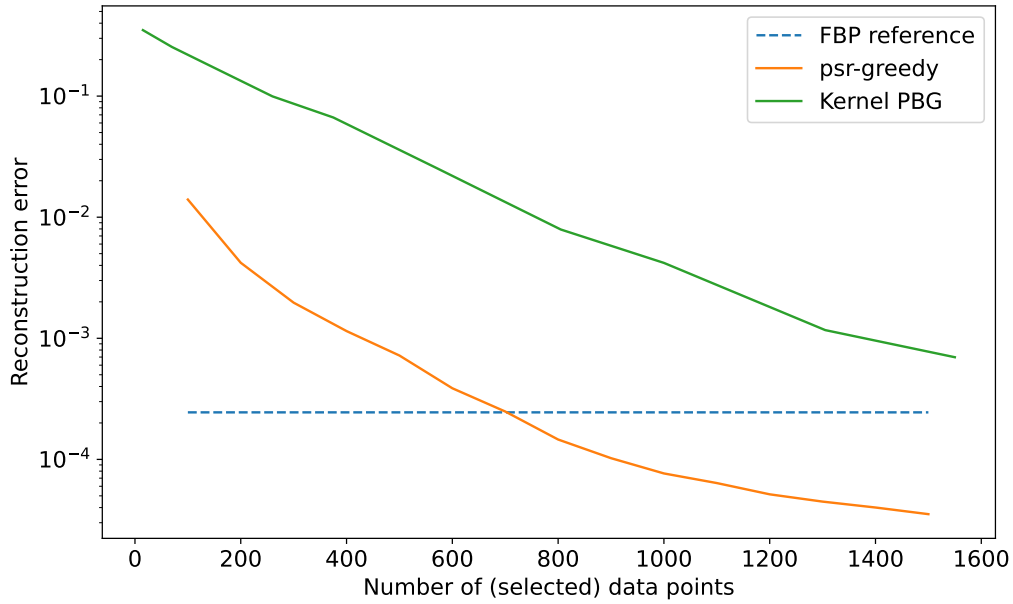

Figure 11.15.: Error decay for psr-greedy algorithm in comparison to FBP reference value and standard kernel-based method on parallel beam geometry data

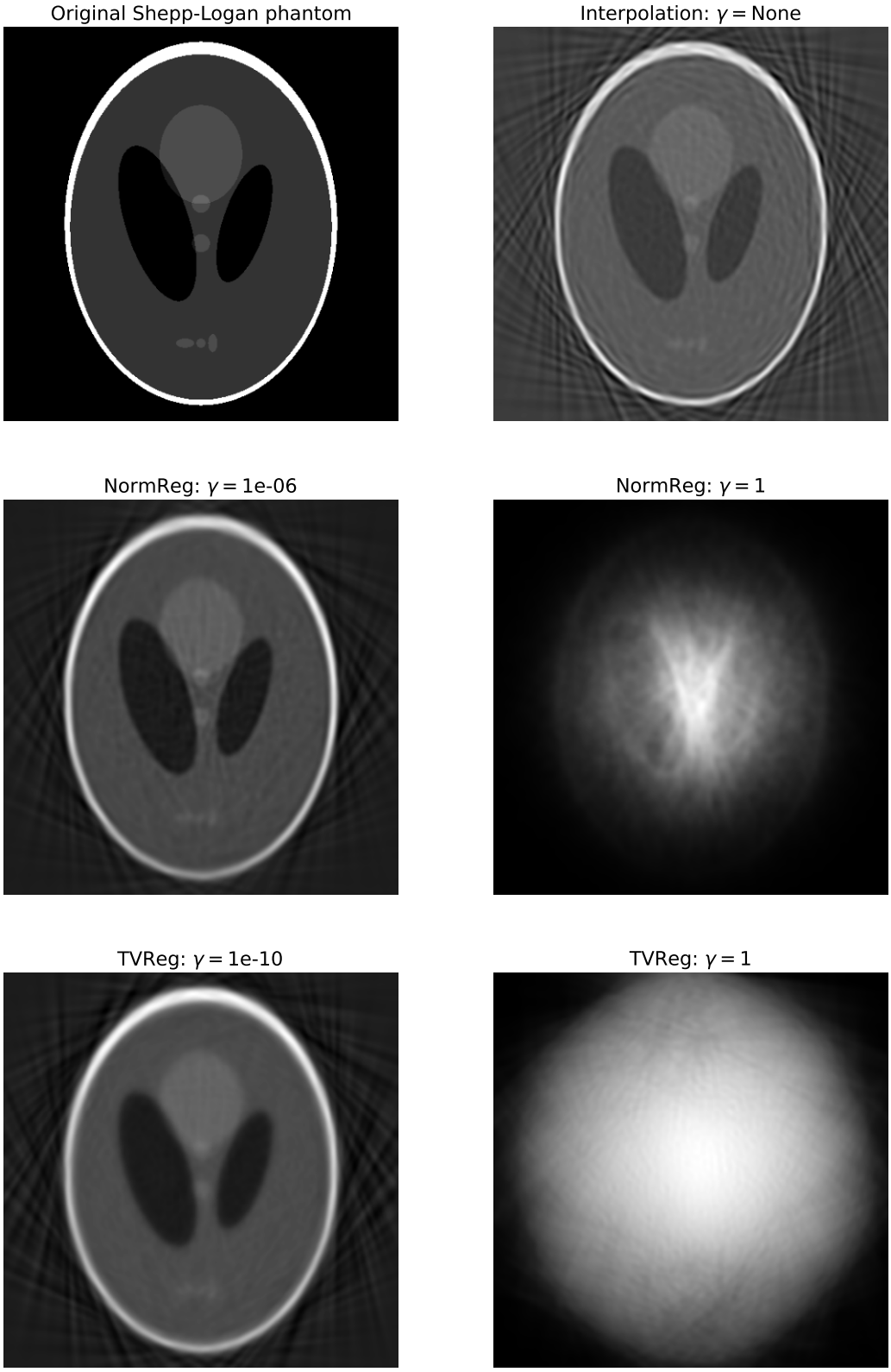

Figure 11.16.: Reconstructions of the Shepp-Logan phantom for different regularization methods

12. Summary and Outlook

Motivated by previous advances in the research field, this thesis aimed to provide an extensive overview and analysis of generalized interpolation problems in reproducing kernel Hilbert spaces and to further elaborate the kernel-based image reconstruction from scattered Radon data. Due to the generality of most theoretical results, we split our work into two parts. While the first part focused on more abstract and general analysis, the second part contained more practical work driven by the application of generalized interpolation to computerized tomography. Thus, interested readers with different applications can mostly focus on the first seven chapters.

In the first part, we discussed the theoretical aspects of generalized interpolation. Besides theoretical results on the linear independence of functionals and the convergence of the interpolation method, we explained and analyzed the generalized Newton basis and greedy data selection algorithms. Moreover, we complemented the greedy algorithms with a simple, flexible regularization approach. The content of the first part can be seen as a framework for treating generalized interpolation problems that arise when discretizing linear problems.

The second part of this thesis started with an introduction to the mathematical problem of computerized tomography and a brief overview of reconstruction methods. With these basics, we were able to discuss the image reconstruction from scattered Radon data via kernel-based generalized interpolation. Based on [38], we analyzed weighted kernel functions and showed that the Radon functionals are elements of the native dual space for suitable combinations of kernels and weight functions. Moreover, we showed that the derived framework of the first part can be applied under suitable conditions. In our numerical experiments, we demonstrated that the choice of Radon functionals is a critical component of the reconstruction method. Recall that this was already well-known for standard interpolation applications. In addition, numerical examples indicated that the kernel-based reconstruction method can compete with other, well-established reconstruction methods in certain scenarios. The numerical results underline the importance of flexibility and the effectiveness of the derived tools.

Note that we do not summarize the theoretical results of this thesis in more detail here, since we already provided a detailed list of our contributions to the research field in the introduction of this thesis. In the following, we will give an outlook on real-world applications and state remaining open problems.

Real-world applications in computerized tomography

Based on the numerical results of Section 11.3, we want to discuss possible real-world applications for the kernel-based reconstruction method. As shown in Figure 11.15, the kernel-based method in combination with proper shape parameters and a suitable greedy algorithm can be used to reduce the number of data points while maintaining the accuracy of the FBP method in certain settings. For the scanning procedure, one could still use common scanners that follow a regular scanning scheme like the parallel beam geometry but only take the measurements for lines that belong to the thinned data set. Figure 12.1 shows a modification of the parallel beam geometry scanning scheme from Figure 9.2, where only a selection of Radon lines (blue) is considered at each angle. The remaining Radon lines (red) are ignored in the scanning procedure. Consequently, the performance of the kernel-based methods in practice could easily be tested in existing systems for computerized tomography.

In medical imaging, the reduction of data points results in lower radiation exposure for patients. However, the thinning process requires the Radon data for a large set of Radon lines to derive a suitable representation of the complete Radon data. A lower radiation exposure is only achieved if the thinning is performed on training data sets prior to the actual examination of patients. To this end, it should be investigated whether the vectorial greedy approach from Subsection 7.2.4 is able to determine functional sets that generalize well to certain groups of examinations. Moreover, it should be investigated whether the sequential thinning approach from Section 10.3 yields a similar accuracy as the usual thinning procedure, since reconstructions are available earlier within the sequential approach.

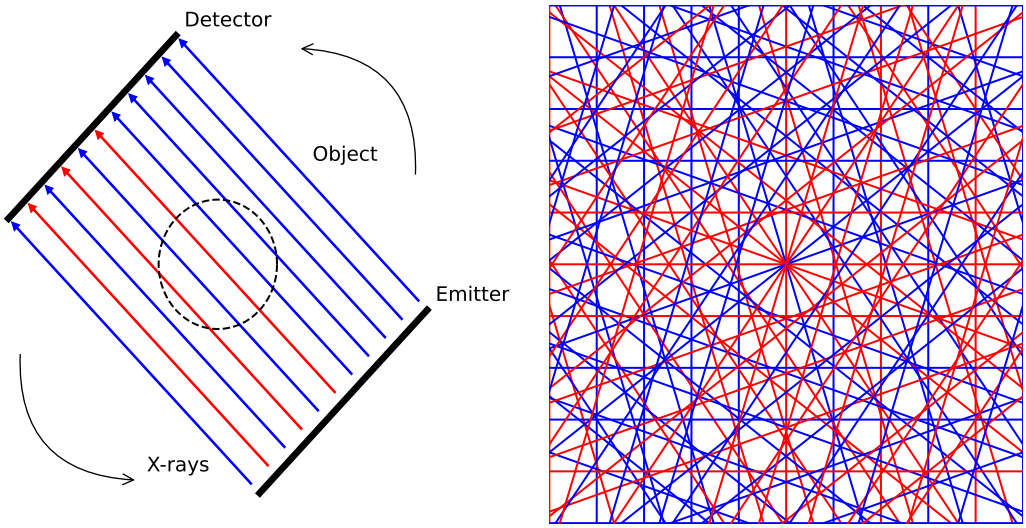

Figure 12.1.: Modification of parallel beam geometry scanning procedure from Figure 9.2

Kernel-based methods for linear inverse problems

The main objective of the first part was to derive an extensive framework for solving generalized interpolation problems with positive definite kernel functions. These problems usually arise when discretizing linear inverse problems such as the reconstruction from line integral values, which served as our main application in the second part of this thesis. Due to the mild assumptions made in the first part, we were able to apply the derived theory and tools to computerized tomography without further modifications. Hence, we assume that the derived framework can be applied successfully to many different linear problems. For the sake of clarity, we want to list the two key requirements for the application of the kernel-based framework.

To this end, let $A : \mathcal{F}_1(X) \to \mathcal{F}_2(Y)$ be a bounded linear operator between two Banach spaces $\mathcal{F}_1(X)$ and $\mathcal{F}_2(Y)$ consisting of real-valued functions on the domains X and Y. Given a target object $f \in \mathcal{F}_1(X)$, we can only perform finitely many measurements, i.e. we only have access to the values $Af(y_i)$, $i = 1, ..., n$, corresponding to a finite point set $\{y_1, ..., y_n\} \subset Y$. In the context of generalized interpolation, this leads to the discretized problem

$$As(y_i) = Af(y_i) \qquad i = 1, ..., n, \tag{12.1}$$

where $s \in \mathcal{F}_1(X)$ is the approximation to f. The requirements of the kernel-based framework can then be described as follows:

(i) The problem (12.1) is a generalized interpolation problem with respect to the functionals $\delta_{y_i} \circ A$, $i = 1, ..., n$, where δ_{y_i} denote the respective Dirac functionals. Hence, we need to choose a kernel function K such that $\delta_y \circ A \in \mathcal{H}_K^*$ for all $y \in Y$. Due to the reproduction property, it is rather easy to find suitable kernels for common problems, see e.g. Proposition 3.14 and Section 4.1. In addition, one can modify standard kernels to enforce the boundedness of these functionals in certain settings. For example, weighted kernels as introduced in Section 10.1 represent a simple modification and guarantee the boundedness of the Radon functionals for appropriate weight functions (cf. Section 10.2).

(ii) For the efficient implementation of the framework, it is essential that the Gram matrix entries

$$\left\langle \delta_{y_i} \circ A,\ \delta_{y_j} \circ A \right\rangle_K \qquad i, j = 1, ..., n$$

can be computed efficiently, since these values are the basis of the whole reconstruction process (cf. Algorithm 1). To this end, we are mostly interested in deriving analytic expressions for the evaluation of the inner product. In the case of partial derivative operators, we just have to

differentiate the chosen kernel. However, integral operators usually require a more careful choice of the kernel. In [148], the author was not able to derive an analytic expression of the inner product due to the complexity of the spherical Radon transform. We face the same problem in Section 10.2, where we have only provided one suitable combination of a kernel and weight function in Example 10.16. Until now, no further combinations are known that deliver suitable analytic expressions. Hence, this requirement is a critical part of the application to inverse problems.

If the considered operator and the chosen kernel meet the previous two requirements, the framework including its interpolation tools can be applied efficiently to determine a reconstruction.

Open problems

At last, we want to highlight relevant problems and questions that arose in our analysis and were not solved. The following list can be used as a guideline for further advances in the research field of kernel-based (generalized) interpolation.

- **Newton basis computation:** As stated in Remark 6.12, the computational costs of evaluating a given functional from the dual space at all Newton basis elements grows quadratically, which is a significant disadvantage in applications with huge underlying data sets. In Section 11.2, we restricted ourselves to data sets of order 10^5 and lower to achieve feasible computation times. For significantly larger sets, it would take hours to compute a reconstruction with the presented kernel-based algorithms. Meanwhile, the FBP method is able to compute a reconstruction from millions of data points in a few seconds on the same hardware. A more efficient computation scheme for the Newton basis or a similar basis would significantly improve the efficiency of the whole reconstruction scheme. Recall that the inefficiency of algebraic reconstruction methods is still one of the major obstacles to their practical use (cf. Section 9.2).

- **Discontinuous kernels:** In Section 10.2, we assumed the weighted kernels to be continuous to ensure that their native spaces can be embedded into $L^1(\mathbb{R}^2)$, see Remark 10.17. Since common kernel functions are usually continuous (cf. Section 2.1), this assumption is not a severe limitation. However, it could be beneficial to investigate the construction of discontinuous kernels that are integrable and provide a proper L^p-embedding for their native space. Recall from Chapter 3 that the smoothness of the kernel determines the smoothness of the functions from the native space. Thus, selecting a smooth kernel leads to a smooth reconstruction model and we might not be able to properly reconstruct relevant features like discontinuities of the target function. At the moment, we are only aware of the approach in [40], where the authors introduced the concept of *Variably Scaled Discontinuous Kernels (VSDK)*. It remains to be investigated whether this type of kernel is useful for image reconstruction from scattered Radon data.

- **Further examples of weighted kernels:** So far, the weighted Gaussian kernel from Example 10.6 is the only known combination of a kernel and weight function that yields an analytic expression of the Gram matrix entries and the evaluation of the Riesz representers in the case of Radon functionals. Similar to the standard interpolation with a Gaussian kernel, the stability and accuracy of the weighted Gaussian model heavily rely on the determination of good shape parameters. Moreover, the smooth weighted Gaussian model is probably not suited for piecewise-constant objects like the Shepp-Logan phantom, see Section 11.2. To enhance the versatility of the kernel-based reconstruction method, it is essential to find other suitable weighted kernels that provide a less smooth and more stable reconstruction model.

- **Linear independence of Radon functionals:** We were not able to provide a linear independence result for the Radon functionals from (10.1) in Subsection 10.2.1. Although we ended up with the integral representation (10.8), which is similar to the representation from 4.4, we could not proceed as in Section 4.1 due to the non-compact support of the Radon functionals. However, we still expect that the Radon functionals are linearly independent for pairwise distinct parameter pairs and suitably weighted kernels, especially since we did not run into major numerical trouble in our experiments. To prove the linear independence, one would have to take a closer look at the Fourier theory of non-compactly supported and tempered distributions. As we have already stated several times before, the linear independence result only has a theoretical value (cf. end of Subsection 10.2.1). In practical cases, we can keep track of the power function to identify linear dependence.

- **Convergence rates:** In Remark 10.25 & 10.26, we already stated a few ideas to obtain convergence rates for the image reconstruction from Radon data via weighted kernel functions. For the convergence analysis of kernel-based interpolation algorithms, it is common to restrict to Sobolev kernels and make use of the Sobolev space theory (see, e.g., [146] or Section A.5). Our numerical results from Section 11.2.1 suggest that the order of convergence depends on the Sobolev order of the weighted kernel. Hence, we assume that corresponding convergence rates can be derived. In general, it is desirable to derive convergence rates in terms of the native space norm, since this implies similar convergence in terms of other well-known norms (cf. Corollary 10.24).

- **Source condition:** For the convergence result concerning the image reconstruction from Radon data (cf. Theorem 10.23), we assumed that the target function is an element of the restricted space $\mathcal{H}_{K_w,\Gamma_R}$ for given radius $R > 0$. Since we do not have a simpler representation of this subspace, it is not clear whether this is a realistic assumption. In the theory of inverse problems, it is a common source condition to assume that the target functions are elements of (fractional) Sobolev spaces, see e.g. [18, Chapter 4]. We saw in Corollary 10.9 that, under suitable conditions, the weighted kernels native space $\mathcal{H}_{K_w}$ is a subspace of a Sobolev space again. It could be highly beneficial to investigate special examples of weighted kernels that lead to reasonable source conditions in the kernel setting.

- **Numerical comparison to other algebraic methods:** Besides the reconstruction via kernel-based generalized interpolation, we have discussed alternative algebraic reconstruction methods in Section 9.2 & 10.4. However, we only compared the theoretical aspects of these different methods and did not conduct any numerical comparison tests. It should be investigated whether the framework for kernel-based generalized interpolation results in significant advantages over previous ART-like approaches with pixel or blob functions. Recall that there is no such tool as greedy data thinning algorithms for these ART-like approaches. Hence, finding a suitable amount of basis functions is a challenging task in general and might require significantly more fine-tuning and computational resources.

- **Application to noisy or incomplete data:** In our numerical comparison with other reconstruction methods (cf. Section 11.3), we mainly focused on the number of data points needed to obtain a reasonable reconstruction. As stated in the introduction of the second part, there are other scenarios with severe data limitations, e.g. the viewing angle can be limited. It should be tested whether the kernel-based reconstruction method is able to produce suitable reconstructions in a large variety of critical scenarios. Additionally, the algorithm needs to be tested on (noisy) real-world measurements, since we only dealt with simulated data in our numerical tests.

A. Mathematical Tools

In the main part of this thesis, we made use of several well-investigated mathematical tools, e.g. the Fourier transform and the theory of distributions. We want to use this chapter as a supplement to fix definitions and notations, which may vary in the literature. Moreover, we list relevant theoretical results that were used in our analysis.

A.1. Standard Analysis

We start with some basic definitions and results from standard analysis. In this section, we do not include detailed explanations. Further information can be found in common analysis books. We just state the relevant statements, and thus try to avoid misconceptions in our analysis.

Distance function

The fill distance (cf. Section 5.2) is based on the distance function, which measures the minimal distance between a point and a subset.

Definition A.1. Let (Z, d_Z) be a metric space and $A \subset Z$. We define the distance to the subset A as

$$\operatorname{dist}(z, A) := \inf_{a \in A} d_Z(z, a) \qquad \text{for } z \in Z,$$

resulting in the `distance function` $\operatorname{dist}(\cdot, A) : Z \to \mathbb{R}$.

In Corollary 5.4, we interpret the power function as a distance to a finite-dimensional linear subspace. Thus, we can rely on the general properties of the distance function.

Lemma A.2. *In the setting of Definition A.1, the following statements hold:*

(1) We have

$$|\operatorname{dist}(z, A) - \operatorname{dist}(\tilde{z}, A)| \leq d_Z(z, \tilde{z}) \qquad \textit{for all } z, \tilde{z} \in Z,$$

i.e. the distance function is Lipschitz continuous on Z with Lipschitz constant 1.

(2) The equation $\operatorname{dist}(z, A) = 0$ *holds if and only if* $z \in \overline{A}$.

Totally bounded sets

For our convergence analysis regarding the greedy selection algorithms in Section 7.2, we restrict to totally bounded subsets of the native dual space.

Definition A.3. Let (Z, d_Z) be a metric space and $A \subset Z$. We call A `totally bounded` if for any $\varepsilon > 0$, there are finitely many points $z_1, ..., z_M \in Z$ such that

$$A \subset \bigcup_{j=1}^{M} B_\varepsilon(z_j).$$

In other words, totally bounded sets can be covered with a finite amount of ε-balls for any given radius. Note that $A \subset Z$ is totally bounded if and only if it is totally bounded in the metric subspace $\big(A, d_{Z|A}\big)$, where $d_{Z|A}$ is the restriction of d_Z to $A \times A$, which leads to the equivalent notion of `totally bounded spaces`. Similar to the concept of compactness, it is not easy to determine whether a subset is totally bounded or not just by applying the definition. We state some useful properties that help us to identify totally bounded subsets.

Lemma A.4. *Let (Z, d_Z) be a metric space and $A \subset Z$.*

(1) If A is totally bounded and $B \subset A$, then B is totally bounded as well.

(2) If A is compact, then A is totally bounded. Conversely, if A is totally bounded and complete with respect to d_Z, then A is compact.

(3) If (Z, d_Z) is complete, then A is totally bounded if and only if it is pre-compact. Hence, if Z is a finite-dimensional real vector space and d_Z is induced by a norm on Z, then A is totally bounded if and only if A is bounded.

(4) Let (Y, d_Y) be another metric space and $\varrho : Z \to Y$ be a uniformly continuous mapping. If A is totally bounded, then its image $\varrho(A)$ is totally bounded in Y.

Forward differences

In Proposition 3.14 part (4), we have used the approximation properties of *forward difference operators* regarding partial derivatives. Here, we want to properly introduce these operators starting with the univariate case. Given a function $f : \mathbb{R} \to \mathbb{R}$, it is common knowledge that its difference quotient at point $x \in \mathbb{R}$ with step width $h \neq 0$ is given by

$$\Delta_h f(x) := \frac{f(x+h) - f(x)}{h}.$$

By iterating this formula, we can define forward differences of arbitrarily high orders.

Definition A.5. For $f : \mathbb{R} \to \mathbb{R}$ and $k \in \mathbb{N}_0$, we define the k-`th forward differences` as

$$\Delta_h^k f(x) := h^{-k} \cdot \sum_{i=0}^{k} (-1)^{k-i} \cdot \binom{k}{i} \cdot f(x + i \cdot h) \qquad \text{for } x \in \mathbb{R}, h \neq 0. \tag{A.1}$$

This defines the `forward difference operator` Δ_h^k, which maps a function f to another function $\Delta_h^k f$ according to (A.1).

For these forward difference operators, one can prove an extended mean value theorem and the convergence to high-order derivatives for sufficiently smooth functions.

Lemma A.6. *Let $f : \mathbb{R} \to \mathbb{R}$ be a function.*

(1) For $k_1, k_2 \in \mathbb{N}_0$, we have

$$\Delta_h^{k_1} \left[\Delta_h^{k_2} f \right] = \Delta_h^{k_1 + k_2} f \qquad \text{for all } h \neq 0.$$

(2) Suppose that f is k-times differentiable for $k \in \mathbb{N}_0$. Then, for every $x \in \mathbb{R}$ and $h \neq 0$, there is $\xi \in Q_{x,h}^{(k)}$ such that

$$\Delta_h^k f(x) = f^{(k)}(\xi).$$

Here, $Q_{x,h}^{(k)}$ is defined as the closed interval

$$Q_{x,h}^{(k)} := \left[x + \min(0, k \cdot h), x + \max(0, k \cdot h) \right]. \tag{A.2}$$

(3) If $f \in \mathscr{C}^k(\mathbb{R}^d)$ for $k \in \mathbb{N}_0$, we have

$$\lim_{h \to 0} \Delta_h^k f(x) = f^{(k)}(x) \qquad \text{for all } x \in \mathbb{R}.$$

Based on the univariate forward differences, we derive a multivariate version for functions $f : \mathbb{R}^d \to \mathbb{R}$. Given an index $l \in \{1, ..., d\}$ and an order $k \in \mathbb{N}_0$, we can apply the univariate operator to the l-th component of the input argument, i.e.

$$\Delta_{h,l}^k f(x) := h^{-k} \cdot \sum_{i=0}^{k} (-1)^{k-i} \cdot \binom{k}{i} \cdot f(x + i \cdot h \cdot e^{(l)}) \qquad \text{for } x \in \mathbb{R}, h \neq 0,$$

where $e^{(l)} \in \mathbb{R}^d$ denotes the l-th standard basis vector. By concatenating forward difference operators acting on the separate components of $\mathbb{R}^d$, we get the multivariate forward difference operators. To simplify our analysis, we recall the multi-index notation from multivariate calculus. A multi-index is given by an array

$$\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d$$

of non-negative integers. The modulus of a multi-index is defined as

$$|\alpha| = \sum_{l=1}^{d} \alpha_l.$$

Note that, since $\mathbb{N}_0^d \subset \mathbb{R}^d$ holds, we can perform the usual componentwise addition, subtraction and scalar multiplication. Moreover, we consider the partial ordering and binomial coefficient

$$j \leq \alpha \ :\Leftrightarrow \ j_l \leq \alpha_l \text{ for all } l \in \{1, ..., d\} \qquad \text{and} \qquad \binom{\alpha}{j} := \prod_{l=1}^{d} \binom{\alpha_l}{j_l}$$

for multi-indices $\alpha, j \in \mathbb{N}_0^d$.

Definition A.7. Let $f : \mathbb{R}^d \to \mathbb{R}$ and $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d$ be a muti-index. We define the respective `forward differences` as

$$\Delta_h^\alpha f(x) := \Delta_{h,1}^{\alpha_1} \circ ... \circ \Delta_{h,d}^{\alpha_d} f(x) \qquad \text{for } x \in \mathbb{R}^d, h \neq 0.$$

As in the univariate case, this defines an operator Δ_h^α for each $h \neq 0$.

For the multivariate case, one can derive results that are similar to Lemma A.6. To this end, we set

$$Q_{x,h}^{(\alpha)} := Q_{x_1,h}^{(\alpha_1)} \times ... \times Q_{x_d,h}^{(\alpha_d)} \subset \mathbb{R}^d \qquad \text{for } x = (x_1, ..., x_d) \in \mathbb{R}^d,\ \alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d,\ h \neq 0, \tag{A.3}$$

where the components $Q_{x_l,h}^{(\alpha_l)} \subset \mathbb{R}$, $l = 1, ..., d$, are defined as in (A.2).

Lemma A.8. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a multivariate function.*

(1) For any multi-index $\alpha \in \mathbb{N}_0^d$, we have

$$\Delta_h^\alpha f(x) = h^{-|\alpha|} \cdot \sum_{\substack{j \in \mathbb{N}_0^d \\ j \leq \alpha}} (-1)^{|\alpha|-|j|} \cdot \binom{\alpha}{j} \cdot f(x + h \cdot j) \qquad \textit{for all } x \in \mathbb{R}^d, h \neq 0.$$

(2) Suppose that $f \in \mathscr{C}^k(\mathbb{R}^d)$ for $k \in \mathbb{N}_0$ and $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq k$. Then, given $x \in \mathbb{R}^d$ and $h \neq 0$, there is $\xi \in Q_{x,h}^{(\alpha)}$ (cf. (A.3)) such that

$$\Delta_h^\alpha f(x) = D^\alpha f(\xi).$$

In particular, we have

$$\lim_{h \to 0} \Delta_h^\alpha f(x) = D^\alpha f(x).$$

A.2. Orthogonal Projection

The construction of the native space in Section 3.1 leads to a Hilbert space of functions, where we can apply various well-known results from functional analysis. For example, we can make use of the Fréchet-Riesz representation theorem, which was already stated in Theorem 3.2. Here, we want to focus on the *orthogonal projection operators* and mainly fix notations. Further explanations can be found in any standard book on functional analysis.

Let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ be a Hilbert space and $S \subset \mathcal{H}$ be a subspace. Given $f \in \mathcal{H}$, we are interested in finding the best approximation to f from V, i.e. an element $s_f \in S$ such that

$$\|f - s_f\|_{\mathcal{H}} = \inf_{s\in S} \|f - s\|_{\mathcal{H}}. \tag{A.4}$$

If S is closed, we can guarantee the unique existence of best approximations.

Theorem A.9. *Let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ be a Hilbert space and $S \subset \mathcal{H}$ be a closed subspace. Then, the following statements hold:*

(1) For any $f \in \mathcal{H}$, there is a unique best approximation $s_f \in S$ satisfying (A.4). The best approximation is equivalently characterized by the orthogonality relation

$$\langle f - s_f, s\rangle_{\mathcal{H}} = 0 \qquad \text{for all } s \in S.$$

(2) The Hilbert space $\mathcal{H}$ can be decomposed via the orthogonal sum

$$\mathcal{H} = S \oplus S^{\perp},$$

where $S^{\perp}$ denotes the orthogonal complement of S. In particular, we have

$$\|s_f\|_{\mathcal{H}} \leq \|f\|_{\mathcal{H}},$$

where equality holds if and only if $f \in S$.

(3) If S is finite-dimensional with orthonormal basis $B = \{b_1, ..., b_n\}$, then s_f can be computed via the orthogonal projection formula

$$s_f = \sum_{i=1}^{n} \langle f, b_i\rangle_{\mathcal{H}} \cdot b_i.$$

Moreover, the squared approximation error is given by

$$\|f - s_f\|_{\mathcal{H}}^2 = \|f\|_{\mathcal{H}}^2 - \sum_{i=1}^{n} |\langle f, b_i\rangle_{\mathcal{H}}|^2.$$

Definition A.10. Let $S \subset \mathcal{H}$ be a closed subspace. According to Theorem A.9, we can define the respective `orthogonal projection operator`

$$\mathcal{P}_S : \mathcal{H} \to V,\ f \mapsto s_f.$$

From the previous discussion, it immediately follows that $\mathcal{P}_S$ is a linear operator with operator norm

$$\|\mathcal{P}_S\|_{\mathcal{H}} = \sup_{f\in\mathcal{H}\setminus\{0\}} \frac{\|\mathcal{P}_S(f)\|_{\mathcal{H}}}{\|f\|_{\mathcal{H}}} = 1.$$

A.3. Fourier Transform

The *Fourier transform* has been an essential tool in several sections of this thesis, e.g. for Bochner's theorem (cf. Theorem 2.7) and the inversion of the Radon transform (cf. Section 8.2). For the reader's convenience, we provide relevant definitions and properties that add further explanation to the analysis of the main part and clarify notations in this section. More details can be found in [131] and [147, Section V.2]. We start with the definition of the Fourier transform on $L^1(\mathbb{R}^d)$ (cf. [147, Definition V.2.1]).

Definition A.11. For $f \in L^1(\mathbb{R}^d)$, we define the `Fourier transform` $\mathcal{F}_{L^1} f$ of f as

$$\mathcal{F}_{L^1} f(\omega) := (2\pi)^{-d/2} \cdot \int_{\mathbb{R}^d} f(x) \cdot e^{-i \cdot \langle x, \omega \rangle_2} \, dx \qquad \text{for all } \omega \in \mathbb{R}^d. \tag{A.5}$$

Here, the factor i denotes the imaginary unit.

Note that the integral (A.5) is well-defined for integrable functions, so that $\mathcal{F}_{L^1}$ defines a linear operator on $L^1(\mathbb{R}^d)$.

Example A.12. In Example 2.8, we have already stated that the Fourier transform of the Gaussian function

$$G_\nu(x) = e^{-\nu \cdot \|x\|_2^2} \qquad \text{for } x \in \mathbb{R}^d$$

with shape parameter $\nu > 0$ is again a Gaussian function. More precisely, its Fourier transform is given by

$$\mathcal{F} G_\nu(\omega) = (2\nu)^{-d/2} \cdot e^{-\frac{\|\omega\|_2^2}{4\nu}} \qquad \text{for all } \omega \in \mathbb{R}^d,$$

see e.g. [143, Theorem 5.18 ff.] or [147, Lemma V.2.6] for computations. We remark that, due to the simple relation between a Gaussian function and its Fourier transform, Gaussian mollifiers are popular tools in Fourier analysis.

In the following, we list relevant properties of the Fourier transform on $L^1(\mathbb{R}^d)$ (see e.g. [131, Theorem 1.1, 1.2 & 1.4] and [147, Satz V.2.2]).

Proposition A.13. *Let $f \in L^1(\mathbb{R}^d)$ be an integrable function.*

(1) The Fourier transform $\mathcal{F}_{L^1} f$ is uniformly continuous on $\mathbb{R}^d$ and bounded by

$$|\mathcal{F}_{L^1} f(\omega)| \leq (2\pi)^{-d/2} \cdot \|f\|_{L^1(\mathbb{R}^d)} \qquad \text{for all } \omega \in \mathbb{R}^d.$$

Consequently, the Fourier transform

$$\mathcal{F}_{L^1} : L^1(\mathbb{R}^d) \to \mathscr{C}_b(\mathbb{R}^d), \; f \mapsto \mathcal{F}_{L^1} f$$

is a bounded linear operator, where

$$\mathscr{C}_b(\mathbb{R}^d) = \left\{ f \in \mathscr{C}(\mathbb{R}^d) \;\middle|\; \|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)| < \infty \right\}$$

is the space of bounded continuous functions on $\mathbb{R}^d$.

(2) The Fourier transform decays at infinity, i.e.

$$\mathcal{F}_{L^1} f(\omega) \to 0 \quad \text{for} \quad \|\omega\|_2 \to \infty.$$

(3) Given another function $g \in L^1(\mathbb{R}^d)$, we have

$$\mathcal{F}_{L^1} [f * g] (\omega) = (2\pi)^{d/2} \cdot \mathcal{F}_{L^1} f(\omega) \cdot \mathcal{F}_{L^1} g(\omega) \qquad \text{for all } \omega \in \mathbb{R}^d.$$

Here, $f * g$ denotes the usual convolution product on $L^1(\mathbb{R}^d)$, i.e.

$$(f * g)(x) := \int_{\mathbb{R}^d} f(x - y) \cdot g(y) \, dy \qquad \text{for } x \in \mathbb{R}^d.$$

Note that part (2) of the previous theorem is known as the *Riemann-Lebesgue lemma* and part (3) is usually called the *convolution theorem*. An important subspace for the analysis of $\mathcal{F}_{L^1}$ is the *Schwartz space*, which is given by

$$\mathscr{S}(\mathbb{R}^n) = \Big\{ f \in \mathscr{C}^\infty(\mathbb{R}^n) \;\Big|\; \sup_{x\in\mathbb{R}^n} (1 + \|x\|_2^m) \cdot |D^\alpha f(x)| < \infty \text{ for all } m \in \mathbb{N}_0,\ \alpha \in \mathbb{N}_0^d \Big\}.$$

It can be shown that $\mathscr{S}(\mathbb{R}^d) \subset L^p(\mathbb{R}^d)$ for all $p \in [1,\infty]$ and

$$\mathcal{F}_{L^1}\left[\mathcal{F}_{L^1} f\right](x) = f(-x) \qquad \text{for all } f \in \mathscr{S}(\mathbb{R}^d),\ x \in \mathbb{R}^d,$$

see e.g. [147, Lemma V.2.7]. With that, the inverse Fourier operator can be derived (cf. [131, Corollary 1.21] and [147, Satz V.2.8]).

Definition A.14. For $f \in L^1(\mathbb{R}^d)$, we define the `inverse Fourier transform` $\mathcal{F}_{L^1}^{-1} f$ of f as

$$\mathcal{F}_{L^1}^{-1} f(x) := (2\pi)^{-d/2} \cdot \int_{\mathbb{R}^d} f(\omega) \cdot e^{i\cdot\langle\omega,x\rangle_2}\, d\omega \qquad \text{for all } x \in \mathbb{R}^d.$$

Again, the factor i denotes the imaginary unit.

Theorem A.15. *Regarding the inversion of the Fourier transform, we have the following properties:*

(1) If $f \in L^1(\mathbb{R}^d)$ and $\mathcal{F}_{L^1} f \in L^1(\mathbb{R}^d)$, then

$$\mathcal{F}_{L^1}^{-1}\left[\mathcal{F}_{L^1} f\right] = f,$$

where the equality holds in L^1-sense. Pointwise equality holds everywhere if f is continuous. In particular, the Fourier transform $\mathcal{F}_{L^1}$ is injective.

(3) The restriction

$$\mathcal{F}_{L^1} : \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d) \tag{A.6}$$

to the Schwartz space is a well-defined automorphism which satisfies

$$\langle \mathcal{F}_{L^1} f, \mathcal{F}_{L^1} g\rangle_{L^2(\mathbb{R}^d)} = \langle f, g\rangle_{L^2(\mathbb{R}^d)} \qquad \text{for all } f, g \in \mathscr{S}(\mathbb{R}^d).$$

With part (3) of the previous theorem, we can construct the Fourier transform on the $L^2(\mathbb{R}^2)$ to benefit from the Hilbert space theory. Recall that $\mathscr{S}(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$, so that we extend (A.6) to an isometric automorphism (see, e.g., [131, Section I.2] or [147, pages 234-235]).

Theorem A.16. *Let $\mathcal{F}_{L^2} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ be the unique continuous extension of (A.6) with respect to $\|\cdot\|_{L^2(\mathbb{R}^d)}$ on $L^2(\mathbb{R}^d)$. Then $\mathcal{F}_{L^2}$ is an automorphism and isometric, i.e.*

$$\langle \mathcal{F}_{L^2} f, \mathcal{F}_{L^2} g\rangle_{L^2(\mathbb{R}^d)} = \langle f, g\rangle_{L^2(\mathbb{R}^d)} \qquad \text{for all } f, g \in L^2(\mathbb{R}^d).$$

The previous result is known as the *Plancherel theorem*. Note that for functions $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, we have the equality

$$\mathcal{F}_{L^1} f(\omega) = \mathcal{F}_{L^2} f(\omega) \qquad \text{for almost every } \omega \in \mathbb{R}^d.$$

Hence, we only denote the Fourier transform as $\mathcal{F}$ without any suffix in our analysis. From the context, it becomes clear which version is considered.

A.4. Distribution Theory

In Section 4.1, we have derived linear independence results for compactly supported distributions, where we made use of well-known results from distribution theory. In the following, we list the most relevant statements and thus provide a very brief and simplified overview. More details can be found in [67], [138] and [154]. We start with the space of *test functions*.

Definition A.17. The `space of test functions` on $\mathbb{R}^d$ is defined as

$$\mathscr{D}(\mathbb{R}^d) := \Big\{ f \in \mathscr{C}^\infty(\mathbb{R}^d) \;\Big|\; f \text{ has compact support} \Big\}.$$

Example A.18. A very common example is the bump function

$$f(x) := \begin{cases} e^{-\left(1-\|x\|_2^2\right)^{-1}} & \text{if } \|x\|_2^2 < 1, \\ 0 & \text{otherwise,} \end{cases}$$

whose support is contained in the closed unit ball. Due to the fast decay of the exponential function, f is infinitely smooth and therefore a test function. In addition, given $\varepsilon > 0$, we can define the modified version

$$g_\varepsilon(x) := \varepsilon^{-d} \cdot \|f\|_{L^1(\mathbb{R}^d)}^{-1} \cdot f(x/\varepsilon) \qquad \text{for } x \in \mathbb{R}^d.$$

This collection of functions, often referred to as a *standard mollifier*, has the following useful properties:

(i) $g_\varepsilon \in \mathscr{D}(\mathbb{R}^d)$

(ii) $\operatorname{supp}(g_\varepsilon) = \overline{B_\varepsilon(0)}$

(iii) $\|g_\varepsilon\|_{L^1(\mathbb{R}^d)} = 1$

Regarding convergence in $\mathscr{D}(\mathbb{R}^d)$, we say that a sequence $(f_n)_{n\in\mathbb{N}}$ of test functions *converges to zero* if there is a compact set $E \subset \mathbb{R}^d$ such that

$$\operatorname{supp}(f_n) \subset E \quad \text{for all } n \in \mathbb{N} \qquad \text{and} \qquad \sup_{x\in E} |D^\alpha f(x)| \xrightarrow{n\to\infty} 0 \quad \text{for all multi-indices } \alpha \in \mathbb{N}_0^d,$$

see e.g. [154, Proposition I.1.7]. Accordingly, the sequence converges to $f \in \mathscr{D}(\mathbb{R}^d)$ if $f_n - f$ converges to zero. Using this notion of convergence, a linear functional $\lambda : \mathscr{D}(\mathbb{R}^d) \to \mathbb{R}$ is *continuous on* $\mathscr{D}(\mathbb{R}^d)$ if it satisfies

$$\lambda(f_n) \xrightarrow{n\to\infty} 0 \qquad \text{for all } (f_n)_{n\in\mathbb{N}} \subset \mathscr{D}(\mathbb{R}^d) \text{ with } f_n \xrightarrow{n\to\infty} 0,$$

see e.g. [67, Theorem 2.1.4] and [154, Proposition I.8.1 ff.].

Definition A.19. The `space of distributions` is defined as

$$\mathscr{D}'(\mathbb{R}^d) := \Big\{ \lambda : \mathscr{D}(\mathbb{R}^d) \to \mathbb{R} \;\Big|\; \lambda \text{ is linear and continuous on } \mathscr{D}(\mathbb{R}^d) \Big\}.$$

Example A.20. In the following, we list some examples of distributions that were already mentioned in Section 4.1:

(i) For $x \in \mathbb{R}^d$, the respective *Dirac functional* given by

$$\delta_x(f) := f(x) \qquad \text{for all } f \in \mathscr{D}(\mathbb{R}^d)$$

is a distribution.

(ii) We have $\delta_x \circ D^\alpha \in \mathscr{D}'(\mathbb{R}^d)$ for all $x \in \mathbb{R}^d$ and $\alpha \in \mathbb{N}_0^d$, where D^α denotes the *partial differential operator*

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} \qquad \text{for } f \in \mathscr{D}(\mathbb{R}^d).$$

(iii) Let $h \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ be locally integrable, i.e. for every compact subset $E \subset \mathbb{R}^d$, we have $h \in L^1(E)$. Then, the induced functional

$$T_h(f) := \int_{\mathbb{R}^d} f(x) \cdot h(x)\, dx \qquad \text{for } f \in \mathscr{D}(\mathbb{R}^d) \tag{A.7}$$

is a well-defined distribution. Distributions of the form (A.7) are called *regular distributions*. In particular, for a compact set $E \subset \mathbb{R}^d$ with positive Lebesgue measure $\mathrm{vol}(E) > 0$, we can choose the modified characteristic function $h = \mathrm{vol}(E)^{-1} \cdot \chi_E$ to get the *cell average functional*

$$\lambda_E(f) = T_h(f) = \frac{1}{\mathrm{vol}(E)} \cdot \int_E f(x)\, dx \qquad \text{for } f \in \mathscr{D}(\mathbb{R}^d).$$

In addition to Definition A.19, we say that $(\lambda_n)_{n\in\mathbb{N}} \subset \mathscr{D}'(\mathbb{R}^d)$ *converges to* $\lambda \in \mathscr{D}'(\mathbb{R}^d)$ if

$$\lambda_n(f) \xrightarrow{n\to\infty} \lambda(f) \qquad \text{for all } f \in \mathscr{D}(\mathbb{R}^d),$$

see e.g. [67, page 38] and [154, Theorem V.1.6].

Distributions with compact support

To make use of the Fourier(-Laplace) transform and its underlying properties, we restrict to the case of distributions with compact support in our analysis. However, we first need to introduce the *support of a distribution*. We say that a distribution $\lambda \in \mathscr{D}'(\mathbb{R}^d)$ *vanishes on an open subset* $U \subset \mathbb{R}^d$ if

$$\lambda(f) = 0 \qquad \text{for all } f \in \mathscr{D}(\mathbb{R}^d) \text{ with } \mathrm{supp}(f) \subset U.$$

The largest open subset on which $\lambda \in \mathscr{D}'(\mathbb{R}^d)$ vanishes is then given by

$$U_\lambda = \bigcup \Big\{ U \subset \mathbb{R}^d \;\Big|\; U \text{ is open and } \lambda \text{ vanishes on } U \Big\}, \tag{A.8}$$

see e.g. [154, Theorem I.13.1], whose complement represents the support.

Definition A.21. Let $\lambda \in \mathscr{D}'(\mathbb{R}^d)$ and U_λ defined as in (A.8). The `support of` λ is defined as

$$\mathrm{supp}(\lambda) := \mathbb{R}^d \setminus U_\lambda.$$

Consequently, we say that λ has `compact support` if $\mathrm{supp}(\lambda) \subset \mathbb{R}^d$ is a compact subset. The set of all compactly supported distributions is denoted by

$$\mathscr{E}'(\mathbb{R}^d) := \Big\{ \lambda \in \mathscr{D}'(\mathbb{R}^d) \;\Big|\; \mathrm{supp}(\lambda) \text{ is compact} \Big\}.$$

Note that all functionals mentioned in Example A.20 have compact support. For compactly supported distributions $\lambda \in \mathscr{E}'(\mathbb{R}^d)$, we can define a special version of the Fourier transform. To this end, let $\omega \in \mathbb{R}^d$ and consider the corresponding basis function

$$e_\omega : \mathbb{R}^d \to \mathbb{C},\ x \mapsto e^{-i\cdot\langle x,\omega\rangle_2}. \tag{A.9}$$

The *Fourier-Laplace transform* is then given by the application of a compactly supported distribution to the C^∞ functions e_ω, $\omega \in \mathbb{R}^d$, see e.g. [67, Section 7.3].

Definition A.22. The `Fourier-Laplace transform` of $\lambda \in \mathscr{E}'(\mathbb{R}^d)$ is defined as

$$\mathcal{F}_{\mathscr{E}'}\lambda(\omega) = (2\pi)^{-d/2} \cdot \lambda(e_\omega) \qquad \text{for } \omega \in \mathbb{R}^d,$$

where e_ω is given by (A.9). This results in a function $\mathcal{F}_{\mathscr{E}'}\lambda : \mathbb{R}^d \to \mathbb{C}$.

Example A.23. Given $x \in \mathbb{R}^d$ and its respective Dirac functional δ_x, the largest open subset on which δ_x vanishes is $U_{\delta_x} = \mathbb{R}^d \setminus \{x\}$. Therefore, its support is given by $\operatorname{supp}(\delta_x) = \{x\}$. The Fourier-Laplace transform of the Dirac functional is given by

$$\mathcal{F}_{\mathscr{E}'}\delta_x(\omega) = (2\pi)^{-d/2} \cdot e^{-i\cdot\langle x,\omega\rangle_2} \qquad \text{for all } \omega \in \mathbb{R}^d.$$

One key property of the Fourier-Laplace transform is that it maps distributions with compact support to entire functions satisfying certain growth conditions. This result is well-known as the *Paley-Wiener-Schwartz theorem*. Here, we state a highly simplified version that is sufficient for our analysis. A detailed version can be found in [67, Theorem 7.3.1], for example.

Theorem A.24. *Let $\lambda \in \mathscr{E}'(\mathbb{R}^d)$. Then its Fourier-Laplace transform $\mathcal{F}_{\mathscr{E}'}\lambda$ can be defined on the whole complex space $\mathbb{C}^d$. This extension of $\mathcal{F}_{\mathscr{E}'}\lambda$ is an entire function, i.e. it is holomorphic on $\mathbb{C}^d$.*

As a consequence, $\mathcal{F}_{\mathscr{E}'}\lambda$ usually denotes the Fourier-Laplace transform of $\lambda \in \mathscr{E}'(\mathbb{R}^d)$ on the whole domain $\mathbb{C}^d$. To end this section, we list further properties that are relevant to our analysis.

Lemma A.25. *The Fourier-Laplace transform satisfies the following additional properties:*

(1) Let $\lambda \in \mathscr{E}'(\mathbb{R}^d)$ such that $\mathcal{F}_{\mathscr{E}'}\lambda$ has compact support. Then we have $\lambda = 0$. In particular, $\mathcal{F}_{\mathscr{E}'}$ is injective on $\mathscr{E}'(\mathbb{R}^d)$.

(2) If $(\lambda_n)_{n\in\mathbb{N}}$ is a sequence of compactly supported distributions that converges to $\lambda \in \mathscr{E}'(\mathbb{R}^d)$, then we have the pointwise convergence

$$\mathcal{F}_{\mathscr{E}'}\lambda_n(\omega) \xrightarrow{n\to\infty} \mathcal{F}_{\mathscr{E}'}\lambda(\omega) \qquad \text{for all } \omega \in \mathbb{R}^d.$$

A.5. Sobolev Spaces

In Theorem 3.18 and the following, we discussed the connection between kernel-based approximation and *Sobolev spaces*. Moreover, Sobolev spaces play a key role in the theory of inverse problems, see e.g. Section 8.3 or Section 9.1. Although our analysis did not heavily rely on the theory of Sobolev spaces, we want to provide a brief overview and highlight respective techniques for the analysis of kernel-based reconstruction methods. For further reading, we refer to [2].

We start with the definition of *weak derivatives* (see, e.g., [2, Definition 1.62]), which is one of the cornerstones of the Sobolev space theory.

Definition A.26. Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ be locally integrable and $\alpha \in \mathbb{N}_0^d$. A function $g \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ is the respective `weak derivative of` f if the condition

$$\int_{\mathbb{R}^d} g(x)\cdot h(x)\,dx = (-1)^{|\alpha|}\cdot\int f(x)\cdot D^\alpha h(x)\,dx \qquad \text{for all } h \in \mathscr{D}'(\mathbb{R}^d)$$

holds, where D^α denotes the respective differential operator. In the weak case, we denote the derivative by $g = D^\alpha f$ as well and call f `weakly differentiable`.

Note that the *fundamental lemma of variational calculus* shows that the weak derivative is unique. For differentiable functions, integration by parts shows that the usual derivative and the weak derivative coincide. The Sobolev spaces consist of functions whose weak derivatives exist and are integrable up to a certain order (see, e.g., [2, Definition 3.1 & 3.2]).

Definition A.27. For $m \in \mathbb{N}$ and $p \in [1,\infty]$, we define the respective `Sobolev space` as

$$W^{m,p}(\mathbb{R}^d) := \left\{ f \in L^p(\mathbb{R}^d) \;\middle|\; D^\alpha f \text{ exists and satisfies } D^\alpha f \in L^p(\mathbb{R}^d) \text{ for all } |\alpha| \le m \right\}$$

and equip this space with the `Sobolev space norm`

$$\|f\|_{W^{m,p}(\mathbb{R}^d)} = \begin{cases} \left(\sum\limits_{|\alpha|\le m} \|D^\alpha f\|^p_{L^p(\mathbb{R}^d)} \right)^{1/p} & \text{if } p < \infty, \\ \max\limits_{|\alpha|\le m} \|D^\alpha f\|_{L^\infty(\mathbb{R}^d)} & \text{if } p = \infty. \end{cases}$$

The Sobolev spaces for $p = 2$ are additionally denoted by

$$\mathcal{H}^m(\mathbb{R}^d) := W^{m,2}(\mathbb{R}^d).$$

We remark that the Sobolev spaces $W^{m,p}(\mathbb{R}^d)$ are Banach spaces in general and $\mathcal{H}^m(\mathbb{R}^d)$ is a Hilbert space for each $m \in \mathbb{N}$, see e.g. [2, Theorem 3.3]. Moreover, we have the following characterization and inclusion (see, e.g., [147, Satz V.2.12 & V.2.14]).

Theorem A.28. *In the case $p = 2$, the following statements hold.*

(1) For $m \in \mathbb{N}$, the respective Sobolev space can be written as

$$\mathcal{H}^m(\mathbb{R}^d) = \Big\{ f \in L^2(\mathbb{R}^d) \;\Big|\; \big(1 + \|\cdot\|_2^2\big)^{m/2} \cdot \mathcal{F}f \in L^2(\mathbb{R}^d) \Big\},$$

where $\mathcal{F}$ denotes the Fourier transform. Moreover, the norm $\|\cdot\|_m$ defined as

$$\|f\|_m := \Big\| \big(1 + \|\cdot\|_2^2\big)^{m/2} \cdot \mathcal{F}f \Big\|_{L^2(\mathbb{R}^d)} \qquad \text{for } f \in \mathcal{H}^m(\mathbb{R}^d)$$

is equivalent to the Sobolev norm defined in Definition A.27.

(2) If $k \in \mathbb{N}_0$ and $m > k + d/2$, then we have the inclusion

$$\mathcal{H}^m(\mathbb{R}^d) \subset \mathscr{C}^k(\mathbb{R}^d),$$

i.e. each equivalence class in $\mathcal{H}^m(\mathbb{R}^d)$ has a representative in $\mathscr{C}^k(\mathbb{R}^d)$.

Note that part (2) of the previous theorem is known as the *Sobolev lemma* and is usually formulated for general $p \in [1, \infty)$, see [2, Chapter 4]. The alternative characterization from part (1) including the equivalent norm is well-suited to define Sobolev spaces of fractional orders $a \geq 0$. However, in order to define Sobolev spaces for negative orders, we have to consider the *space of tempered distributions*, which is denoted by $\mathscr{S}'(\mathbb{R}^d)$ (see, e.g., [93, page 75 ff.]).

Definition A.29. Let $a \in \mathbb{R}$. The respective `Sobolev space of fractional order` is defined as the Hilbert space

$$\mathcal{H}^a(\mathbb{R}^d) := \Big\{ f \in \mathscr{S}'(\mathbb{R}^d) \;\Big|\; \|f\|_{\mathcal{H}^a(\mathbb{R}^d)} < \infty \Big\},$$

where the norm is given by

$$\|f\|^2_{\mathcal{H}^\alpha(\mathbb{R}^d)} := \int_{\mathbb{R}^d} \big(1 + \|\omega\|_2^2\big)^\alpha \cdot |\mathcal{F}f(\omega)|^2 \, d\omega.$$

From the previous definition, it immediately follows that

$$\|f\|_{\mathcal{H}^{a_1}(\mathbb{R}^d)} \leq \|f\|_{\mathcal{H}^{a_2}(\mathbb{R}^d)} \qquad \text{for all } f \in \mathcal{H}^{a_2}(\mathbb{R}^d),$$

for $a_1 \leq a_2$, so that we have the inclusion $\mathcal{H}^{a_2}(\mathbb{R}^d) \subset \mathcal{H}^{a_1}(\mathbb{R}^d)$ in this case. In particular, we have

$$\mathcal{H}^a(\mathbb{R}^d) \subset \mathcal{H}^0(\mathbb{R}^d) = L^2(\mathbb{R}^d) \qquad \text{for all } a \geq 0.$$

Regarding the multiplication of Sobolev functions, we have used the following result in Corollary 10.9 to show the inclusion of the weighted kernel's native space into a suitable Sobolev space (cf. [20, Theorem 5.1 & 7.3]).

Proposition A.30. *Let $a_1, a_2, \sigma \in \mathbb{R}$ satisfy the following conditions:*

(i) $0 \leq \sigma \leq \min(a_1, a_2)$

(ii) $a_1 + a_2 - \sigma > d/2$

Then, the bilinear multiplication mapping

$$\Pi : \mathcal{H}^{a_1}(\mathbb{R}^d) \times \mathcal{H}^{a_2}(\mathbb{R}^d) \to \mathcal{H}^\sigma(\mathbb{R}^d), \; (f_1, f_2) \mapsto f_1 \cdot f_2$$

is well-defined and continuous, i.e. there is a constant $C > 0$ such that

$$\|f_1 \cdot f_2\|_{H^\sigma(\mathbb{R}^d)} \leq C \cdot \|f_1\|_{H^{a_1}(\mathbb{R}^d)} \cdot \|f_2\|_{H^{a_2}(\mathbb{R}^d)} \qquad \textit{for all } f_1 \in \mathcal{H}^{a_1}(\mathbb{R}^d), f_2 \in \mathcal{H}^{a_2}(\mathbb{R}^d).$$

To end this introduction to Sobolev spaces, we remark that the previous definitions and results can be generalized to more complex domains $\Omega \subset \mathbb{R}^d$.

Sampling inequalities

As already pointed out in [146, page 7], one key advantage of Sobolev spaces is the availability of so-called *sampling inequalities*. Using these inequalities, the Sobolev norm of the interpolation residual can be estimated in terms of the fill distance. We follow the discussion in [143, Section 11.6], [146, Section 2.2] here and state an example (cf. [82, Theorem 2.2]).

Theorem A.31. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain that satisfies an interior cone condition and has a Lipschitz boundary. Moreover, assume that*

(i) $a = k + s$ *with* $k \in \mathbb{N}_0$, $0 \leq s < 1$

(ii) $1 \leq p < \infty$, $1 \leq q \leq \infty$

(iii) $m \in \mathbb{N}_0$ *with* $k > m + d/p$ *if* $p > 1$ *or* $k \geq m + d/p$ *if* $p = 1$.

Then, there is $C > 0$ such that for any discrete set $X \subset \Omega$ with sufficiently small fill distance $h_{X,\Omega}$, we have the norm inequality

$$\|f\|_{W^{m,q}(\Omega)} \leq C \cdot \left(h_{X,\Omega}^{a-m-d\cdot(1/p-1/q)_+} \cdot \|f\|_{W^{a,p}(\Omega)} + h_{X,\Omega}^{-m} \cdot \|f_X\|_\infty \right) \qquad \text{for all } f \in W^{a,p}(\Omega), \quad \text{(A.10)}$$

where $(1/p - 1/q)_+ = \max(1/p - 1/q, 0)$ and $f_X \in \mathbb{R}^{|X|}$ is the restriction of f to X.

If the kernel K satisfies the assumptions of Corollary 3.19 for $a > d/2$, we have $\mathcal{H}_K = W^{a,2}(\Omega)$ when restricting the native space to the bounded domain Ω, where the native space and Sobolev norm are equivalent. Given $f \in \mathcal{H}_K$ and $X \subset \Omega$ with a sufficiently small fill distance, let $I_{K,\Lambda_X}(f) \in S_{K,\Lambda_X}$ denote the standard interpolant to f on X. In this case, we can simply plug in the interpolation residual $f - I_{K,\Lambda_X}(f) \in W^{a,2}(\Omega)$ into the inequality (A.10) and get

$$\|f - I_{K,\Lambda_X}(f)\|_{W^{m,q}(\Omega)} \leq C' \cdot h_{X,\Omega}^{a-m-d\cdot(1/2-1/q)_+} \cdot \|f\|_K$$

for suitable $C' > 0$, where we used that $\|f - I_{K,\Lambda_X}(f)\|_K \leq \|f\|_K$ holds due to the orthogonal projection properties of the interpolation operator, see Theorem 4.10. Note that the last inequality yields estimates for the L^∞ norm in the special case $m = 0$, $q = \infty$.

Besides the application of this technique to the standard interpolation case, it was demonstrated in [146, Section 3] how the norm inequality (A.10) can be used in the context of differential operators. Lastly, we remark that similar sampling inequalities were used in [101, Theorem IV.2.2] to derive general L^2-error estimates in the context of computerized tomography. It remains to be investigated whether these inequalities can be adapted to the analysis of the kernel-based reconstruction from scattered Radon data.

Declaration of Contributions

This declaration highlights the parts of this thesis which are based on collaborative work with other researchers. In particular, the contributions of the involved persons are described. All parts of this thesis that are not mentioned below are due to the candidate with no contribution from other people.

- **Anisotropic product kernels:** The paragraph on *anisotropic product kernels* from Subsection 2.1.3 and the content of Section 3.3 are based on the preprint [6], which is the result of joint research with Juliane Entzian under the supervision of Armin Iske. In particular, the relevant statements of Theorem 2.18, Corollary 2.19, Theorem 2.20, Theorem 3.20 and Theorem 3.25 from this thesis correspond to the statements of Theorem 4.4, Corollary 4.5, Theorem 4.6, Theorem 2.3 and Theorem 3.4 from the preprint [6]. The aforementioned statements including proofs were jointly developed by Juliane Entzian and the candidate in weekly research meetings with equal contributions from both co-authors. The co-author and supervisor Armin Iske provided valuable scientific advice and improved the preprint by editing the initial draft texts.

  In order to properly include the content of [6, Sections 2-4], the candidate provides individual texts and partly changes the notation as well as the ordering of the statements in this thesis. Note that the last three sections of the preprint [6] which discuss the numerical advantages of anisotropic product kernels are not included in this thesis.

- **Convergence results on generalized interpolation:** The analysis of Chapter 5 & 7 is based on the preprint [8], which is a result of the candidate's research on generalized interpolation theory conducted under the supervision of Armin Iske. All results from [8, Section 3 & 4] are due to the candidate and included in this thesis. The co-author and supervisor Armin Iske provided valuable scientific advice and improved the preprint by editing the initial draft texts.

  In comparison to the preprint, the candidate provides individual texts and adds more depth to some parts of the analysis in this thesis. Note that the numerical results from [8, Section 5] are not included. Instead, the candidate provides a new implementation of the involved algorithms and a more sophisticated numerical investigation.

The following comments explain the relation between this thesis and further publications:

- The proceedings paper [7], co-authored by Armin Iske and the candidate, is based on a talk given at *GAMM 2020@21 conference* and contains an initial numerical investigation of greedy data selection algorithms in the context of computerized tomography. Although the respective numerical tests were implemented and conducted by the candidate, no parts of this proceedings paper are incorporated into this thesis. As mentioned before, this thesis contains new and more sophisticated numerical investigations.

- The candidate contributed to the proceedings paper [5], which is joint work with Juliane Entzian and Armin Iske. More precisely, the candidate incorporated the *psr-greedy algorithm* into the reconstruction via different kernels (cf. [5, Subsection 4.6.2]), and therefore only contributed to the numerical investigations (cf. [5, Subsection 4.6]). Since the content of the proceedings paper does not match the purpose of this thesis, it is not included.

List of Publications

During his doctorate, the candidate has contributed to the following research publications:

[5] K. Albrecht, J. Entzian, and A. Iske. Anisotropic Kernels for Particle Flow Simulation. In A. Iske and T. Rung, editors, *Modeling, Simulation and Optimization of Fluid Dynamic Applications*, Lecture Notes in Computational Science and Engineering: Volume 148, pages 57–76. Springer.

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

[7] K. Albrecht and A. Iske. Greedy algorithms for image approximation from scattered Radon data. *PAMM*, 21(1):e202100223, 2021.

[8] K. Albrecht and A. Iske. On the convergence of generalized kernel-based interpolation by greedy data selection algorithms. arXiv preprint: 2407.03840, 2024.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben.

Darüber hinaus versichere ich, dass diese Dissertation nicht in einem früheren Promotionsverfahren eingereicht wurde.

Ort, Datum	Kristof Albrecht

Bibliography

[1] T. Aboiyar, E. H. Georgoulis, and A. Iske. Adaptive ADER methods using kernel-based polyharmonic spline WENO reconstruction. *SIAM Journal on Scientific Computing*, 32(6):3251–3277, 2010.

[2] R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics*. Academic Press, 2nd edition, 2003.

[3] N. Albers. Anwendung des Remez-Algorithmus auf Gaussians. Bachelor's thesis, Universität Hamburg, 2018.

[4] K. Albrecht. Orthogonalsysteme in kernbasierten Approximationsräumen. Master's thesis, Universität Hamburg, 2019.

[5] K. Albrecht, J. Entzian, and A. Iske. Anisotropic Kernels for Particle Flow Simulation. In A. Iske and T. Rung, editors, *Modeling, Simulation and Optimization of Fluid Dynamic Applications*, Lecture Notes in Computational Science and Engineering: Volume 148, pages 57–76. Springer, 2023.

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

[7] K. Albrecht and A. Iske. Greedy algorithms for image approximation from scattered Radon data. *PAMM*, 21(1):e202100223, 2021.

[8] K. Albrecht and A. Iske. On the convergence of generalized kernel-based interpolation by greedy data selection algorithms. arXiv preprint: 2407.03840, 2024.

[9] A. H. Andersen and A. C. Kak. Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1):81–94, 1984.

[10] A.H. Andersen. Algebraic Reconstruction in CT from Limited Views. *IEEE Transactions on Medical Imaging*, 8(1):50–55, 1989.

[11] N. Aronszajn. La théorie des noyaux reproduisants et ses applications. Première partie. *Mathematical Proceedings of the Cambridge Philosophical Society*, 39(3):133–153, 1943.

[12] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

[13] R. Askey. Radial characteristic functions. Technical Report 1262, University of Wisconsin, Mathematics Research Center, 1973.

[14] K. E. Atkinson. *An Introduction to Numerical Analysis*. Wiley, 2nd edition, 1989.

[15] D. O. Baguer, J. Leuschner, and M. Schmidt. Computed tomography reconstruction using deep image prior and learned reconstruction methods. *Inverse Problems*, 36(9):094004, 2020.

[16] R. Beatson, O. Davydov, and J. Levesley. Error bounds for anisotropic RBF interpolation. *Journal of Approximation Theory*, 162(3):512–527, 2010. Special Issue: Bommerholz Proceedings.

[17] R. K. Beatson and W. zu Castell. Scattered data interpolation of Radon data. *Calcolo*, 48:5–19, 2011.

[18] M. Beckmann. *Error Estimates and Convergence Rates for Filtered Back Projection Reconstructions*. PhD thesis, Universität Hamburg, 2018.

[19] A. Beer. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Annalen der Physik und Chemie*, 86:78–88, 1852.

[20] A. Behzadan and M. Holst. Multiplication in Sobolev spaces, revisited. *Arkiv för Matematik*, 59(2):275–306, 2021.

[21] M. Beister, D. Kolditz, and W. A. Kalender. Iterative reconstruction methods in X-ray CT. *Physica Medica*, 28(2):94–108, 2012.

[22] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, 2004.

[23] S. Bochner. *Vorlesung über Fouriersche Integrale*. Mathematik und ihre Anwendungen: Volume 12. Akademische Verlagsgesellschaft, 1932.

[24] S. Bochner. Montone Funktionen, Stieltjes Integrale und harmonische Analyse. *Mathematische Annalen*, 108:378–410, 1933.

[25] J. P. Boyd and K. W. Gildersleeve. Numerical experiments on the condition number of the interpolation matrices for radial basis functions. *Applied Numerical Mathematics*, 61(4):443–459, 2011.

[26] D. S. Broomhead and D. Lowe. Multivariable Functional Interpolation and Adaptive Networks. *Complex Systems*, 2:321–355, 1988.

[27] M. D. Buhmann. *Radial Basis Functions: Theory and Implementations*. Cambridge Monographs on Applied and Computational Mathematics: Volume 12. Cambridge University Press, 2003.

[28] M. H. Buonocore, W. R. Brody, and A. Macovski. A Natural Pixel Decomposition for Two-Dimensional Image Reconstruction. *IEEE Transactions on Biomedical Engineering*, BME-28(2):69–78, 1981.

[29] T. M. Buzug. *Computed Tomography*. Springer, 2008.

[30] G. Casciola, D. Lazzaro, L. B. Montefusco, and S. Morigi. Shape Preserving Surface Reconstruction Using Locally Anisotropic Radial Basis Function Interpolants. *Computers & Mathematics with Applications*, 51(8):1185–1198, 2006.

[31] T. Chang and G. T. Herman. A Scientific Study of Filter Selection for a Fan-Beam Convolution Reconstruction Algorithm. *SIAM Journal on Applied Mathematics*, 39(1):83–105, 1980.

[32] W. Cheney and W. Light. *A Course in Approximation Theory*. Graduate Studies in Mathematics: Volume 101. American Mathematical Society, 2009.

[33] A. M. Cormack. Representation of a Function by Its Line Integrals, with Some Radiological Applications. *Journal of Applied Physics*, 34(9):2722–2727, 1963.

[34] A. M. Cormack. Representation of a Function by Its Line Integrals, with Some Radiological Applications. II. *Journal of Applied Physics*, 35(10):2908–2913, 1964.

[35] P. C. Curtis Jr. n-parameter families and best approximation. *Pacific Journal of Mathematics*, 9(4):1013–1027, 1959.

[36] C. De Boor. *A Practical Guide to Splines*. Applied Mathematical Sciences: Volume 27. Springer, Revised edition, 2001.

[37] S. De Marchi. On optimal center locations for radial basis function interpolation: computational aspects. *Rendiconti del Seminario Matematico*, 61(3):343–358, 2003.

[38] S. De Marchi, A. Iske, and G. Santin. Image reconstruction from scattered Radon data by weighted positive definite kernel functions. *Calcolo*, 55(2), 2018.

[39] S. De Marchi, A. Iske, and A. Sironi. Kernel-based Image Reconstruction from Scattered Radon Data. *Dolomites Research Notes on Approximation*, 9(2):19–31, 2016.

[40] S. De Marchi, F. Marchetti, and E. Perracchione. Jumping with variably scaled discontinuous kernels (VSDKs). *BIT Numerical Mathematics*, 60:441–463, 2020.

[41] S. De Marchi, R. Schaback, and H. Wendland. Near-optimal data-independent point locations for radial basis function interpolation. *Advances in Computational Mathematics*, 23:317–330, 2005.

[42] R. DeVore, G Petrova, and P. Wojtaszczyk. Greedy Algorithms for Reduced Bases in Banach Spaces. *Constructive Approximation*, 37:455–466, 2013.

[43] S. Dutta, M. W. Farthing, E. Perracchione, G. Savant, and M. Putti. A greedy non-intrusive reduced order model for shallow water equations. *Journal of Computational Physics*, 439:110378, 2021.

[44] N. Dyn, M. S. Floater, and A. Iske. Adaptive thinning for bivariate scattered data. *Journal of Computational and Applied Mathematics*, 145(2):505–517, 2002.

[45] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications: Volume 375. Kluwer Academic Publishers, 2000.

[46] M. Esmaeilbeigi, O. Chatrabgoun, and M. Cheraghi. The Role of Hilbert-Schmidt SVD basis in Hermite-Birkhoff interpolation in fractional sense. *Computational and Applied Mathematics*, 38, 2019.

[47] G. E. Fasshauer. Solving partial differential equations by collocation with radial basis functions. In A. Le Méhauté, C. Rabut, and L. L. Schumaker, editors, *Surface Fitting and Multiresolution Methods*, volume 1997, pages 131–138. Vanderbilt University Press, 1997.

[48] G. E. Fasshauer. *Meshfree Approximation Methods using MATLAB*. Interdisciplinary Mathematical Sciences: Volume 6. World Scientific, 2007.

[49] G. E. Fasshauer. Positive definite kernels: past, present and future. *Dolomites Research Notes on Approximation*, 4(2):21–63, 2011.

[50] G. E. Fasshauer and M. McCourt. *Kernel-based Approximation Methods using MATLAB*. Interdisciplinary Mathematical Sciences: Volume 19. World Scientific, 2015.

[51] G. E. Fasshauer and M. J. McCourt. Stable Evaluation of Gaussian Radial Basis Function Interpolants. *SIAM Journal on Scientific Computing*, 34(2):A737–A762, 2012.

[52] T. G. Feeman. *The Mathematics of Medical Imaging*. Springer Undergraduate Texts in Mathematics and Technology. Springer, 2nd edition, 2015.

[53] C. Franke and R. Schaback. Convergence order estimates of meshless collocation methods using radial basis functions. *Advances in Computational Mathematics*, 8:381–399, 1998.

[54] C. Franke and R. Schaback. Solving partial differential equations by collocation using radial basis functions. *Applied Mathematics and Computation*, 93(1):73–82, 1998.

[55] P. Gilbert. Iterative Methods for the Three-dimensional Reconstruction of an Object from Projections. *Journal of Theoretical Biology*, 36(1):105–117, 1972.

[56] R. Gordon, R. Bender, and G. T. Herman. Algebraic Reconstruction Techniques (ART) for Three-dimensional Electron Microscopy and X-ray Photography. *Journal of theoretical Biology*, 29(3):471–481, 1970.

[57] V. Griem. Weighted Positive Definite Kernels. Master's thesis, Universität Hamburg, 2017.

[58] A. Haar. Die Minkowskische Geometrie und die Annäherung an stetige Funktionen. *Mathematische Annalen*, 78:294–311, 1917.

[59] C. Hamaker and D. C. Solmon. The Angles between the Null spaces of X Rays. *Journal of Mathematical Analysis and Applications*, 62(1):1–23, 1978.

[60] S Helgason. *The Radon Transform*. Progress in Mathematics: Volume 5. Birkhäuser, 2nd edition, 1999.

[61] G. T. Herman. *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*. Advances in Computer Vision and Pattern Recognition. Springer, 2nd edition, 2009.

[62] G. T. Herman and A. Lent. Iterative reconstruction algorithms. *Computers in Biology and Medicine*, 6(4):273–294, 1976.

[63] G. T. Herman and L. B. Meyer. Algebraic Reconstruction Techniques Can Be Made Computationally Efficient. *IEEE Transactions on Medical Imaging*, 12(3):600–609, 1993.

[64] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[65] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2013.

[66] G. N Hounsfield. Computerized transverse axial scanning (tomography): Part I. Description of system. *British Journal of Radiology*, 46(552):1016–1022, 1973.

[67] Lars Hörmander. *The Analysis of Linear Partial Differential Operators I*. Classics in Mathematics. Springer, 2003.

[68] A. Iske. Reconstruction of Functions From Generalized Hermite-Birkhoff Data. In C. K. Chui and L. L. Schumaker, editors, *Approximation Theory VIII: Approximation and Interpolation (Volume 1)*, Series in Approximations and Decompositions: Volume 6, pages 257–264. World Scientific, 1995.

[69] A. Iske. *Multiresolution Methods in Scattered Data Modelling*. Lecture Notes in Computational Science and Engineering: Volume 37. Springer, 2004.

[70] A. Iske. *Approximation Theory and Algorithms for Data Analysis*. Texts in Applied Mathematics: Volume 68. Springer, 2018.

[71] A. Iske and T. Sonar. On the structure of function spaces in optimal recovery of point functionals for eno-schemes by radial basis functions. *Numerische Mathematik*, 74:177–201, 1996.

[72] M. Jiang and G. Wang. Convergence of the Simultaneous Algebraic Reconstruction Technique (SART). *IEEE Transactions on Image Processing*, 12(8):957–961, 2003.

[73] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[74] S. I. Kabanikhin. Definitions and examples of inverse and ill-posed problems. *Journal of Inverse and Ill-posed Problems*, 16(4):317–357, 2008.

[75] S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bulletin International de l'Académie Polonaise des Sciences et des Lettres. Classe des Sciences Mathématiques et Naturelles*, 35:355–357, 1937.

[76] R. V. Kadison and J. R. Ringrose. *Fundamentals of the Theory of Operator Algebras - Volume I: Elementary Theory*. Pure and Applied Mathematics: Volume 100. Academic Press, 1983.

[77] J. P. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Applied Mathematical Sciences: Volume 160. Springer, 2005.

[78] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. Classics in Applied Mathematics: Volume 33. Society for Industrial and Applied Mathematics, 2001.

[79] E. J. Kansa. Multiquadrics — A scattered data approximation scheme with applications to computational fluid-dynamics — II: Solutions to parabolic, hyperbolic and elliptic partial differential equations. *Computers & Mathematics with applications*, 19(8-9):147–161, 1990.

[80] V. Kolehmainen, S. Siltanen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo. Statistical inversion for medical x-ray tomography with few radiographs: II. Application to dental radiology. *Physics in Medicine & Biology*, 48(10):1465–1490, 2003.

[81] S. Le Borne. Factorization, Symmetrization, and Truncated Transformation of Radial Basis Function-GA Stabilized Gaussian Radial Basis Functions. *SIAM Journal on Matrix Analysis and Applications*, 40(2):517–541, 2019.

[82] Q. T. Le Gia, F. J. Narcowich, J. D. Ward, and H. Wendland. Continuous and discrete least-squares approximation by radial basis functions on spheres. *Journal of Approximation Theory*, 143(1):124–133, 2006.

[83] R. M. Lewitt. Multidimensional digital image representations using generalized Kaiser–Bessel window functions. *Journal of the Optical Society of America A*, 7(10):1834–1846, 1990.

[84] R. M. Lewitt. Alternatives to voxels for image representation in iterative reconstruction algorithms. *Physics in Medicine & Biology*, 37(3):705–716, 1992.

[85] A.K. Louis and F. Natterer. Mathematical problems of computerized tomography. *Proceedings of the IEEE*, 71(3):379–389, 1983.

[86] Y. Maday, O. Mula, A. T. Patera, and M. Yano. The Generalized Empirical Interpolation Method: Stability theory on Hilbert spaces with an application to the Stokes equation. *Computer Methods in Applied Mechanics and Engineering*, 287:310–334, 2015.

[87] W. R. Madych and S. A. Nelson. Multivariate interpolation: A variational theory. Unpublished manuscript, Iowa State University, 1983.

[88] W. R. Madych and S. A. Nelson. Multivariate interpolation and conditionally positive definite functions. *Approximation Theory and its Applications*, 4:77–89, 1988.

[89] W. R. Madych and S. A. Nelson. Multivariate interpolation and conditionally positive definite functions II. *Mathematics of Computation*, 54:211–230, 1990.

[90] W. R. Madych and S. A. Nelson. Bounds on Multivariate Polynomials and Exponential Error Estimates for Multiquadric Interpolation. *Journal of Approximation Theory*, 70(1):94–114, 1992.

[91] J. C. Mairhuber. On Haar's theorem concerning Chebysheff problems having unique solutions. *Proceedings of the American Mathematical Society*, 7(4):609–615, 1956.

[92] R. Marabini, G. T. Herman, and J. M. Carazo. 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs). *Ultramicroscopy*, 72(1-2):53–65, 1998.

[93] W. McLean. *Strongly Elliptic Systems and Boundary Integral Equations*. Cambridge University Press, 2000.

[94] J. Mercer. Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 209:415–446, 1909.

[95] D. Micieli, T. Minniti, V. Formoso, W. Kockelmann, and G. Gorini. A comparative study of reconstruction methods applied to Neutron Tomography. *Journal of Instrumentation*, 13(6):C06006, 2018.

[96] E. H. Moore. On properly positive Hermitian matrices. *Bulletin of the American Mathematical Society*, 23, 1916.

[97] E. H. Moore. *General Analysis, Part I.* Memoirs of the American Philosophical Society: Volume 1. The American Philosophical Society, 1935.

[98] S. Müller. *Komplexität und Stabilität von kernbasierten Rekonstruktionsmethoden*. PhD thesis, Georg-August-Universität Göttingen, 2008.

[99] S. Müller and R. Schaback. A Newton basis for Kernel spaces. *Journal of Approximation Theory*, 161(2):645–655, 2009.

[100] F. J. Narcowich and J. D. Ward. Generalized Hermite interpolation via matrix-valued conditionally positive definite functions. *Mathematics of Computation*, 63(208):661–687, 1994.

[101] F. Natterer. *The Mathematics of Computerized Tomography*. Classics in Applied Mathematics: Volume 32. Society for Industrial and Applied Mathematics, 2001.

[102] F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction*. SIAM Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics, 2001.

[103] J. Neveu. *Processus aléatoires gaussiens*. Séminaire de Mathématiques Supérieures. Les presses de l'Université de Montréal, 1968.

[104] M. Pazouki. *Stable Bases for Kernel Based Methods*. PhD thesis, Georg-August-Universität Göttingen, 2012.

[105] M. Pazouki and R. Schaback. Bases for kernel-based spaces. *Journal of Computational and Applied Mathematics*, 236(4):575–588, 2011.

[106] A. Pinkus. *n-Widths in Approximation Theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete - 3. Folge: Volume 7. Springer, 1985.

[107] L. Pronzato and A. Zhigljavsky. Quasi-uniform designs with optimal and near-optimal uniformity constant. arXiv preprint: 2112.10401, 2021.

[108] E. T. Quinto. Exterior and limited-angle tomography in non-destructive evaluation. *Inverse Problems*, 14(2):339–353, 1998.

[109] J. Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften*, 69:262–277, 1917.

[110] G. N. Ramachandran and A. V. Lakshminarayanan. Three-dimensional Reconstruction from Radiographs and Electron Micrographs: Application of Convolutions instead of Fourier Transforms. *Proceedings of the National Academy of Sciences*, 68(9):2236–2240, 1971.

[111] A. Rieder and A. Faridani. The Semidiscrete Filtered Backprojection Algorithm Is Optimal for Tomographic Inversion. *SIAM Journal on Numerical Analysis*, 41(3):869–892, 2003.

[112] Y. Saatchi. *Scalable Inference for Structured Gaussian Process Models*. PhD thesis, University of Cambridge, 2011.

[113] G. Santin. VKOGA. GitHub repository: `https://github.com/GabrieleSantin/VKOGA`.

[114] G. Santin and B. Haasdonk. Convergence rate of the data-independent P-greedy algorithm in kernel-based approximation. *Dolomites Research Notes on Approximation*, 10(2):68–78, 2017.

[115] G. Santin, D. Wittwar, and B. Haasdonk. Greedy regularized kernel interpolation. arXiv preprint: 1807.09575, 2018.

[116] R. Schaback. Error estimates and condition numbers for radial basis function interpolation. *Advances in Computational Mathematics*, 3:251–264, 1995.

[117] R. Schaback. Native Hilbert Spaces for Radial Basis Functions I. In M. W. Müller, M. D. Buhmann, D. H. Mache, and M. Felten, editors, *New Developments in Approximation Theory*, International Series of Numerical Mathematics: Volume 132, pages 255–282. Birkhäuser, 1999.

[118] R. Schaback. A Greedy Method for Solving Classes of PDE problems. arXiv preprint: 1903.11536, 2019.

[119] R. Schaback and H. Wendland. Adaptive greedy techniques for approximate solution of large RBF systems. *Numerical Algorithms*, 24:239–254, 2000.

[120] R. Schaback and J. Werner. Linearly constrained reconstruction of functions by kernels with applications to machine learning. *Advances in Computational Mathematics*, 25:237–258, 2006.

[121] I. J. Schoenberg. Metric Spaces and Completely Monotone Functions. *Annals of Mathematics*, 39(4):811–841, 1938.

[122] B. Schölkopf, R. Herbrich, and A. J. Smola. A Generalized Representer Theorem. In D. Helmbold and B. Williamson, editors, *Computational Learning Theory. COLT 2001*, Lecture Notes in Computer Science: Volume 2111, pages 416–426. Springer, 2001.

[123] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation and Machine Learning. MIT Press, 2002.

[124] L. A. Shepp and B. F. Logan. The Fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science*, 21(3):21–43, 1974.

[125] S. Siltanen, V. Kolehmainen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo. Statistical inversion for medical x-ray tomography with few radiographs: I. General theory. *Physics in Medicine & Biology*, 48(10):1437–1463, 2003.

[126] A. Sironi. Medical Image Reconstruction using Kernel Based Methods. Master's thesis, University of Padova, 2011. arXiv: 1111.5844.

[127] K. T. Smith, D. C. Solmon, and S. L. Wagner. Practical and mathematical aspects of the problem of reconstructing objects from radiographs. *Bulletin of the American Mathematical Society*, 83(6):1227–1270, 1977.

[128] T. Sonar. Optimal recovery using thin plate splines in finite volume methods for the numerical solution of hyperbolic conservation laws. *IMA Journal of Numerical Analysis*, 16:549–581, 1996.

[129] T. Sonar. *Mehrdimensionale ENO-Verfahren*. Advances in Numerical Mathematics. Teubner, 1997.

[130] T. Sonar. On Families of Pointwise Optimal Finite Volume ENO Approximations. *SIAM Journal on Numerical Analysis*, 35:2350–2369, 1998.

[131] E. M. Stein and G. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton Mathematical Series. Princeton University Press, 1971.

[132] I. Steinwart and A. Christmann. *Support Vector Machines*. Information Science and Statistics. Springer, 2008.

[133] T. Strohmer and R. Vershynin. A Randomized Solver for Linear Systems with Exponential Convergence. In J. Díaz, K. Jansen, J. D. P. Rolim, and U. Zwick, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques. APPROX, RANDOM 2006*, Lecture Notes in Computer Science: Volume 4110, pages 499–507. Springer, 2006.

[134] K. Tanabe. Projection Method for Solving a Singular System of Linear Equations and its Applications. *Numerische Mathematik*, 17:203–214, 1971.

[135] V. N. Temlyakov. Greedy approximation. *Acta Numerica*, 17:235–409, 2008.

[136] A. I. Tolstykh and D. A. Shirobokov. On using radial basis functions in a "finite difference mode" with applications to elasticity problems. *Computational Mechanics*, 33:68–79, 2003.

[137] G. Wahba. *Spline Models for Observational Data*. CBMS-NSF Regional Conference series in appliedmathematics: Volume 59. Society for Industrial and Applied Mathematics, 1990.

[138] W. Walter. *Einführung in die Theorie der Distributionen*. BI-Wissenschaftsverlag, 1994.

[139] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[140] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4:389–396, 1995.

[141] H. Wendland. Error Estimates for Interpolation by Compactly Supported Radial Basis Functions of Minimal Degree. *Journal of Approximation Theory*, 93(2):258–272, 1998.

[142] H. Wendland. On the Convergence of a General Class of Finite Volume Methods. *SIAM Journal on Numerical Analysis*, 43:987–1002, 2005.

[143] H. Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics: Volume 17. Cambridge University Press, 2005.

[144] T. Wenzel, G. Santin, and B. Haasdonk. A novel class of stabilized greedy kernel approximation algorithms: Convergence, stability and uniform point distribution. *Journal of Approximation Theory*, 262, 2021.

[145] T. Wenzel, G. Santin, and B. Haasdonk. Analysis of Target Data-Dependent Greedy Kernel Algorithms: Convergence Rates for f-, $f \cdot P$- and f/P-Greedy. *Constructive Approximation*, 57:45–74, 2023.

[146] T. Wenzel, D. Winkle, G. Santin, and B. Haasdonk. Adaptive meshfree solution of linear partial differential equations with PDE-greedy kernel methods. arXiv preprint: 2207.13971, 2022.

[147] D. Werner. *Funktionalanalysis*. Springer-Lehrbuch. Springer, 8th edition, 2018.

[148] T. Wiatowski. Kernel Based Image Reconstruction from Spherical Radon Data. Master's thesis, Technische Universität München, 2012.

[149] D. Wirtz and B. Haasdonk. A vectorial kernel orthogonal greedy algorithm. *Dolomites Research Notes on Approximation*, 6(2):83–100, 2013.

[150] D. Wittwar. *Approximation with matrix-valued kernels and highly effective error estimators for reduced basis approximations*. PhD thesis, Universität Stuttgart, 2022.

[151] Z. Wu. Hermite-Birkhoff interpolation of scattered data by radial basis functions. *Approximation Theory and its Applications*, 8:1–10, 1992.

[152] Z. Wu and R. Schaback. Local error estimates for radial basis function interpolation of scattered data. *IMA Journal of Numerical Analysis*, 13(1):13–27, 1993.

[153] K. Yao. Applications of Reproducing Kernel Hilbert Spaces – Bandlimited Signal Models. *Information and Control*, 11(4):429–444, 1967.

[154] K. Yosida. *Functional Analysis*. Classics in Mathematics. Springer, 1995.