Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Matematica

_____

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE MATEMATICHE
INDIRIZZO MATEMATICA COMPUTAZIONALE
CICLO XXVIII

Approximation in kernel-based spaces, optimal subspaces and approximation of eigenfunctions

Direttore della Scuola : Ch.mo Prof. Pierpaolo Soravia
Coordinatore d'indirizzo: Ch.mo Prof. Michela Redivo Zaglia
Supervisore : Ch.mo Prof. Stefano De Marchi

Dottorando: Gabriele Santin

ABSTRACT

Kernel-based approximation methods provide optimal recovery procedures in the *native* Hilbert spaces in which they are reproducing. Among other, kernels in the notable class of continuous and strictly positive definite kernels on compact sets possess a series decomposition in L_2 - orthonormal eigenfunctions of a particular integral operator.

The interest for this decomposition is twofold. On one hand, the subspaces generated by eigenfunctions, or *eigenbasis* elements, are L_2-optimal trial spaces in the sense of *widths*. On the other hand, such expansion is the fundamental tool of some of the state of the art algorithms in kernel approximation. Despite these reasons motivate a great interest in the eigenbasis, for a given kernel this decomposition is generally completely unknown.

In this view, this thesis faces the problem of approximating the eigenbasis of general continuous and strictly positive definite kernels on general compact sets of the Euclidean space, for any space dimension.

We will at first define a new kind of optimality that is based on a error measurement closer to the one of standard kernel interpolation. This new width is then analyzed, and we will determine its value and characterize its optimal subspaces, which are spanned by the eigenbasis. Moreover, this optimality result is suitable to be scaled to some particular subspace of the native space, and this restriction allows us to prove new results on the construction of computable optimal trial spaces. This situation covers the standard case of point-based interpolation, and will provide algorithms to approximate the eigenbasis by means of standard kernel techniques. On the basis of the new theoretical results, asymptotic estimates on the convergence of the method will be proven. These computations will be translated into effective algorithms, and we will test their behavior in the approximation of the eigenspaces. Moreover, two applications of kernel-based methods will be analyzed.

Riassunto

I metodi kernel forniscono procedure di miglior approssimazione negli spazi di Hilbert *nativi*, ovvero gli spazi in cui tali kernel sono *reproducing kernel*. Nel caso notevole di kernel continui e strettamente definiti positivi su insiemi compatti, è nota l'esistenza di una decomposizione in una serie data dalle autofunzioni (ortonormali in L_2) di un particolare operatore integrale.

L'interesse per questa espansione è motivata da due ragioni. Da un lato, i sottospazi generati dalle autofunzioni, o elementi della *eigenbasis*, sono i trial space L_2-ottimali nel senso delle *widhts*. D'altro canto, tale espansione è lo strumento fondamentale alla base in alcuni degli algoritmi di riferimento utilizzati nell'approssimazione con kernel. Nonostante queste ragioni motivino decisamente l'interesse per le eigenbasis, la suddetta decomposizione è generalmente sconosciuta.

Alla luce di queste motivazioni, la tesi affronta il problema dell'approssimazione delle eigenbasis per generici kernel continui e strettamente definiti positivi su generici insiemi compatti dello spazio euclideo, per ogni dimensione.

Inizieremo col definire un nuovo tipo di ottimalità basata sulla misura dell'errore tipica dell'interpolazione kernel standard. Il nuovo concetto di width sarà analizzato, ne sarà calcolato il valore e caratterizzati i rispettivi sottospazi ottimali, che saranno generati dalla eigenbasis. Inoltre, questo risultato di ottimalità risulterà essere adatto ad essere ristretto ad alcuni particolari sottospazi dello spazio nativo. Questa restrizione ci permetterá di dimostrare nuovi risultati sulla costruzione di trial space ottimali che siano effettivamente calcolabili. Questa situazione include anche il caso dell'interpolazione kernel basata su valutazioni puntuali, e fornirà algoritmi per approssimare le autofunzioni tramite metodi kernel standard. Forniremo inoltre stime asintotiche di convergenza del metodo basate sui nuovi risultati teorici. I metodi presentati saranno implementati in algoritmi numerici, e ne testeremo il comportamento nell'approssimazione degli autospazi. Infine analizzeremo l'applicazioni dei metodi kernel a due diversi problemi di approssimazione.

ACKNOWLEDGMENTS

First and foremost, I wish to thank my Advisor Prof. Stefano De Marchi, who introduced me to this fascinating topic and pushed me toward interesting questions and problems. He has been a mentor in a broad sense, and he gave me many great opportunities.

My gratitude goes also to all the members of the research group *Constructive Approximation and Applications* at the University of Padova and Verona. Since the writing of my Bachelor thesis I found there an exciting and open environment, always committed to motivating and promoting young students.

The central part of this thesis has been written thanks to the supervision of Prof. Robert Schaback. Working with him has been a pleasure and an honor.

I would like to thank Prof. Gregory Fasshauer, who accepted to be the Referee of this thesis, and whose comments and suggestions on the first draft contributed to improve this work.

Many thanks also to the members of my new research group *Numerical Mathematics* at the University of Stuttgart, and especially to Prof. Bernard Haasdonk, who gave me all the time I needed to finalize this thesis.

Not mentioning everything my family did and do for me, my parents Maria Antonia and Angelo teached and always encouraged me to follow my aspirations, while putting them in the right perspective. Special thanks to them, and to my brother and sisters Caterina, Francesco e Cecilia, with Laura, Tobia and Riccardo.

I wish to thank also Emilia, Lucia and Fulvio, who embraced me as in a second family.

The years in Padova would not have had the same meaning without my friends and colleagues Alessio, Anna, Andrea, Cristina, Daniele, Daria, Davide, Davide, Federico, Francesco, Giovanni, Laura, Luigi, Luisa, Mariano, Marta, Matteo, Michele and Valentina, with whom I shared countless beautiful moments. Many thanks to each of them, and in particular to Laura for her priceless friendship, and to Federico, who shared with me his passion for Mathematics, and who often believed in me much more than how I did by myself.

Finally and above all, thanks to Maria.

Contents

A Anita, Arcangelo,
Elide e Appollonio.

Introduction

Kernel methods are successfully employed in various fields and applications. In Machine Learning they are the basis of methods such as *support vector machines* and *RBF-neural networks*, which are used to solve complex classification and regression problems. In Geosciences they appear as the method of *kriging*, which provides approximations of minimal expected deviation of random fields. More generally in Statistics, kernels are strictly related to covariance functions, and kernel-based methods are used to produce optimal estimators of stochastic processes. Moreover, kernels are used in Optimization and Geometric Modeling.

Among the various fields of application, of notable interest is the problem of approximation of functions, where kernel-based methods are extensively applied thanks to their flexibility. Namely, through kernels it is possible to treat scattered-data problems, problems defined in many space dimensions and problems defined over manifolds. This feature is a clear advantage over other methods, such as polynomial approximation. Moreover, all the approximation methods that arise from the use of kernels have an essential functional structure in common, and, to some extent, they can be analyzed independently from the particular situation.

Beyond the recovery from point-wise samples, in kernel-based spaces it is also possible to consider *generalized interpolation* problems defined by *weak data*, i.e., approximation problems where the data come from the evaluation of general functionals on a unknown function. The application of these techniques to the solution of partial differential equations leads to the construction of *meshless* algorithms, and to the solution of PDE problems that are hardly faced by methods such as *finite elements*.

Moreover, and more important, kernel-based methods constitute the error optimal recovery strategy in the *native* Hilbert spaces where they are *reproducing*. Such spaces are fairly general, and they can all be treated in the same way by means of their kernel. This optimality is of particular interest for some kernels, whose associated native spaces can be proven to be equivalent to certain Sobolev spaces.

These methods can deal with scattered data, and they produce mean-

1

ingful approximations from completely general samplings. Nevertheless, the success of the recovering deeply depends on the choice of the structure of the samplings, and poor data lead to a poor recovery. In this view, it is of great interest the problem of finding, characterizing and constructing good or even optimal *trial spaces*, to be used both for interpolation and generalized interpolation. The quality of these trial spaces is measured in different way, and we will concentrate in this thesis on *error optimal* subspaces.

Moreover, it is known that the kernel approximation process is both *stable* and *convergent* (see [18]). Nevertheless, it has been observed throughout the literature on kernel methods that a plain, or *direct*, application of kernel-based interpolation methods often leads to unsatisfactory results and to an unstable computation of the approximants. This awareness has led in the recent year to move the attention of the research on the development of alternative computation methods. They are designed to make stable computations in kernel-based spaces, and to obtain the theoretically expected results by removing the limitations due to the wrong use of kernel techniques.

In both these contexts, it has become clear that a special role is played by a particular decomposition of the kernel, due to J. Mercer [39]. This decomposition comes from the spectral decomposition of an integral operator defined through the kernel, and it defines a basis of the native space, named the *eigenbasis*.

In particular, the subspaces generated by the eigenbasis have been proven to provide error optimal trial spaces (see [54]), whenever the error is measured by L_2 norms. This result constitutes a theoretical foundation for the interest in the eigenbasis, and it is deeply connected to the theory of *widths* in Approximation Theory ([46]). Furthermore, the eigenbasis is the fundamental tool of many of the techniques developed in the recent years (see e.g. [12, 24, 26, 28, 37]), and the new algorithms are in fact able to solve recovering problems with extremely high precision, when the decomposition is explicitly known.

Despite these reasons motivate a great interest in the eigenbasis, for a given kernel the Mercer decomposition is generally completely unknown.

In this view, the present thesis faces the problem of approximating the eigenbasis of general *strictly positive definite* kernels on general compact sets of the Euclidean space, for any space dimension.

We will at first define a new kind of optimality that is strictly connected to other widths, but which is instead based on a error measurement closer to standard kernel interpolation. This new width is then analyzed, and we will find its value and characterize its optimal subspaces, which are again given by the spaces spanned by the eigenbasis, in accordance to the aforementioned results.

Moreover, this optimality result turns out to be suitable to be scaled to some particular subspace of the native space. This restriction, on the contrary of the case of other widths, allows to prove new results on the construction of actually computable optimal trial spaces.

This situation covers the standard case of *point-based* interpolation, and will provide algorithms to approximate the eigenbasis by means of standard kernel techniques. On the basis of the new theoretical results, asymptotic estimates on the convergence of the method will be proven.

These computations will be translated into effective algorithms, and we will test their behavior in the approximation of the eigenspaces.

This thesis is organized as follows. Chapter 1 is intended to review some preliminary fact in the theory of kernel-based approximation and introduce the notation used throughout this work. In particular, we recall the definition of a *kernel* and the construction of its *native space*, and we discuss the approximation process in this space. Moreover, we recall the definition of the *eigenbasis*, which will play a central role in the following Chapters. This Chapter is not intended to be an introduction to kernel-based approximation but only to set the background of the forthcoming results.

Chapter 2 contains the main theoretical results of this thesis. At first we generalize the approximation method of the first Chapter to the case of general subspaces of the native space. Then, we state the problem of determining error-optimal subspaces and discuss its framework in the field of Approximation Theory. A new kind of *width* will be introduced, and we will prove that it is minimized by the eigenbasis, in accordance to other known results. This new result allow to consider also a constrained optimization problem, and in this way we are able to take into account the usual situation of kernel-based interpolation. Moreover, some result are proven about the approximation of the eigenbasis trough *point-based* techniques. We conclude this Chapter by presenting two partial extension of this work to the case of different error measurements, and to the case of the recovery of the solution of linear functional equations.

In Chapter 3 we move our attention to the algorithmic counterpart of these results. First, thanks to the analysis of closed subspaces of the previous Chapter, it is possible to construct optimal subspaces by Linear Algebra techniques, and to use these subspaces as an approximation of the eigenbasis. Second, a direct discretization of the integral equation defining the eigenbasis leads to another algorithm, that is discussed in details. In fact, the two algorithms have been developed separately, but we prove here that they are closely related both from a theoretical and a computational point of view, when a certain discretization of the L_2 inner product is used.

Although these two algorithms give a first insight in the actual computation of the eigenbasis, they are not suitable for a direct use. Instead, two

approximated version of both of them are presented in Chapter 4. They are designed to avoid the direct computations of the previous Chapter, which are replaced by greedy techniques and by projection methods. The two algorithms of this Chapter are now ready to be implemented, and we will test them to confirm some of the theoretical results developed so far.

Chapter 5 is devoted to present two different applications of kernel-based approximation. The first one couples one of the algorithms developed in this thesis to a fast domain decomposition technique, and the resulting method is able to solve large approximation problems in a fast and stable way. The second application, instead, is quite independent from the rest of the thesis, and analyses the recovery of medical images by means of kernel techniques.

In Appendix A we set up some further detail on the relation between different native spaces, that will be useful to derive one of the results of the previous Chapters.

Most of the contents of this thesis is published or accepted for publication. In particular, the results of Chapter 2 are published in [52] except for the extensions of the last Section. Chapter 3 presents some of the contents of [16] and [52], while the connection between the two algorithms is a new result. The algorithms of Chapter 4 are published in [52] and in [17]. Generalizations of the contents of Chapter 5 are the object of the two papers [8] and [14], which are currently in preparation, while the present version of the first method is published in [9].

All the algorithms introduced and tested in this thesis have been implemented as Matlab programs, and they can be freely downloaded from [49–51], except for the first algorithm of Chapter 5, which is partially still under development.

Chapter 1

Preliminaries

In this Chapter we will give a brief introduction to the field of *kernel-based approximation*.

The aim of this part of the thesis is to set up the ground floor where the following results are rooted, and to clarify the notation used throughout this work. In this view, we will take into account only the results that are relevant to understand what follows, while plenty of results and approaches are not discussed. The extensive study of the theory of kernels and of reproducing kernel Hilbert spaces partially presented in the first two Sections dates back to the seminal paper [1]. A detailed discussion of the theory and practice of kernel-based approximation can be retrieved from the monographs [5, 22, 23, 60] and from the survey papers [4, 56].

We will concentrate here only on the basic facts that constitute the fundamental tools of kernel-based approximation. In the subsequent Chapters, we will recall some more advanced results in the context where they are needed.

The following results will be mainly presented according to the fashion and the notation of [22] and [60].

1.1   Kernels and kernel-based spaces

We start by defining the main object of this thesis, i.e., the kernel K.

Definition 1.1. *Let $\Omega \subset \mathbb{R}^d$ be a nonempty set. A positive definite kernel K on Ω is a positive definite and symmetric function $K : \Omega \times \Omega \to \mathbb{R}$.*

Positive definiteness is understood in the sense that, for any natural number $N \in \mathbb{N}$, for any vector of coefficients $c \in \mathbb{R}^N$ and for any set $X_N = \{x_1, \ldots, x_N\}$ of N pairwise distinct elements in Ω, we have

$$\sum_{i,j=1}^{N} c_i c_j K(x_i, x_j) \geq 0. \tag{1.1}$$

5

This is the same as requiring that the $N \times N$ kernel matrix A,

$$A_{ij}(X_N) = K(x_i, x_j), \ 1 \leq i, j \leq N,$$

is positive semidefinite for any N and $X_N \subset \Omega$.

We will consider in this thesis only strictly positive definite kernels (hence positive definite kernel matrices), i.e., equality in (1.1) holds if and only if $c \equiv 0$. For this reasons we will simply use the term *kernel* to mean a strictly positive definite kernel. Moreover, even if kernels can be defined on quite general sets, we will consider in the following only the case $\Omega \subset \mathbb{R}^d$.

In many applications, the most used kernels are *translational invariant* or even *radial*. Namely, there exist a univariate function $\Phi : \mathbb{R}^d \to \mathbb{R}$ such that

$$K(x, y) = \Phi(x - y), \ x, y \in \Omega$$

or $\Phi : \mathbb{R}_{\geq 0} \to \mathbb{R}$ such that

$$K(x, y) = \Phi(\|x - y\|_2), \ x, y \in \Omega$$

where $\| \cdot \|_2$ denotes the Euclidean norm on $\mathbb{R}^d$. Radial kernels, which take the name of *Radial Basis Functions* (RBFs), are usually defined up to a factor $\varepsilon > 0$, which is called the *shape parameter* and is employed to control the scale of the kernel, i.e., $K(x, y) = \Phi(\varepsilon \|x - y\|_2)$. Figure 1.1 shows the shape of the commonly used Gaussian kernel $K(x, y) = e^{-\varepsilon^2 \|x - y\|_2^2}$ on $\Omega = [-1, 1]$, for various values of ε.
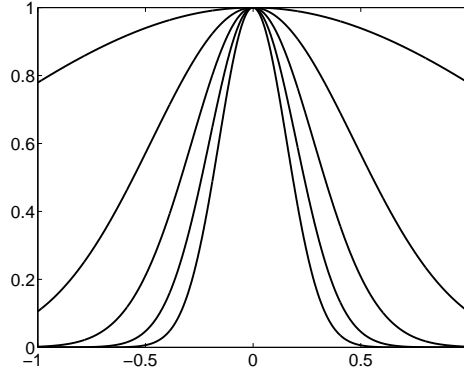
Figure 1.1: The Gaussian kernel on $[-1, 1]$ for different values of ε.

Associated with the kernel K there is a unique *native space* $\mathcal{H}(\Omega)$, that is a separable Hilbert space of functions $f : \Omega \to \mathbb{R}$ where K is the *reproducing kernel* in the following sense.

Definition 1.2. *A kernel K is a reproducing kernel of the Hilbert space $(H, (\cdot, \cdot)_H)$ of real valued functions defined on $\Omega \subset \mathbb{R}^d$ if the following conditions hold:*

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

(ii) $(f, K(\cdot, x))_H = f(x)$ *for all* $x \in \Omega$, $f \in H$.

A Hilbert space possessing a reproducing kernel is called a reproducing kernel Hilbert space (RKHS), and any RKHS has a unique kernel.

The *reproducing property* (ii) is equivalent to state that, for $x \in \Omega$, the x-translate of the kernel, $K(\cdot, x)$, is the Riesz representer of the evaluation functional δ_x, that is hence a continuous functional in H. Also the converse holds.

Theorem 1.3. *If for all* $x \in \Omega$ *the evaluation functionals* $\{\delta_x\}$ *are continuous in a Hilbert space* $(H, (\cdot, \cdot)_H)$ *of real valued functions defined on* $\Omega \subset \mathbb{R}^d$, *H is a reproducing kernel Hilbert space.*

We will use in the following the notation $\mathcal{H}(\Omega)$ to denote the native space of K, and $(\cdot, \cdot)$, without subscript, to denote the inner product of $\mathcal{H}(\Omega)$.

One remarkable way to construct the native space is as follows. First one considers the space $\mathcal{H}_0(\Omega) = \operatorname{span}\{K(\cdot, x), x \in \Omega\}$ and then equips it with the positive definite and symmetric bilinear form

$$\left(\sum_j c_j K(\cdot, x_j), \sum_i d_i K(\cdot, x_i) \right) := \sum_{j,i} c_j d_i K(x_j, x_i).$$

The native space $\mathcal{H}(\Omega)$ then is the closure of $\mathcal{H}_0(\Omega)$ with respect to the inner product defined by this form.

1.2 Kernel-based interpolation and approximation

With these tools in hand, we can face the problem of recovering a function $f \in \mathcal{H}(\Omega)$ from its samples. Namely, given a set $X_N = \{x_1, \ldots, x_N\} \subset \Omega$ of distinct points, our goal is to find a function $s_{f,X_N} \in \mathcal{H}(\Omega)$ such that

$$s_{f,X_N}(x_i) = f(x_i), \ 1 \le i \le N. \tag{1.2}$$

The way this problem is solved in $\mathcal{H}(\Omega)$ is by considering an interpolant in the form

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

generated by the linear combination of the kernel translates on the points X_N with unknown coefficients, that need to be computed to meet the interpolation conditions (1.2).

From a linear algebra point of view, finding the coefficient vector $c \in \mathbb{R}^N$ is equivalent to solving a linear equation involving the kernel matrix $A(X_N)$, which we will denote simply as A from now on when no confusion is possible. The equation to solve is

$$Ac = b,$$

where $b_i = f(x_i)$, and it has a unique solution thanks to the positive definiteness of the kernel. This means that, using kernel-based techniques, we can always compute an interpolant function s_{f,X_N} to any set of N values defined on any set of N pairwise distinct points in Ω.

From a functional point of view, instead, we are assigning an N - dimensional *trial space* $V(X_N) \subset \mathcal{H}(\Omega)$ spanned by the kernel translates $K(\cdot, x_1)$, ..., $K(\cdot, x_N)$. We can then view the above interpolation process as the action of an interpolation operator that acts from $\mathcal{H}(\Omega)$ to the finite subspace $V(X_N)$. We have the following.

Proposition 1.4. *The interpolation operator that maps a function $f \in \mathcal{H}(\Omega)$ to its kernel interpolant $s_{f,X_N} \in V(X_N)$ is the orthogonal projector*

$$\Pi_{\mathcal{H},V(X_N)} : \mathcal{H}(\Omega) \to V(X_N).$$

In particular, the kernel interpolant s_{f,X_N} is also the minimal $\mathcal{H}(\Omega)$-norm approximant of f from $V(X_N)$.

Apart from being a functional counterpart of the result on the existence and the uniqueness of the solution of the above linear system, this result tells us that the interpolation process is also a optimal recovery process in $\mathcal{H}(\Omega)$. Moreover, it will be the key to generalize the approximation process to trial spaces in $\mathcal{H}(\Omega)$ more general that the points based subspaces $V(X_N)$.

1.3   Error bounds

The usual way to study the interpolation error is to consider the *Power Function* $P_N(x)$ of X_N at x, that is the $\mathcal{H}(\Omega)$-norm of the pointwise error functional

$$f \mapsto f(x) - s_{f,X_N}(x)$$

at x. More precisely, we have the following definition.

Definition 1.5. *The Power Function $P_N(x)$ of X_N is defined as*

$$P_N(x) = \sup_{\|f\| \leq 1} |f(x) - s_{f,X_N}(x)| \text{ for all } x \in \Omega.$$

It is immediate to conclude that the point-wise interpolation error for a function $f \in \mathcal{H}(\Omega)$ can be bounded by the Power Function as

$$|f(x) - s_{f,X_N}(x)| \leq P_N(x)\|f\| \; \forall x \in \Omega.$$

Many efforts have been devoted to the study of the behavior of P_N on Ω, in order to obtain precise estimates on the decay of the interpolation error in terms of K, Ω and X_N. Before recalling some estimates on the decay of P_N that will be used in the following, we show in the next Lemma that the Power Function can be actually computed using linear combinations of functions of $V(X_N)$.

Lemma 1.6. *For any set of points $X_N \subset \Omega$, there exists a unique cardinal basis $\{l_j\}_{j=1}^N$ of the point based subspace $V(X_N) \subset \mathcal{H}(\Omega)$, i.e., a set of functions such that*

$$l_j(x_i) = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}.$$

Moreover, the Power Function of X_N at $x \in \Omega$ can be computed by the formula

$$P_N(x)^2 = K(x,x) - \sum_{j=1}^N l_j(x)K(x,x_j).$$

The estimate on the Power Function we recall here relates the interpolation error with the smoothness of the kernel and the distribution of the points X_N. They apply to translational invariant kernels $K(x,y) = \Phi(x-y)$, where Φ has a generalized Fourier transform on $\mathbb{R}^d$, which will be denoted as $\hat{\Phi}$. We assume also that $\Omega \subset \mathbb{R}^d$ is bounded and satisfies an interior cone condition.

In this situation we have the following error bounds (see [53, 55]), that relate the approximation error with the *fill distance*

$$h_N = \sup_{x \in \Omega} \min_{x_j \in X_N} \|x - x_j\|.$$

Theorem 1.7. *If the kernel K has finite smoothness, we have that*

$$\hat{\Phi}(\omega) \sim (1 + \|\omega\|)^{-\beta-d}$$

for $\|\omega\| \to \infty$, and

$$\|f - s_{f,X_N}\|_\infty \leq c h_N^{\beta/2}\|f\|, \;\; \text{for all } f \in \mathcal{H}(\Omega), \qquad (1.3)$$

while for infinitely smooth kernels we have

$$\|f - s_{f,X_N}\|_\infty \leq c \exp(-c/h_N)\|f\|, \;\; \text{for all } f \in \mathcal{H}(\Omega). \qquad (1.4)$$

Both bounds are in fact bounds on the $L_\infty(\Omega)$ norm of the Power Function, properly multiplied by the $\mathcal{H}(\Omega)$-norm of f.

A second part of this Theorem, involving estimates in the $L_2(\Omega)$ norm of the interpolation error, is postponed to Theorem 1.7.

The results discussed so far state that functions of $\mathcal{H}(\Omega)$ can be recovered by kernel-based interpolation. Indeed, a partial converse holds, i.e., if a function can be recovered by kernel-based interpolation with a sufficiently fast decaying $L_\infty(\Omega)$ error, then it must belong to $\mathcal{H}(\Omega)$ (see [55]).

We conclude this Section by pointing out that, although theoretically convergent, and even if it has been proven (see [18]) that the approximation process is stable in $\mathcal{H}(\Omega)$, a straightforward application of the *direct* interpolation method described in this Chapter is severely affected by stability and conditioning issues, in particular in the *flat limit* $\varepsilon \to 0$. Since this problem is not the main topic of this thesis, we refer to the aforementioned monographs for a detailed analysis of the stability of the interpolation process, and to the recent papers [6, 12, 20, 24, 26–28, 37] for effective methods to treat the situation $\varepsilon \to 0$.

1.4   Embedding in $L_2(\Omega)$ and the eigenbasis

We conclude this introductory Chapter by analyzing the connection between the native space $\mathcal{H}(\Omega)$ and the space $L_2(\Omega)$.

We make the additional assumptions that Ω is a compact set in $\mathbb{R}^d$ and the kernel is continuous on $\Omega \times \Omega$. This ensures that the native space has a continuous embedding into $L_2(\Omega)$. Indeed, using the reproducing property (ii) we have

$$\|f\|_{L_2} \leq \left(\int_\Omega K(x,x)dx \right)^{1/2} \|f\| \quad \text{for all } f \in \mathcal{H}(\Omega)$$

where the integral of the kernel is finite.

This allows to define a compact and self-adjoint integral operator $T : L_2(\Omega) \to L_2(\Omega)$ that will play a central role in the following.

Definition 1.8. *For $f \in L_2(\Omega)$ we define*

$$Tf(x) = \int_\Omega K(x,y)f(y)dy, \quad x \in \Omega. \tag{1.5}$$

It can be shown that the range $T(L_2(\Omega))$ of T is dense in $\mathcal{H}(\Omega)$, and that the operator is the adjoint of the embedding operator of $\mathcal{H}(\Omega)$ into $L_2(\Omega)$, i.e.,

$$(f,g)_{L_2} = (f,Tg) \text{ for all } f \in \mathcal{H}(\Omega), \ g \in L_2(\Omega). \tag{1.6}$$

This operator is the key tool in the next Theorem (see e.g. [47, Ch. 5]), which applies to our situation, and provides a way to represent the kernel as an *expansion* (or *Hilbert - Schmidt* or *Mercer*) kernel.

Theorem 1.9 (Mercer). *If K is a continuous and positive definite kernel on a compact set Ω, the operator T has a countable set of positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots > 0$ and eigenfunctions $\{\varphi_j\}_{j \in \mathbb{N}}$ with $T\varphi_j = \lambda_j \varphi_j$. The eigenfunctions are orthonormal in $L_2(\Omega)$ and orthogonal in $\mathcal{H}(\Omega)$ with $\|\varphi_j\|^2 = \lambda_j^{-1}$. Moreover, the kernel can be decomposed as*

$$K(x,y) = \sum_{j=1}^{\infty} \lambda_j \, \varphi_j(x) \, \varphi_j(y) \quad x, y \in \Omega, \tag{1.7}$$

where the sum is absolutely and uniformly convergent.

From now on we will call $\{\sqrt{\lambda_j}\varphi_j\}_{j \in \mathbb{N}}$ the *eigenbasis*, and use the notation $E_n = \mathrm{span}\{\sqrt{\lambda_j}\varphi_j, j = 1, \ldots, n\}$.

Furthermore, using this decomposition the native space can be characterized as

$$\mathcal{H}(\Omega) = \left\{ f \in L_2(\Omega) : \sum_j \frac{(f, \varphi_j)_{L_2}}{\lambda_j} < \infty \right\}.$$

We recall that it is also possible to go the other way round and define a positive definite and continuous kernel starting from a given sequence of functions $\{\varphi_j\}_j$ and weights $\{\lambda_j\}_j$, provided some mild conditions of summability and linear independence. Further details on this construction are discussed in Appendix A.

We remark that the operator T can be defined using any positive and finite measure μ with full support on Ω (see [57]) and the same properties still hold, but we will concentrate here on the Lebesgue measure.

Moreover, apart from the results of this thesis, note that eigenfunction expansions play a central role in most of the papers cited in the last Section, and especially in the RBF-QR methods.

Finally, since we can deal with $L_2(\Omega)$-norms of functions in $\mathcal{H}(\Omega)$, we can also state the last part of Theorem 1.7, which can also be found in [53].

Theorem 1.10. *Under the hypothesis of Theorem 1.7 for kernels with finite smoothness the estimate* (1.3) *can be improved as follows*

$$\|f - s_{f,X_N}\|_{L_2} \leq c h_N^{(\beta+d)/2}\|f\|, \text{ for all } f \in \mathcal{H}(\Omega). \tag{1.8}$$

1.5 Generalized interpolation

This Section is devoted to the problem of generalized interpolation, where a function $f \in \mathcal{H}(\Omega)$ is recovered from weak-data. Namely, given a finite set of linear and continuous functionals $L_N = \{\ell_i\}_{i=1}^N \subset \mathcal{H}(\Omega)^*$, we want to find a function $s_{f,L_N} \in \mathcal{H}(\Omega)$ such that

$$\ell_i(s_{f,L_N}) = \ell_i(f), \quad 1 \leq i \leq N.$$

The function s_{f,L_N}, as in the case of point-wise interpolation, denotes the generalized interpolant defined by the above weak interpolation conditions.

This problem, in particular, arises from two different collocation methods for the solution of operator equations, e.g.,

$$L(f) = g, \ \ g \in \mathcal{H}(\Omega), \ \ L : \mathcal{H}(\Omega) \to \mathcal{H}(\Omega).$$

Once a set of collocation points $X_N \subset \Omega$ is chosen (in some applications, a distinction between points in $\mathrm{int}(\Omega)$ and in $\partial\Omega$ is needed), one collects data

$$L(f)(x_i) = g(x_i), \ \ 1 \le i \le N,$$

or, in other words, one constructs the set of linear functionals $L_N = \{\ell_i\}_{i=1}^N$, $\ell_i = \delta_{x_i} \circ L$. Under some assumptions on L, X_N and Ω, we have in fact $L_N \subset \mathcal{H}(\Omega)^*$, and we are in the situation described above.

The choice of the functional form of s_{f,L_N} characterize the method. Namely, one has

$$s_{f,L_N} = \sum_{j=1}^N c_j K(\cdot, x_j)$$

for unsymmetric collocation method, and

$$s_{f,L_N} = \sum_{j=1}^N c_j \ell_j^y (K(\cdot, y))$$

for symmetric collocation methods, where the superscript y, respectively x, means that ℓ_j acts on the second variable of the kernel, respectively on the first.

The interpolation conditions then read

$$\sum_{j=1}^N c_j \ell_i^x (K(x, x_j)) = \ell_i(f) = g(x_i)$$

in the not symmetric case, and

$$\sum_{j=1}^N c_j \ell_i^x \ell_j^y (K(x, y)) = \ell_i(f) = g(x_i)$$

in the symmetric one. While in the second case conditions of linear independence of $L_N \subset \mathcal{H}(\Omega)^*$ ensure the existence of a unique solution s_{f,L_N} (see [21]), the not symmetric case needs more care in the choice of the collocation points X_N (see [33]).

A thorough discussion of the theory and numerics of these methods can be found, e.g., in [60, Chapter 16] and [22, Chapters 36-41]. We recall

only that, even if the unsymmetric formulation, when well-posed, usually produces better approximant (error-wise) and it is faster to implement, the symmetric formulation is optimal in the sense of the $\mathcal{H}(\Omega)$ norm. In particular, the optimality of symmetric collocation is related to the fact that, for $\ell \in \mathcal{H}(\Omega)^*$, its Riesz representer is

$$\ell^y(K(\cdot, y)). \tag{1.9}$$

These are exactly the basis elements used in symmetric collocation.

Chapter 2

Optimal subspaces and approximation of eigenfunctions

The present Chapter is devoted to develop the main theoretical results of this thesis. It focuses on the problem of finding and characterizing error-optimal n-dimensional subspaces of $\mathcal{H}(\Omega)$, and to deduce from them some informations on the approximation of the eigenbasis.

To this end, at first general finite dimensional subspaces of $\mathcal{H}(\Omega)$ are analyzed, together with their associated approximation operators and generalized Power Functions.

Then, the eigenspaces are compared with other n - dimensional subspaces, and some results are proven about their optimality with respect to the minimization of the $L_2(\Omega)$ norm of the generalized Power Function. These results have strong connections to already known results on n-widths (see e.g. [46]), which will be also briefly recalled.

Moreover, the new results are suitable to scale the theory to the case of finite dimensional subspaces of $\mathcal{H}(\Omega)$, and they allow to take into account the usual situation of point-based kernel approximation as presented in the previous Chapter. In this setting a constrained optimization problem is solved to characterize a new kind of optimal subspaces, and we will in particular provide error bounds for the approximation of the true eigensystem in terms of point-based techniques.

Some extensions of these results are presented in Section 2.5. They aim at replacing $L_2(\Omega)$ with a kernel-based space, and to have a first look at the optimal subspace problem in the case of the recovery of linear functionals.

The contents of this Chapter are published in [52] and are partially taken from this paper, except for Section 2.5, which presents unpublished results.

2.1 General subspaces of the native space

At first, we generalize here part of the construction of the previous Chapter. We are interested in considering fully general subspaces V_n of $\mathcal{H}(\Omega)$ of dimension n, generated by any set of n basis functions, and the corresponding kernel-based approximation scheme, i.e., the approximation that uses V_n as a trial space. The following discussion includes as a particular case the usual situation of point-based subspaces and the usual kernel-based interpolation, as presented in Chapter 1.

2.1.1 Kernel-based approximation

Given a subspace V_n, we can define a generalized kernel-based approximation based on it. Namely, since the interpolation operator is a projector, for $f \in \mathcal{H}(\Omega)$ we can consider in this case the approximating function

$$\Pi_{\mathcal{H},V_n} f \quad = \quad \sum_{j=1}^{n} (f, v_j) v_j,$$

where the v_j are $\mathcal{H}(\Omega)$- orthonormal basis functions of V_n, i.e., the $\mathcal{H}(\Omega)$-projection of $f \in \mathcal{H}(\Omega)$ into V_n. Observe that, given a function $f \in \mathcal{H}(\Omega)$, the approximant $\Pi_{\mathcal{H},V_n} f$ can be also defined as the unique function in V_n that satisfies the *weak interpolation* conditions

$$(\Pi_{\mathcal{H},V_n} f, v_j) = (f, v_j), \ 1 \le j \le n,$$

and these conditions reduce to the usual interpolation conditions when one considers the usual point-based sets.

    We are now interested in an error analysis of the approximation by functions from these general subspaces. To this end, we generalize in the following the definition of the Power Function (see 1.5) to the case of a general linear operator Π in a general Hilbert space H.

Definition 2.1. *For a normed linear space H of functions on Ω and a linear operator Π on H such that all the functionals $\delta_x - \delta_x \circ \Pi$ are continuous for some norm $\| \cdot \|_H$, the generalized Power Function in $x \in \Omega$ is the norm of the error functional at x, i.e.,*

$$P_{\Pi,\|.\|_H}(x) := \sup_{\|f\|_H \le 1} |f(x) - (\Pi f)(x)|. \tag{2.1}$$

    The definition fits our situation, because we are free to take $\Pi = \Pi_{\mathcal{H},V_n}$ with $\|.\|_H = \|.\|$, the normed linear space H being $\mathcal{H}(\Omega)$. In the following, when no confusion is possible, we will use the simplified notation $P_{V_n,\mathcal{H}}$ or just P_{V_n} to denote the Power Function of $\Pi_{\mathcal{H},V_n}$ with respect to $\| \cdot \|$.

    We need now to look at the relation between the generalized Power Function, a subspace and its bases. We start with the following Lemma, which allows to safely deal with infinite sums of (squares of) basis functions.

Lemma 2.2. *If a separable Hilbert space H of functions on Ω has continuous point evaluation, then each H-orthonormal basis $\{v_j\}_j$ satisfies*

$$\sum_j v_j^2(x) < \infty.$$

Conversely, the above condition ensures that all point evaluation functionals are continuous.

Proof. The formula

$$f = \sum_j (f, v_j)_H v_j$$

holds in the sense of limits in H. If point evaluations are continuous, we can write

$$f(x) = \sum_j (f, v_j)_H v_j(x) \text{ for all } x \in \Omega$$

in the sense of limits in $\mathbb{R}$. Since the sequence $\{(f, v_j)_H\}_j$ is in ℓ_2 and can be an arbitrary element of that space, the sequence $\{v_j(x)\}_j$ must be in ℓ_2, because the above expression is a continuous linear functional on ℓ_2.

For the converse, observe that for any $n \in \mathbb{N}$, $x \in \Omega$ and $\sum_{j=1}^n c_j^2 \leq 1$ the term $\sum_{j=1}^n c_j v_j(x)$ is bounded above by $\sum_j v_j^2(x)$, which is finite for any $x \in \Omega$. Hence, for any $x \in \Omega$, $\sup_{\|f\|_H \leq 1} |f(x)|$ is uniformly bounded for $f \in H$. $\qquad\square$

Given this Lemma, it is possible to compute the generalized Power Function in terms of an orthonormal basis of the subspace.

Proposition 2.3. *For projectors Π_{V_n} within separable Hilbert spaces H of functions on some domain Ω onto finite-dimensional subspaces V_n generated by H-orthonormal functions $v_1, \ldots, v_n$ that are completed, we have*

$$P^2_{\Pi_{V_n}, \|.\|_H}(x) = \sum_{k>n} v_k^2(x)$$

provided that all point evaluation functionals are continuous.

Proof. The pointwise error at x is

$$f(x) - \Pi_{V_n} f(x) = \sum_{k>n} (f, v_k)_H v_k(x),$$

and, thanks to the previous Lemma, we can safely bound its norm as

$$|f(x) - \Pi_{V_n} f(x)|^2 \leq \sum_{k>n} (f, v_k)_H^2 \sum_{j>n} v_j^2(x)$$

$$= \|f - \Pi_{V_n} f\|_H^2 \sum_{j>n} v_j^2(x) \leq \|f\|_H^2 \sum_{j>n} v_j^2(x),$$

with equality if $f \in V_n^\perp$. $\qquad\square$

This framework includes also the usual Power Function defined in Definition 1.5.

Corollary 2.4. *Let X_n be a set of n points in a compact domain Ω, and let $V(X_n)$ be spanned by the X_n-translates of K. Then the above notion of $P_{V(X_n)}$ coincides with the standard notion of the interpolatory Power Function w.r.t. X_n.*

Proof. The interpolation operator coincides with the projector $\Pi_{\mathcal{H}, V(X_n)}$, and the Power Function is in both cases defined as the norm of the pointwise error functional. $\square$

2.1.2   Other approximation processes

With these tools in hand, we can now study the kernel-based approximation based on a general subspace $V_n \subset \mathcal{H}(\Omega)$, and the previous construction generalizes in a natural way the usual kernel-based interpolation. Nevertheless, it is also possible to define different linear approximation processes based on a subspace V_n. In particular, in the following we will need to consider also the $L_2(\Omega)$ projector

$$\Pi_{L_2, V_n} f \;\; = \;\; \textstyle\sum_{j=1}^{n} (f, w_j)_{L_2} w_j, \quad f \in \mathcal{H}(\Omega),$$

the w_j being a $L_2(\Omega)$-orthonormal basis of V_n, which is well defined thanks to the embedding described in Section 1.4, and it is defined also on all of $L_2(\Omega)$.

We conclude this Section by establishing a fundamental connection between the two projectors defined so far. They do not coincide in general, but there is a special case. For the sake of clarity we present here the proof of the following Lemma, even if it relies on a result that is proven in Section 2.3.1.

Lemma 2.5. *If the projectors coincide on $\mathcal{H}(\Omega)$ for an n-dimensional space V_n, then V_n is spanned by n eigenfunctions. The converse also holds.*

Proof. We start with the converse. For each fixed φ_j, thanks to (1.6), we have

$$(f, \varphi_j)_{L_2} = \lambda_j (f, \varphi_j) \text{ for all } f \in \mathcal{H}(\Omega).$$

Then

$$\Pi_{L_2, V_n} f = \sum_{j=1}^{n} (f, \varphi_j)_{L_2} \varphi_j = \sum_{j=1}^{n} \lambda_j (f, \varphi_j) \varphi_j$$

$$= \sum_{j=1}^{n} (f, \sqrt{\lambda_j} \varphi_j) \sqrt{\lambda_j} \varphi_j = \Pi_{\mathcal{H}, V_n} f.$$

Assume now that the projectors coincide on $\mathcal{H}(\Omega)$. We can choose a basis $\{v_j\}_j$ of V_n which is $L_2(\Omega)$-orthonormal and $\mathcal{H}(\Omega)$-orthogonal (see Lemma

2.9), with $\|v_j\|^2 = 1/\sigma_j$. Since $\Pi_{L_2,V_n} f = \Pi_{\mathcal{H},V_n} f$ for any $f \in \mathcal{H}(\Omega)$, necessarily
$$(f, v_j)_{L_2} = \sigma_j(f, v_j) \text{ for all } f \in \mathcal{H}(\Omega) \text{ and } j = 1, \ldots, n,$$
and in particular for $f = K(\cdot, x)$, $x \in \Omega$. Consequently, $\{v_j\}_j$ and $\{\sigma_j\}_j$ are eigenfunctions / eigenvalues of T. $\qquad\square$

We remark that the approximation operators $\Pi_{\mathcal{H},V_n}$ and Π_{L_2,V_n} are not the only ones that can be defined on V_n, but they are sufficient to develop the results of this Chapter. Despite that, in Section 2.5.1 we will extend and generalize part of the following discussion by taking into account also other projectors.

2.2 Optimal subspaces with respect to the Power Function

In this Section we present the fundamental result of this thesis. Our goal is to analyze the behavior of the approximation error, depending on the n-dimensional subspace V_n. Roughly speaking, we want to characterize the n-dimensional subspaces of minimal error, where the error is measured in different ways.

We will at first review some known results. These results are part of the theory of *widths* (see e.g. the comprehensive monograph [46], and in particular Chapter 4 for the theory in Hilbert spaces)), and they deal with the $L_2(\Omega)$ error associated to the two approximation schemes defined in the previous Section.

We will then prove a Theorem concerning the $L_2(\Omega)$ norm of the pointwise error. Besides providing a new point of view on optimal subspaces, this result allows to scale the theory to the case of closed subspaces dealt with in Section 2.3, and this scaling will be central in this thesis.

2.2.1 Review of the known results

We will concentrate first on the n-width of Kolmogorov. The Kolmogorov n-width $d_n(A; H)$ of a subset A in an Hilbert space H is defined as

$$d_n(A; H) := \inf_{\substack{V_n \subset H \\ \dim(V_n)=n}} \sup_{f \in A} \inf_{v_n \in V_n} \|f - v_n\|_H.$$

It measures how n-dimensional subspaces of H can approximate a given subset A. If the infimum is attained by a subspace, this is called an *optimal subspace*. The interest is in characterizing optimal subspaces and to compute or estimate the asymptotic behavior of the width, usually letting A to be the unit ball $S(H)$ of H.

The first result that introduces and studies n-widths for native spaces was presented in [54]. The authors consider the n-width $d_n(S(\mathcal{H}(\Omega)); L_2(\Omega))$, simply d_n from now on, and they prove the following Theorem. The proof is presented here since analogous proofs will be repeated in this thesis. We recall that E_n, as defined in Section 1.4, is the n-dimensional subspace of $\mathcal{H}(\Omega)$ spanned by the first n eigenfunctions.

Theorem 2.6. *For any $n > 0$ we have*

$$d_n = \inf_{\substack{V_n \subset L_2 \\ \dim(V_n) = n}} \sup_{f \in S(\mathcal{H})} \|f - \Pi_{L_2, V_n} f\|_{L_2} = \sqrt{\lambda_{n+1}}, \tag{2.2}$$

and E_n is an optimal subspace.

Proof. Let $f \in \mathcal{H}(\Omega)$ be a function with $\|f\| = 1$, i.e., since $\{\sqrt{\lambda_j}\varphi_j\}_j$ is an $\mathcal{H}(\Omega)$-orthonormal basis of $\mathcal{H}(\Omega)$,

$$f = \sum_j c_j \sqrt{\lambda_j}\varphi_j, \quad \sum_j c_j^2 = 1.$$

In the case of $V_n = E_n$, since $\{\varphi_j\}_j$ are normalized in $L_2(\Omega)$ we have

$$\|f - \Pi_{L_2, E_n} f\|_{L_2}^2 = \left\| \sum_{j>n} c_j \sqrt{\lambda_j}\varphi_j \right\|_{L_2}^2 = \sum_{j>n} c_j^2 \lambda_j \le \lambda_{n+1} \sum_{j>n} c_j^2,$$

thus $\sup_{f \in S(\mathcal{H})} \|f - \Pi_{L_2, E_n} f\|_{L_2} = \sqrt{\lambda_{n+1}}$ and $d_n \le \sqrt{\lambda_{n+1}}$.

To prove the converse inequality, consider any n-dimensional subspace $V_n \subset L_2(\Omega)$, and observe that the supremum $\sup_{f \in S(\mathcal{H})} \|f - \Pi_{L_2, V_n} f\|_{L_2}$ is obtained for $f \in V_n^\perp$, where the orthogonal complement is taken in $L_2(\Omega)$. Since $\text{codim}(V_n^\perp) = n$, $V_n^\perp$ has nonempty intersection with E_{n+1}. Letting then $g \in V_n^\perp \cap E_{n+1}$, $\|g\| = 1$, we have in particular $g \in E_{n+1}$ and

$$\|g\|_{L_2}^2 = \left\| \sum_{j=1}^{n+1} c_j \sqrt{\lambda_j}\varphi_j \right\|_{L_2}^2 = \sum_{j=1}^{n+1} c_j^2 \lambda_j \ge \lambda_{n+1} \sum_{j=1}^{n+1} c_j^2 = \lambda_{n+1}.$$

Putting all together,

$$\sup_{f \in S(\mathcal{H})} \|f - \Pi_{L_2, V_n} f\|_{L_2} = \sup_{f \in S(V_n^\perp)} \|f\|_{L_2} \ge \|g\|_{L_2} \ge \sqrt{\lambda_{n+1}},$$

thus $d_n \ge \sqrt{\lambda_{n+1}}$. $\square$

This result is a direct application of the general theory of width in Hilbert spaces (see e.g. [46, Corollary 2.6]). Nevertheless, it is the first that exactly

highlights the importance of analyzing the expansion of the operator T to better understand the process of approximation in $\mathcal{H}(\Omega)$.

A further step in the study of optimal subspaces in native spaces have been made in [42]. Namely, it is possible to replace the $L_2(\Omega)$ projector Π_{L_2,V_n} by the $\mathcal{H}(\Omega)$ projector $\Pi_{\mathcal{H},V_n}$, while still keeping the $L_2(\Omega)$ norm to measure the error. The $\mathcal{H}(\Omega)$ projector is the standard interpolation projector, and it differs from the $L_2(\Omega)$ projector unless the space V_n is an eigenspace, as proven in Lemma 2.5.

We consider the $L_2(\Omega)$ norm of the error functional in $\mathcal{H}(\Omega)$ for the projection $\Pi_{\mathcal{H},V_n}$ into a subspace $V_n \subset \mathcal{H}(\Omega)$, i.e.,

$$\sup_{\|f\|_{\mathcal{H}} \leq 1} \|f - \Pi_{\mathcal{H},V_n} f\|_{L_2}$$

and we look for the subspace which minimizes this quantity. In other words, following the definition of the Kolmogorov n-width, it is possible to define

$$\kappa_n := \inf_{\substack{V_n \subset \mathcal{H} \\ \dim(V_n)=n}} \sup_{f \in S(\mathcal{H})} \|f - \Pi_{\mathcal{H},V_n} f\|_{L_2}.$$

We recall in the next Theorem that κ_n is equivalent to d_n, i.e., the best approximation in $L_2(\Omega)$ of $S(\mathcal{H}(\Omega))$ with respect to $\|\cdot\|_{L_2}$ can be achieved using $\mathcal{H}(\Omega)$ itself and the projector $\Pi_{\mathcal{H},V_n}$.

Theorem 2.7. *For any $n > 0$ we have*

$$\kappa_n = \sqrt{\lambda_{n+1}},$$

and E_n is an optimal subspace.

Proof. Since $\mathcal{H}(\Omega) \subset L_2(\Omega)$ and since $\Pi_{L_2,V_n} f$ is the best approximation from V_n of $f \in \mathcal{H}(\Omega)$ wrt. $\|\cdot\|_{L_2}$, we have

$$d_n = \inf_{\substack{V_n \subset L_2 \\ \dim(V_n)=n}} \sup_{f \in S(\mathcal{H})} \|f - \Pi_{L_2,V_n} f\|_{L_2}$$

$$\leqslant \inf_{\substack{V_n \subset \mathcal{H} \\ \dim(V_n)=n}} \sup_{f \in S(\mathcal{H})} \|f - \Pi_{L_2,V_n} f\|_{L_2}$$

$$\leqslant \inf_{\substack{V_n \subset \mathcal{H} \\ \dim(V_n)=n}} \sup_{f \in S(\mathcal{H})} \|f - \Pi_{\mathcal{H},V_n} f\|_{L_2} = \kappa_n.$$

On the other hand, since $\Pi_{L_2,E_n} = \Pi_{\mathcal{H},E_n}$ on $\mathcal{H}(\Omega)$ (Lemma 2.5),

$$\kappa_n \leqslant \sup_{f \in S(\mathcal{H})} \|f - \Pi_{\mathcal{H},E_n} f\|_{L_2}$$

$$= \sup_{f \in S(\mathcal{H})} \|f - \Pi_{L_2,E_n} f\|_{L_2} = d_n,$$

since E_n is optimal for d_n. Hence $\kappa_n = d_n = \sqrt{\lambda_{n+1}}$. $\qquad\square$

2.2.2 Optimality with respect to the Power Function

Given these results on optimal error subspaces, we move now to another way of studying the approximation error. Instead of directly considering the $L_2(\Omega)$ norm of approximants, we first take the norm of the pointwise error of the $\Pi_{\mathcal{H},E_n}$ projector and then minimize its $L_2(\Omega)$ norm over Ω. This means to find a subspace which minimizes the $L_2(\Omega)$ norm $\|P_{V_n}\|_{L_2}$ of the Power Function P_{V_n} among all n-dimensional subspaces $V_n \subset \mathcal{H}(\Omega)$. Using the definition of the generalized Power Function, we can rephrase the problem in the fashion of the previous results by defining

$$p_n := \inf_{\substack{V_n \subset \mathcal{H} \\ \dim(V_n)=n}} \left\| \sup_{f \in S(\mathcal{H})} |f(\cdot) - \Pi_{\mathcal{H},V_n} f(\cdot)| \right\|_{L_2}.$$

This quantity is of interest for at least two reasons. First, the Power Function is the usual way to measure the approximation error and it can be easily and directly computed also for the standard, point-based interpolation process, thus it is useful to extend the previous results to this situation. Second, as will be clear in the next Section, results on p_n can be employed when dealing with closed or even finite dimensional subspaces of the native space, despite what one can do with d_n and κ_n.

In the following Theorem we compute and characterize the values of p_n. It is relevant that, also in this case, the optimal n-dimensional subset is E_n, and p_n can be expressed in terms of the eigenvalues, in a way that involves all the indexes $j > n$. The proof mimics [34, Theorem 1].

Theorem 2.8. *For any $n > 0$ we have*

$$p_n = \sqrt{\sum_{j>n} \lambda_j},$$

and E_n is an optimal subspace.

Proof. For a subset V_n we can consider an $\mathcal{H}(\Omega)$-orthonormal basis $\{v_k\}_{k=1}^n$ and complete it to an orthonormal basis $\{v_k\}_{k \in \mathbb{N}}$ of $\mathcal{H}(\Omega)$. We can move from the eigenbasis to this basis using a matrix $A = (a_{ij})$ as

$$v_k = \sum_{j=1}^{\infty} a_{jk} \sqrt{\lambda_j} \varphi_j, \tag{2.3}$$

where $\sum_{j=1}^{\infty} a_{jk}^2 = \sum_{k=1}^{\infty} a_{jk}^2 = 1$. Hence, the power function of V_n is

$$P_{V_n}(x)^2 = \sum_{k>n} v_k(x)^2 = \sum_{k>n} \left(\sum_{j=1}^{\infty} a_{jk} \sqrt{\lambda_j} \varphi_j(x) \right)^2$$

$$= \sum_{i,j=1}^{\infty} \sqrt{\lambda_i} \varphi_i(x) \sqrt{\lambda_j} \varphi_j(x) \sum_{k>n} a_{ik} a_{jk},$$

and, defining $q_j = \sum_{k=1}^n a_{jk}^2$, we can compute its norm as

$$\|P_{V_n}\|_{L_2}^2 = \int_\Omega \sum_{k>n} \left(\sum_{j=1}^\infty a_{jk} \sqrt{\lambda_j} \varphi_j(x) \right)^2 dx$$

$$= \sum_{k>n} \sum_{j=1}^\infty a_{jk}^2 \lambda_j = \sum_{j=1}^\infty \lambda_j - \sum_{j=1}^\infty q_j \lambda_j.$$

Now we need to prove that $\sum_{j=1}^\infty q_j \lambda_j \leqslant \sum_{j=1}^n \lambda_j$.

Let $m = \lceil \sum_j q_j \rceil \leq n$. We split the cumulative sum over the q_j into integer ranges

$$i - 1 < \sum_{j=1}^{j_i} q_j \leqslant i, \ 1 \leqslant i \leq m.$$

Then j_m can be infinite, but j_{m-1} is finite, and since $0 \leq q_j \leq 1$ we get stepwise

$$\begin{aligned} 0 < \textstyle\sum_{j=j_{i-1}+1}^{j_i} q_j &\leqslant 1, \\ j_i - j_{i-1} &\geq 1, \\ j_i &\geq i, \\ j_i \geq j_{i-1} + 1 &\geq i \end{aligned}$$

for $1 \leq i \leq m$, using $j_0 = 0$. Since the sequence of the eigenvalues is non negative and non increasing, this implies

$$\begin{aligned} \sum_{j=1}^\infty q_j \lambda_j &\leq \sum_{i=1}^{m-1} q_{j_{i-1}+1} \lambda_{j_{i-1}+1} + \lambda_{j_{m-1}+1} \sum_{j=j_{m-1}+1}^{j_m} q_j \\ &\leq \sum_{i=1}^m \lambda_{j_{i-1}+1} \leq \sum_{i=1}^m \lambda_i \leq \sum_{i=1}^n \lambda_i. \end{aligned}$$

If we take $V_n = E_n$ and $\{\sqrt{\lambda_j} \varphi_j\}_{j=1}^n$ as its basis, the matrix A in (2.3) is the infinite identity matrix. Thus equality holds in the last inequality. $\square$

2.3 Computable optimal subspaces

The previous results motivate the interest for the knowledge and study of the eigenbasis. Nevertheless, there are some details that need to be taken into account when dealing with actual approximation schemes.

First, the eigenbasis is not known in general, and, even if known, it can not be used for truly scattered data approximation. In fact, by the Mairhuber-Curtis Theorem [13, 38] (see e.g. [60, Theorem 2.3]), there exists at least a set of points $X_n \subset \Omega$ such that the collocation matrix of E_n on X_n is singular. Thus, we need to use a trial space V_n that is close, in some sense, to E_n, but that needs to be point-generated.

Second, in many situations we are given a prescribed set of data collected at fixed points $X_N \subset \Omega$, where N can be large. In this case we are interested in techniques that allow to extract from the data a suitable, low dimensional subspace to be used as a trial space or, in the language of the previous Section, to select the width-minimizing n-dimensional subspace in $V(X_N)$.

To deal with both these problems, we start by considering subspaces of $\mathcal{H}(\Omega)$ of the form $V(X_N) = \text{span}\{K(\cdot, x) : x \in X_N\}$, where $X_N = \{x_1, \ldots, x_N\}$ is a possibly large but finite set of points in Ω. The basic idea is to replace $\mathcal{H}(\Omega)$ by $V(X_N)$ in order to get, in the first case, a good point-dependent approximation to the true eigenbasis with respect to $\mathcal{H}(\Omega)$ and, in the second case, to take into account the given data.

We will then repeat the constructions of the previous Section for a finite-dimensional native space, i.e., the problem of finding, for $n < N$, an n-dimensional subset V_n which minimizes the error, in some norm, among all the subspaces of $V(X_N)$ of dimension n. Observe that, despite what happens for the true eigenbasis, any subspace in $V(X_N)$ can now be exactly constructed through the usual point-based techniques and their linear algebra counterparts, as will be done in Chapter 3.

We remark that one could expect that the optimal subspace in $V(X_N)$ is the projection of E_n into $V(X_N)$. In fact, as we will see, this is not the case, but the optimal subspace will still approximate the true eigenspaces in a near-optimal sense.

2.3.1   Restriction to a closed subspace

The analysis of such finite and point-based subspaces can be carried out by looking at general closed subspaces of the native space. Indeed, although these subspaces have no direct interest here, in what follows there is no need to assume the dimension of the subspace to be finite.

It can be proven (see [40, Th. 1]) that, if V is a closed subspace of $\mathcal{H}(\Omega)$, it is the native space on Ω of the kernel $K_V(x, y) = \Pi^x_{\mathcal{H},V} \Pi^y_{\mathcal{H},V} K(x, y)$, with inner product given by the restriction of the one of $\mathcal{H}(\Omega)$.

Being V itself a native space, we can repeat the same construction as before. In particular, we have the embedding $V \hookrightarrow L_2(\Omega)$ and it is possible to define the restricted operator $T_V : L_2(\Omega) \to L_2(\Omega)$ as

$$T_V f(x) = \int_\Omega K_V(x, y) f(y) dy, \quad x \in \Omega, \tag{2.4}$$

which maps $L_2(\Omega)$ into V.

The Mercer Theorem 1.9 applied to this operator gives the eigenbasis for V on Ω and the corresponding eigenvalues, which will be denoted as $\{\varphi_{j,V}\}_j$, $\{\lambda_{j,V}\}_j$. They can be finitely or infinitely many, depending on the

dimension of V. For $n \leq \dim(V)$, we will use the notation

$$E_{n,V} = \mathrm{span}\{\sqrt{\lambda_{j,V}}\varphi_{j,V}, \ j = 1, \ldots, n\}.$$

As an immediate result, this decomposition of T_V proves the following Lemma, which was already used in the proof of Lemma 2.5 in Section 2.1.2. Uniqueness is understood here like stating uniqueness of the eigenvalue expansion of the integral operator defined by the kernel, i.e., the eigenspaces are unique.

Lemma 2.9. *Any closed subspace V of the native space has a unique basis which is $\mathcal{H}(\Omega)$-orthogonal and $L_2(\Omega)$-orthonormal.*

Moreover, since V is a native space itself, the results of the previous Section hold also for V. Namely, we can define in V the analogous notions of d_n, κ_n and p_n, say $d_{n,V}$, $\kappa_{n,V}$, $p_{n,V}$,

$$d_{n,V} = \inf_{\substack{V_n \subset L_2 \\ \dim(V_n)=n}} \sup_{f \in S(V)} \|f - \Pi_{L_2,V_n}f\|_{L_2}$$

$$\kappa_{n,V} = \inf_{\substack{V_n \subset V \\ \dim(V_n)=n}} \sup_{f \in S(V)} \|f - \Pi_{\mathcal{H},V_n}f\|_{L_2}$$

$$p_{n,V} = \inf_{\substack{V_n \subset V \\ \dim(V_n)=n}} \left\| \sup_{f \in S(V)} |f(\cdot) - \Pi_{\mathcal{H},V_n}f(\cdot)| \right\|_{L_2},$$

and we immediately have the following Corollary of Theorems 2.7, 2.8.

Corollary 2.10. *We have*

$$d_{n,V} = \kappa_{n,V} = \sqrt{\lambda_{n+1,V}}, \ \ p_{n,V} = \sqrt{\sum_{j>n} \lambda_{j,V}}$$

where $E_{n,V}$ is an optimal subspace.

The previous results deal with the best approximation of the unit ball $S(V)$, that is not exactly what we are looking for. Instead, we would like to know how to choose a minimal error subspace in V to approximate *any* function in $\mathcal{H}(\Omega)$, not only in V, i.e., we want to minimize the error of approximation of $S(\mathcal{H}(\Omega))$ using only subspaces of V. That is, we are interested in the widths $\tilde{d}_{n,V}$, $\tilde{\kappa}_{n,V}$, $\tilde{p}_{n,V}$ defined as follows

$$\tilde{d}_{n,V} = \inf_{\substack{V_n \subset L_2 \\ \dim(V_n)=n}} \sup_{f \in S(\mathcal{H})} \|f - \Pi_{L_2,V_n}f\|_{L_2}$$

$$\tilde{\kappa}_{n,V} = \inf_{\substack{V_n \subset V \\ \dim(V_n)=n}} \sup_{f \in S(\mathcal{H})} \|f - \Pi_{\mathcal{H},V_n}f\|_{L_2}$$

$$\tilde{p}_{n,V} = \inf_{\substack{V_n \subset V \\ \dim(V_n)=n}} \left\| \sup_{f \in S(\mathcal{H})} |f(\cdot) - \Pi_{\mathcal{H},V_n}f(\cdot)| \right\|_{L_2}.$$

The first notion is not useful, since we obviously have $\tilde{d}_{n,V} = d_n$ due to the fact that the subspace V_n in these definitions is chosen in $L_2(\Omega)$. Instead, we are interested in $\tilde{\kappa}_{n,V}$ and $\tilde{p}_{n,V}$. Here we will compute the second one, while nothing can be said about the first one. The fact that the Power Function-based width scales also to this situation is one of the reasons that motivates the interest in it.

2.3.2   The case of p_n

We start by establishing a connection between the kernels and Power Functions of V and $\mathcal{H}(\Omega)$.

Lemma 2.11. *If $V \subset \mathcal{H}(\Omega)$ is closed,*

$$P_{V,\mathcal{H}}(x)^2 = K(x,x) - K_V(x,x) \quad \text{for all } x \in \Omega. \tag{2.5}$$

Moreover, if $U \subset V \subset \mathcal{H}(\Omega)$ are closed, the Power Functions are related as

$$P_{U,\mathcal{H}}(x)^2 = P_{U,V}(x)^2 + P_{V,\mathcal{H}}(x)^2 \quad \text{for all } x \in \Omega.$$

Proof. The Power Function is the norm of the error functional. For $f \in \mathcal{H}(\Omega)$, $\|f\| \leq 1$, and $x \in \Omega$ we have

$$
\begin{aligned}
|f(x) - \Pi_{\mathcal{H},V} f(x)| &= |(f, K(\cdot,x) - K_V(\cdot,x))| \\
&\leq \|f\| \|K(\cdot,x) - K_V(\cdot,x)\| \leq \sqrt{K(x,x) - K_V(x,x)},
\end{aligned}
$$

with equality if f is the normalized difference of the kernels. This proves (2.5), and the relation between the Power Functions easily follows. $\qquad\square$

Thanks to this Lemma, we can prove the following corollary of Theorem 2.8, that will be the base for the estimates of Section 2.4.

Corollary 2.12. *Let $V \subset \mathcal{H}(\Omega)$ be a closed subspace of $\mathcal{H}(\Omega)$. For any $n \leq \dim(V)$ we have*

$$\tilde{p}_{n,V} = \sqrt{\sum_{j=1}^{\infty} \lambda_j - \sum_{j=1}^{n} \lambda_{j,V}}$$

and $E_{n,V}$ is an optimal subspace. In particular

$$\|P_V\|_{L_2} = \sqrt{\sum_{j=1}^{\infty} \lambda_j - \sum_{j=1}^{\dim V} \lambda_{j,V}}.$$

Proof. Thanks to Lemma 2.11, we know that for any $V_n \subset V$ and for any $x \in \Omega$, the squared power functions of $\mathcal{H}(\Omega)$ and of V differ by an additive constant. This means that the minimality of $E_{n,V}$ stated in Corollary 2.10

does not change if we consider the standard power function on $\mathcal{H}(\Omega)$ instead of the one of V. Moreover, the value of the width can be computed as

$$
\int_\Omega P^2_{E_n,\mathcal{H}(\Omega)}(x)dx = \int_\Omega P^2_{E_n,V}(x)dx + \int_\Omega P^2_{V,\mathcal{H}(\Omega)}(x)dx
$$

$$
= \sum_{j=1}^m \lambda_{j,V} - \sum_{j=1}^n \lambda_{j,V} + \sum_{j=1}^\infty \lambda_j - \sum_{j=1}^m \lambda_{j,V}
$$

$$
= \sum_{j=1}^\infty \lambda_j - \sum_{j=1}^n \lambda_{j,V}.
$$

$\square$

This result has two consequences. At one hand, if we want to have a small Power Function, we need to choose a subspace V which provides a good approximation of the true eigenvalues. On the other hand, when dealing with certain point based subspaces, we can control the decay of the Power Function depending on the number of points we are using, and this bound will provide a bound also on the convergence of the discrete eigenvalues to the true one. The last fact will be discussed in more detail in Section 2.4.

2.3.3 The case of κ_n

We briefly outline here why we are not able to obtain similar results when considering the width κ_n. As we saw, for a closed subspace $V \subset \mathcal{H}(\Omega)$ we have $\kappa_{n,V} = \sqrt{\lambda_{n+1,V}}$, with optimality on the space $E_{n,V}$ spanned by the first n eigenfunctions. If we try to compute $\tilde{\kappa}_{n,V}$, for a given $V_n \subset V$ we have

$$
\sup_{f\in S(\mathcal{H})} \|f - \Pi_{\mathcal{H},V_n}f\|_{L_2} = \sup_{f\in S(V_n^\perp)} \|f\|_{L_2} = \sqrt{\lambda_{1,V_n^\perp}},
$$

where $V_n^\perp$ is the orthogonal complement taken in $\mathcal{H}(\Omega)$, not in V.

On the other hand,

$$
\sup_{f\in S(\mathcal{H})} \|f - \Pi_{\mathcal{H},V_n}f\|_{L_2} \le \sup_{f\in S(\mathcal{H})} \|f - \Pi_{\mathcal{H},V}f\|_{L_2} + \sup_{f\in S(\mathcal{H})} \|\Pi_{\mathcal{H},V}f - \Pi_{\mathcal{H},V_n}f\|_{L_2}
$$

$$
= \sup_{f\in S(\mathcal{H})} \|f - \Pi_{\mathcal{H},V}f\|_{L_2} + \sup_{f\in S(V)} \|f - \Pi_{\mathcal{H},V_n}f\|_{L_2}
$$

and since the first term in the right hand sides is independent of n, we have

$$
\tilde{\kappa}_{n,V} \le \sup_{f\in S(\mathcal{H})} \|f - \Pi_{\mathcal{H},V}f\|_{L_2} + \kappa_{n,V}
$$

$$
= \sup_{f\in S(V^\perp)} \|f\|_{L_2} + \sqrt{\lambda_{n+1,V}} = \sqrt{\lambda_{1,V^\perp}} + \sqrt{\lambda_{n+1,V}}.
$$

So the problem is how, for a given V_n, $\lambda_{1,V_n^\perp}$ splits between V and $V^\perp$ and, in particular, the term depending on $V^\perp$ is not dependent on n. Moreover, there are different situations: if $V \supset E_{n+1}$, the optimal subspace is E_n, $\lambda_{1,E_n^\perp} = \lambda_{n+1}$ and there is no contribution from $V^\perp$. Otherwise we should have a contribution from both V and $V^\perp$. Thus it seems not to exist a general pattern that allows to deduce relations between the true and the discrete eigenvalues.

2.3.4 Approximation of the true eigenspaces

We remark that there is a relation between E_n and $E_{n,V}$: as mentioned before, the optimal subspace $E_{n,V}$ is not the projection of E_n into V, but is near to be its optimal approximation from V. To see this, observe that the operator T_V is the projection of T into V. In fact, given $K_V(x,y) = \sum_j \lambda_{j,V} \varphi_{j,V}(x)\varphi_{j,V}(y)$, we have for any $f \in L_2(\Omega)$

$$
T_V f(x) = \int_\Omega K_V(x,y) f(y) dy = \sum_j \sqrt{\lambda_{j,V}} \varphi_{j,V}(x)(\sqrt{\lambda_{j,V}} \varphi_{j,V}, f)_{L_2}
$$
$$
= \sum_j \sqrt{\lambda_{j,V}} \varphi_{j,V}(x)(\sqrt{\lambda_{j,V}} \varphi_{j,V}, Tf) = \Pi_{\mathcal{H},V} Tf(x).
$$

This means that the couples $(\lambda_{j,V}, \varphi_{j,V})$ are precisely the Bubnov - Galerkin approximations (see e.g. [36, Sec. 18.4]) of the solutions (λ_j, φ_j) of the eigenvalue problem for the restricted operator $T : \mathcal{H}(\Omega) \to \mathcal{H}(\Omega)$ (which is still a compact, positive and self-adjoint operator). We can then use the well known estimates on the convergence of the Bubnov - Galerkin method to express the distance between E_n and $E_{n,V}$.

The following Proposition collects convergence results which follow from [36, Th. 18.5, Th. 18.6].

Proposition 2.13. *Let $V_1 \subset V_2 \subset \cdots \subset V_n \subset \ldots$ be a sequence of closed subspaces which become dense in $\mathcal{H}(\Omega)$. For $1 \leqslant j \leqslant \dim V_n$ we have*

(i) $\lambda_{j,V_n} \leqslant \lambda_{j,V_{n+1}} \leqslant \lambda_j$,

(ii) *Let $r \in \mathbb{N}$ be the multiplicity of λ_j and*
    $F_{j,n} = \{f \in V_n : T_{V_n} f = \lambda_{i,V_n} f$ *and* $\lim_{n\to\infty} \lambda_{i,V_n} = \lambda_j\}$. *For n sufficiently large, $\dim F_{j,n} = r$ and there exists $c_{j,n} > 1/\lambda_j$, $c_{j,n} \to 1/\lambda_j$ as $n \to \infty$, s.t.*

$$
\|\varphi_j - \Pi_{\mathcal{H},F_{j,n}} \varphi_j\| \leqslant c_{j,n} \lambda_j \|\varphi_j - \Pi_{\mathcal{H},V_n} \varphi_j\|. \tag{2.6}
$$

Inequality (2.6) proves in particular that $E_{n,V}$ is an asymptotically optimal approximation of E_n. Indeed, under the assumptions of the last Proposition, we have

$$
\|\varphi_j - \Pi_{\mathcal{H},V_n} \varphi_j\| \leq \|\varphi_j - \Pi_{\mathcal{H},F_{j,n}} \varphi_j\| \leqslant c\|\varphi_j - \Pi_{\mathcal{H},V_n} \varphi_j\|, \tag{2.7}
$$

with $c \to 1$ as $n \to \infty$.

To conclude this Section we point out that, in addition to the point based sets, there is another remarkable way to produce closed subspaces of the native space. Namely, if $\Omega_1 \subset \Omega$ is any Lipschitz subdomain of Ω, $\mathcal{H}(\Omega_1)$ is a closed subset of $\mathcal{H}(\Omega)$. Point (i) of Proposition 2.13 implies that the eigenvalues are decreasing with respect to the inclusion of the base domain (as can be proven also by the min/max characterization of the eigenvalues).

2.4   Convergence of the discrete eigenvalues

In Theorem 2.12 we established a relation between the approximation error and the difference between the true and the discrete eigenvalues. This allows us to use the well known bounds on the approximation error to give corresponding bounds on the convergence of the discrete eigenvalues.

This convergence result relies on a technique developed in [54] to deal with a very similar problem, namely the decay rate of the true eigenvalues, which we recall here for completeness. The idea is as follows. Under the assumptions of Theorems 1.7 and 1.10 on the kernel and on the domain, we can consider a set of asymptotically uniformly distributed points X_n in Ω, i.e., points such that the fill distance behaves like $h_n \leq cn^{-1/d}$. We then have the bounds (1.8) and (1.4) on the $L_2(\Omega)$-norm of the approximation error given by the kernel interpolation on X_n and we can then control, with the same bound, the $L_2(\Omega)$-norm of the approximation error provided by the best n-dimensional subspace, that is, we have a bound on the width κ_n. Observe that inequality (1.4) is in fact a bound on the $L_\infty(\Omega)$ norm, but the bound on the $L_2(\Omega)$ norm easily follows by the compactness of Ω. This reasoning leads to the following Theorem, that is a summary of [54, Th.13, Th.15].

Theorem 2.14. *Under the above assumptions of Theorem 1.7 on K and Ω, the eigenvalues decay at least like*

$$\sqrt{\lambda_{n+1}} < c_1 n^{-(\beta+d)/2d}$$

for a kernel with smoothness β, and at least like

$$\sqrt{\lambda_{n+1}} < c_2 \exp(-c_3 n^{1/d}),$$

for kernels with unlimited smoothness. The constants c_1, c_2, c_3 are independent of n, but dependent on K, Ω, and the space dimension.

It is important to notice that the asymptotics of the eigenvalues is known for the kernels of limited smoothness, and on $\mathbb{R}^d$. If the kernel is of order β, its native space on $\mathbb{R}^d$ is norm equivalent to the Sobolev space $H^{(\beta+d)/2}$. In these spaces the n-width, and hence the eigenvalues, decay like $\Theta(n^{-\frac{\beta+d}{2d}})$

(see [35]). This means that in Sobolev spaces one can recover (asymptotically) the best order of approximation using kernel spaces.

We can now apply the same technique to Theorem 2.12. Since this Theorem involves the $L_2(\Omega)$ norm of the Power Function, we need in this case to consider bound (1.3) in the case of a kernel of limited smoothness, the latter, in fact, being a bound on the Power Function. This proves the following new Corollary of Theorem 2.12.

Corollary 2.15. *Under the above assumptions on K and Ω, we have*

$$0 \le \lambda_j - \lambda_{j,V(X_n)} < c_1 n^{-\beta/d}, \ \ 1 \le j \le n,$$

for a kernel with smoothness β, and

$$0 \le \lambda_j - \lambda_{j,V(X_n)} < c_2 \exp(-c_3 n^{1/d}), \ \ 1 \le j \le n,$$

for kernels with unlimited smoothness. The constants c_1, c_2, c_3 are independent of n, but dependent on K, Ω, and the space dimension.

This Corollary and the previous Theorem proves that, using point based sets with properly chosen points, one can achieve at the same time a fast decay of the true eigenvalues and a fast convergence of the discrete ones.

Both results in this section raise some (open) questions about the converse implication. From Theorem 2.14 we know that the smoothness of the kernel guarantees a fast decay of the eigenvalues. But we can also start from a given expansion to construct a kernel. Is it possible to conclude smoothness of the kernel from fast decay of the eigenvalues?

Corollary 2.15 tells us that uniformly distributed point based sets provide a good approximation of the eigenvalues. We will see in Chapters 3 that one can numerically construct point based sets whose eigenvalues are close to the true ones. Is it possible to prove that these sets are necessarily asymptotically uniformly distributed?

2.5   Extensions

We conclude this Chapter by presenting two extensions, in two different directions, of some of the previous results. The first one aims at replacing the $L_2(\Omega)$ norm with a stronger one. The second one shows that some results on widths can be proven also in the case of the recovery of general linear functionals.

2.5.1   Intermediate spaces

In the previous Sections, we studied the dependence on $V_n \subset \mathcal{H}(\Omega)$ of the error

$$f - \Pi_{\mathcal{H}, V_n} f$$

for functions $f \in \mathcal{H}(\Omega)$, and we found and characterized the minimal error subspace V_n. The approximation operator being a projection, we could not use the $\mathcal{H}(\Omega)$ norm to measure the error, so we needed a weaker norm, and we have chosen to use $\| \cdot \|_{L_2}$. But we can also consider an intermediate norm, i.e., the norm of a Hilbert space where $\mathcal{H}(\Omega)$ is embedded in and which is in turn embedded into $L_2(\Omega)$. In this Section we consider the special situation of a family of native spaces that satisfies these embeddings, and we show how the previous results are consequently modified.

In particular, given a sequence of positive and non increasing *weights* $\{\mu_j\}_j \subset \mathbb{R}$, $\lim_{j \to \infty} \mu_j = 0$, we consider a kernel K_μ (and its native space $\mathcal{H}_\mu(\Omega)$ over Ω) whose Mercer decomposition (1.7) has the same eigenfunctions as K, while the eigenvalues are replaced by the sequence $\{\mu_j\}_j$. Appendix A analyses in details the conditions that need to be satisfied by the weights to give rise to a kernel in this way and to have $\mathcal{H}(\Omega) \hookrightarrow \mathcal{H}_\mu(\Omega) \hookrightarrow L_2(\Omega)$, and proves the equivalents of Lemmas 2.9 and 2.5 when $L_2(\Omega)$ is replaced by $\mathcal{H}_\mu(\Omega)$, that is, Lemmas A.1 and A.2. From now on we assume that $\{\mu_j\}_j$ satisfies these conditions. Moreover, we will slightly change the notation used so far. Namely, as in Appendix A, the kernel K and the native space $\mathcal{H}(\Omega)$ will be denoted as K_λ, $\mathcal{H}_\lambda(\Omega)$ respectively, and to distinguish the two norms we will use the subscripts $\| \cdot \|_\lambda$, $\| \cdot \|_\mu$.

With these tools we can look at the modified optimal subspace problems, where we replace $L_2(\Omega)$ with $\mathcal{H}_\mu(\Omega)$. We consider

$$d_n(\mu) = \inf_{\substack{V_n \subset \mathcal{H}_\mu \\ \dim(V_n) = n}} \sup_{f \in S(\mathcal{H}_\lambda)} \|f - \Pi_{\mathcal{H}_\mu, V_n} f\|_\mu$$

in place of d_n, and

$$\kappa_n(\mu) = \inf_{\substack{V_n \subset \mathcal{H}_\lambda \\ \dim(V_n) = n}} \sup_{f \in S(\mathcal{H}_\lambda)} \|f - \Pi_{\mathcal{H}_\lambda, V_n} f\|_\mu$$

in place of κ_n.

We remark that the same problem for p_n is not well defined in general, because it is an open problem to prove if the Power Function of $\mathcal{H}_\lambda(\Omega)$ is or not in $\mathcal{H}_\mu(\Omega)$. Nevertheless, it is for sure in $L_2(\Omega)$, Ω being compact, so we can expect that it is also in some intermediate space.

In this setting, the results of (2.2) and of Theorem 2.7 read as follows.

Theorem 2.16. *For any $n > 0$ we have*

$$d_n(\mu) = \kappa_n(\mu) = \sqrt{\frac{\lambda_{n+1}}{\mu_{n+1}}}$$

and in both cases an optimal subspace is E_n.

Proof. The proof is the same of the equivalent results of Section 2.2.1 (which can be found in the cited papers), where, in this case, one needs to consider the operator $T_{\lambda/\mu}$ defined in A.3 in place of the operator T defined in 1.5.

$\square$

In principle, part of the results of the previous Sections concerning closed subspaces of $\mathcal{H}_\lambda(\Omega)$ can be extended in a similar way. But since, as explained before, we can not extend the results concerning the Power Function, and since the results on p_n are the key for our results on closed subspaces, we do not give further details here.

Finally, we remark that this extension can be useful in developing numerical algorithms. Namely, replacing $L_2(\Omega)$ with a native space means to being able to avoid to compute integrals, and instead compute kernel-based norms. For this approach to be applicable, one needs to consider a fixed set of eigenfunctions and two set of weights (eigenvalues) $\{\lambda_j\}_j$ and $\{\mu_j\}_j$, such that both the kernels K_λ and K_μ are known as a closed form expression. Although this kind of expansion is not known for commonly used kernels, this is the case for example of the *Chebyshev kernels*, that can be constructed using Chebyshev polynomials as eigenfunctions (hence, by using a weighted L_2 space) and with different possible choices for the weights (with either algebraic or geometric decay rate). Further details on these kernels can be found in [23, Section 3.9.2].

2.5.2   Recovery of linear functionals

Here we look at the problem of generalized interpolation, as discussed in Section 1.5.

We look at a generalized interpolant in a subspace $V_n \subset \mathcal{H}(\Omega)$, where V_n can be either in the form $V_n = V(X_n)$ as in unsymmetric collocation, or

$$V_n = \operatorname{span}\{\ell_1^y(K(\cdot, y)), \dots, \ell_n^y(K(\cdot, y)\},$$

as in symmetric collocation, or spanned by fully general functions in $\mathcal{H}(\Omega)$.

For a given V_n, we want to consider the error

$$\sup_{\|f\| \leq 1} |\ell(f) - \ell(\Pi_{\mathcal{H}, V_n})|,$$

for certain $\ell \in \mathcal{H}(\Omega)^*$, and then to consider the worst-approximable functional

$$\sup_{\|\ell\|_{\mathcal{H}^*} \leq 1} \sup_{\|f\| \leq 1} |\ell(f) - \ell(\Pi_{\mathcal{H}, V_n})|.$$

Unfortunately, this normalization on the functionals would lead to orthogonality problems similar to the one already encountered in the previous Sections, that is, the value of this quantity would be 1 for any V_n.

A possible solution is to impose a weaker normalization on the functionals, namely, instead of (see 1.9)

$$\|\ell\|^2 = \ell^x \ell^y K(x,y) = \sum_{j=1}^{\infty} \lambda_j \ell^2(\varphi_j) \leq 1,$$

we impose

$$\|\ell\|_2^2 := \sum_{j=1}^{\infty} \ell^2(\varphi_j) \leq 1,$$

where the norm $\|\cdot\|_2$ is defined to resemble the $L_2(\Omega)$ norm.

We have the following result, that enlightens once again the special role of the Mercer decomposition of the kernel. In what follows, $\|\cdot\|_{\ell_2}$, when computed on a square-summable sequence, denotes the standard ℓ_2 norm.

Theorem 2.17. *We have*

$$\inf_{\substack{V_n \subseteq \mathcal{H} \\ \dim V_n = n}} \sup_{\|\ell\|_2 \leq 1} \sup_{f \in S(\mathcal{H})} |\ell(f) - \ell(\Pi_{\mathcal{H}, V_n} f)| = \sqrt{\lambda_{n+1}}$$

with equality on the eigenspace E_n.

Proof. For $\ell \in \mathcal{H}(\Omega)^*$, the squared norm of the error functional on V_n is

$$\ell^x \ell^y K(x,y) - \sum_{j=1}^{n} \ell^2(v_j) = \sum_{j>n} \ell^2(v_j)$$

if we work on a completed orthonormal basis $\{v_j\}_j$ of $\mathcal{H}(\Omega)$, and we work with the normalization by

$$\|\ell\|_2^2 = \sum_{j=1}^{\infty} \ell^2(\varphi_j) \leq 1.$$

We can move from the eigenbasis to this basis using a matrix $A = (a_{ij})$ as

$$v_k = \sum_{j=1}^{\infty} a_{jk} \sqrt{\lambda_j} \varphi_j, \tag{2.8}$$

where $\sum_{j=1}^{\infty} a_{jk}^2 = \sum_{k=1}^{\infty} a_{jk}^2 = 1$ and $\sum_{k=1}^{\infty} a_{ik} a_{jk} = \delta_{ij}$.

Hence, the squared norm of the error functional on V_n is

$$\begin{aligned}
&\ell^x \ell^y K(x,y) - \sum_{j=1}^{n} \ell^2(v_j) \\
=\ & \sum_{k>n} \ell^2(v_k) \\
=\ & \sum_{i,j=1}^{\infty} \sqrt{\lambda_i} \ell(\varphi_i) \sqrt{\lambda_j} \ell(\varphi_j) \sum_{k>n} a_{ik} a_{jk}.
\end{aligned}$$

where we used the representation (2.8) to move from a norm to a bilinear form.

To play it back to Linear Algebra and sequence spaces, we define $c_j :=$ $\ell(\varphi_j)$, $A_{ij} := a_{ij}$, $(A_n)_{ij}; = \begin{cases} a_{ij} & if \quad 1 \leq j \leq n \\ 0 & if \quad j > n \end{cases}$ and $\Lambda_{ij} = \lambda_j \delta_{ij}$ so that

$$\sum_{i,j=1}^{\infty} c_j c_i \left(\sum_{k=1}^{\infty} \sqrt{\lambda_i}\sqrt{\lambda_j} a_{ik} a_{jk} - \sum_{k=1}^{n} \sqrt{\lambda_i}\sqrt{\lambda_j} a_{ik} a_{jk} \right) =$$
$$= c^T (\sqrt{\Lambda} A A^T \sqrt{\Lambda} - \sqrt{\Lambda} A_n A_n^T \sqrt{\Lambda})c = \|(A^T\sqrt{\Lambda} - A_n^T\sqrt{\Lambda})c\|_{\ell_2}^2,$$

i.e., we come back from the bilinear form to the norm.

Now we want to maximize $\|(A^T\sqrt{\Lambda} - A_n^T\sqrt{\Lambda})c\|_{\ell_2}^2$ under the constraint $\sum_{j=1}^{\infty} c_j^2 \leq 1$ and then minimize the result over the a_{jk}, where, more generally, $\sum_{k=1}^{\infty} a_{ik} a_{jk} = \delta_{ij}$.

So we are left with the classical problem

$$\inf_{\operatorname{rank} A_n = n} \ \sup_{\|c\|_{\ell_2} \leq 1} \ \|(A^T\sqrt{\Lambda} - A_n^T\sqrt{\Lambda})c\|_{\ell_2},$$

whose solution is known (see e.g. [30, Theorem 2.1]). As in the finite dimensional setting, the singular value decomposition provides the best n-rank approximation of a matrix in the 2-norm, i.e., the value of the optimization problem is the square root of the $(n+1)$-th singular value of the matrix $A^T\sqrt{\Lambda}$, and the optimal subspace is spanned by the first n singular vectors. Since $(A^T\sqrt{\Lambda})^T(A^T\sqrt{\Lambda}) = \Lambda$, the singular values are precisely the kernel eigenvalues $\{\lambda_k\}_k$ and the singular vectors are the elements of the coordinate basis. Going back to the functional setting, these vectors correspond to the elements of the eigenbasis. $\square$

Finally, even if to prove this result we needed to consider functionals that are bounded with respect to $\|\cdot\|_2$, while assuming boundedness on $\mathcal{H}(\Omega)$ is not enough, as in the case of standard approximation it is in fact enough to require boundedness on a space intermediate between $\mathcal{H}(\Omega)$ and $L_2(\Omega)$, i.e.,

$$\|\ell\|_\mu^2 = \sum_{j=1}^{\infty} \mu_j \ell^2(\varphi_j) \leq 1,$$

with the same notation of the previous Section. In this case the value of the width is $\sqrt{\lambda_{n+1}/\mu_{n+1}}$ and the optimal subspace is still E_n.

Chapter 3

Algorithms

This Chapter introduces two algorithms for the construction of a suitable approximation of the eigenbasis.

We will work in a subspace $V(X_N)$ of $\mathcal{H}(\Omega)$, with X_N a set of points in Ω. In this subspace it is possible to treat the kernel through Linear Algebra operations, thus most of the results of the previous Chapter readily translate to actual computational processes.

The algorithm of Section 3.2 is a direct implementation of the results of Section 2.3, and its goal is to explicitly construct the eigenbasis of $V(X_N)$. The algorithm of Section 3.3 is based instead on a different strategy, that relies on the Nyström method (see e.g. [2, Section 11.4]), and it is intended to approximate the true eigenbasis. This algorithm is the first one that has been developed, and it will be presented as it was intended. But it can be reinterpreted, to some extent, as a particular instance of the first one. This connection will be explained in Section 3.4.

To construct their approximation of the eigenbasis, the algorithms are in fact designed to produce a suitable basis of the subspace $V(X_N)$. We will then start this Chapter by recalling some background facts on the general theory of change of basis for $\mathcal{H}(\Omega)$, as developed in [44].

Moreover, both the algorithms are at a nearly theoretical level. To be more precise, they are useful to understand how to perform actual computations in $\mathcal{H}(\Omega)$, but they are not completely suitable for a plain implementation. In particular, the first one can suffer of a severe instability, while both of them are computationally too expensive when N becomes large. To this end, in the next Chapter we will present approximated versions of both of them. These new versions solve these problems and are the ones which are actually used in practice.

The main part of the contents of this Chapter has been published in the papers [16] and [52], and it is partially taken from these works. We remark that part of the results of the first paper about the *WSVD basis* has been developed in the author's Master's Thesis. Nevertheless, we recall here

the main ideas of that work because they are the basis of one of the new computational strategies of the next Chapter, and also because they can be interpreted in a new way thanks to the results of the previous Chapter. This connection, described in Section 3.4, is not present in the aforementioned paper.

3.1   Background facts on change of basis

The general theory on change of basis for native spaces has been introduced in [44, 45]. The first paper deals with positive definite kernels and it is the one we will refer to, while the second one analyses the conditionally positive definite case, and it is not connected to the following results.

Among the results discussed in details in this paper, we stress here the connection between a change of basis and a decomposition of the kernel matrix A, together with a characterization of such bases.

For a given basis $\{v_j\}_{j=1}^N$ of $V(X_n)$ and $x \in \Omega$, we will denote as $\mathcal{T}(x)$ and $\mathcal{V}(x)$ the row vectors containing the evaluations of the standard basis of kernel translates and of this basis, i.e.,

$$\mathcal{T}(x) = [K(x, x_1), \ldots, K(x, x_N)] \in \mathbb{R}^N,$$
$$\mathcal{V}(x) = [v_1(x), \ldots, v_N(x)] \in \mathbb{R}^N.$$

A complete characterization of all the possible bases of $V(X_N)$ can be described by linear algebra operations on the matrix A. Namely, since $V(X_N)$ is a finite dimensional subspace, there exist a square, N dimensional, and invertible matrix C_v that represents the new basis in terms of the basis of translates. If we also define the square matrix of evaluation of the new basis on the points X_N, say $V_v = [v_j(x_i)]_{i,j=1}^N$, we have the following Theorem, proven in [44].

Theorem 3.1. *Any basis $\{v_j\}_{j=1}^N$ of $V(X_N)$ arises from a factorization of the kernel matrix A, i.e.,*

$$A = V_v C_v^{-1},$$

where $V_v = [v_j(x_i)]_{i,j=1}^N$ and $\mathcal{V}(x) = \mathcal{T}(x) \cdot C_v$, $x \in \Omega$.

Given this result, it is possible to classify, in some sense, all the bases in terms of their corresponding factorizations. In particular, we are interested in the following cases, that are treated in details in the cited paper.

The main class of bases we are interested in are obviously the ones that are orthonormal in $\mathcal{H}(\Omega)$.

Theorem 3.2. *Each $\mathcal{H}(\Omega)$ - orthonormal basis $\{v_j\}_{j=1}^N$ of $V(X_N)$ arises from a symmetric decomposition*

$$A = B^T B,$$

with $V_v = B^T$ and $C_v = B^{-1}$.

Moreover, it is possible to characterize also the bases that are orthonormal with respect to the discrete inner product defined by X_N, that is to say, for $u, v \in \mathcal{H}(\Omega)$,

$$(u, v)_{\ell_2(X_N)} = \sum_{i=1}^{N} u(x_i)v(x_i).$$

They can be characterized as follows.

Theorem 3.3. *Each $\ell_2(X_N)$-orthonormal basis $\{v_j\}_{j=1}^{N}$ of $V(X_N)$ arises from a decomposition*

$$A = QB,$$

where Q is an orthonormal matrix, $V_v = Q$ and $C_v = B^{-1}$.

Finally, it is possible to construct bases that enjoy both the above orthogonality properties.

Theorem 3.4. *Each basis $\{v_j\}_{j=1}^{N}$ of $V(X_N)$ that is $\mathcal{H}(\Omega)$ - orthonormal and $\ell_2(X_N)$-orthogonal arises from a singular value decomposition*

$$A = Q\Sigma^2 Q,$$

where $V_v = Q\Sigma$ and $C_v = Q^T\Sigma^{-1}$. Moreover, the $\ell_2(X_N)$-norms of the basis elements correspond to the square roots of the singular values, i.e., $\|v_j\|_{\ell_2(X_N)} = \sigma_j$.

We will keep the same notation in the following, thus a basis $\{v_j\}_{j=1}^{N}$ will be defined by simply exhibit its matrix C_v. We will drop the subscript v when no confusion is possible. Moreover, we will prove that Theorem 3.4 can be generalized by taking into account any couple of inner products on $V(X_N)$.

3.2 Orthogonal basis with respect to two general inner products

As a first step, we generalize here Theorem 3.4 to being able to deal with general inner products. We have the following result.

Theorem 3.5. *Let $(\cdot, \cdot)_a$ and $(\cdot, \cdot)_b$ be any couple of inner products on $V(X_N)$. There exists a basis $\{v_j\}_{j=1}^{N}$ of $V(X_N)$ which is b-orthonormal and a-orthogonal with norms $\{\sigma_j\}_{j=1}^{N}$.*

Proof. Consider the two Gramians of the standard basis with respect to the two inner products, i.e.,

$$G_a = [(K(\cdot, x_i), K(\cdot, x_j))_a]_{i,j=1}^{N},$$
$$G_b = [(K(\cdot, x_i), K(\cdot, x_j))_b]_{i,j=1}^{N}.$$

Following Theorem 3.1, we need to construct an invertible matrix C to express this new basis with respect to the standard one. To have the right orthogonality properties means to prove $G(V)_b = I$ and $G(V)_a = \Sigma^2$, where $G(V)_a, G(V)_b$ are the Gramians of $\{v_j\}_{j=1}^N$ with respect to the two inner products, Σ is the diagonal matrix having on the diagonal the a-norms of the new basis, and I is the identity matrix. Since

$$(v_i, v_j)_a = \sum_{h,k=1}^{N} C_{ih} C_{jk} (K(\cdot, x_h), K(\cdot, x_k))_a,$$

and the same holds for $(\cdot, \cdot)_b$, we have $G(V)_a = C^T G_a C$ and $G(V)_b = C^T G_b C$. So we are left to the problem of finding a matrix C such that $C^T G_a C = \Sigma^2$ and $C^T G_b C = I$, i.e., we need to compute a simultaneous diagonalization of the matrix pencil (G_a, G_b) with the right scaling of the diagonals.

Since the Gramians are symmetric and positive definite matrices this is always possible, e.g. in the following way:

- $G_b = LL^T$ be a Cholesky decomposition,

- define $M = L^{-1} G_a L^{-T}$ (which is symmetric and positive definite),

- let $M = U \Sigma^2 U^T$ be a SVD decomposition,

- define $C = L^{-T} U$.

In fact, this C is the correct one, since we have

$$G(V)_a = C^T G_a C = U^T L^{-1} G_a L^{-T} U = U^T M U = \Sigma^2,$$

and

$$G(V)_b = C^T G_b C = U^T L^{-1} G_b L^{-T} U = U^T L^{-1} LL^T L^{-T} U = U^T U = I.$$

$\square$

This construction generalizes the ones of Theorem 3.4 in the following sense. If the two inner products are the one of $\mathcal{H}(\Omega)$ and $\ell_2(X_N)$, according to [44] we have

$$\begin{aligned}
G_{\ell_2(X_N)} &= [(K(\cdot, x_i), K(\cdot, x_j))_{\ell_2(X_N)}]_{i,j=1}^N = A^2, \\
G_{\mathcal{H}} &= [(K(\cdot, x_i), K(\cdot, x_j))_{\mathcal{H}}]_{i,j=1}^N = A.
\end{aligned}$$

In this particular case, the SVD of A is also a simultaneous diagonalization of the pencil (A^2, A), A being symmetric, and this implies the results of Theorem 3.4.

3.2.1 Direct construction of the discrete eigenbasis

The first algorithm is a particular instance of the construction of the previous Section and aims at a direct construction of the discrete eigenbasis $\{\varphi_{j,N}\}_{j=1}^N$ of $V(X_N)$.

Thanks to Lemma 2.9, we know that the discrete eigenbasis is the unique set of N functions in $V(X_N)$ which is orthonormal in $L_2(\Omega)$ and orthogonal in $\mathcal{H}(\Omega)$, where uniqueness is understood in the sense of uniqueness of the eigendecomposition of the integral operator (2.4). This means that we can simply apply the change of basis described above, with the two inner products being the ones of $\mathcal{H}(\Omega)$ and of $L_2(\Omega)$. In our case G_a is the usual kernel matrix, while

$$(G_b)_{ij} = (K(\cdot, x_i), K(\cdot, x_j))_{L_2}.$$

Thus, provided we know G_b, we can explicitly construct the basis. Observe that the entries of the $L_2(\Omega)$-Gramian of the kernel translates can be expressed in terms of the eigenbasis as

$$(G_b)_{ij} = (K(\cdot, x_i), K(\cdot, x_j))_{L_2} = \sum_{k=1}^{\infty} \lambda_k^2 \varphi_k(x_i)\varphi_k(x_j), \qquad (3.1)$$

or, in other words, as an *iterated kernel*, that is in particular an instance of the situation discussed in Section 2.5.1. Nevertheless, for a general kernel it is not the case that we can explicitly compute the latter Gramian, so one can use a large set of points $X_M \subset \Omega$, $M >> N$ to approximate the $L_2(\Omega)$ inner product by its discrete counterpart.

Moreover, observe that, for practical use, it is more convenient to swap the role of G_a and G_b, i.e., of $\mathcal{H}(\Omega)$ and $L_2(\Omega)$, in the algorithm. In this way we construct the basis $\{\sqrt{\lambda_{j,N}}\varphi_{j,N}\}_{j=1}^N$, which is $\mathcal{H}(\Omega)$-orthonormal hence more suitable for approximation purposes, and, moreover, we obtain directly the eigenvalues of order N as $\Sigma = \text{diag}(\lambda_{j,N}, j = 1, \ldots, N)$.

As an example, in Figure 3.1 we show the first four elements of this basis. The basis is constructed using the Gaussian kernel with $\varepsilon = 1$, on the unit disk in $\mathbb{R}^2$, with X_N a regular grid of the unit square restricted to the disk. The figures were produced with a Matlab implementation of the present algorithm, that can be downloaded from [51]. Further computational details and numerical aspects are analyzed in the next Chapter.
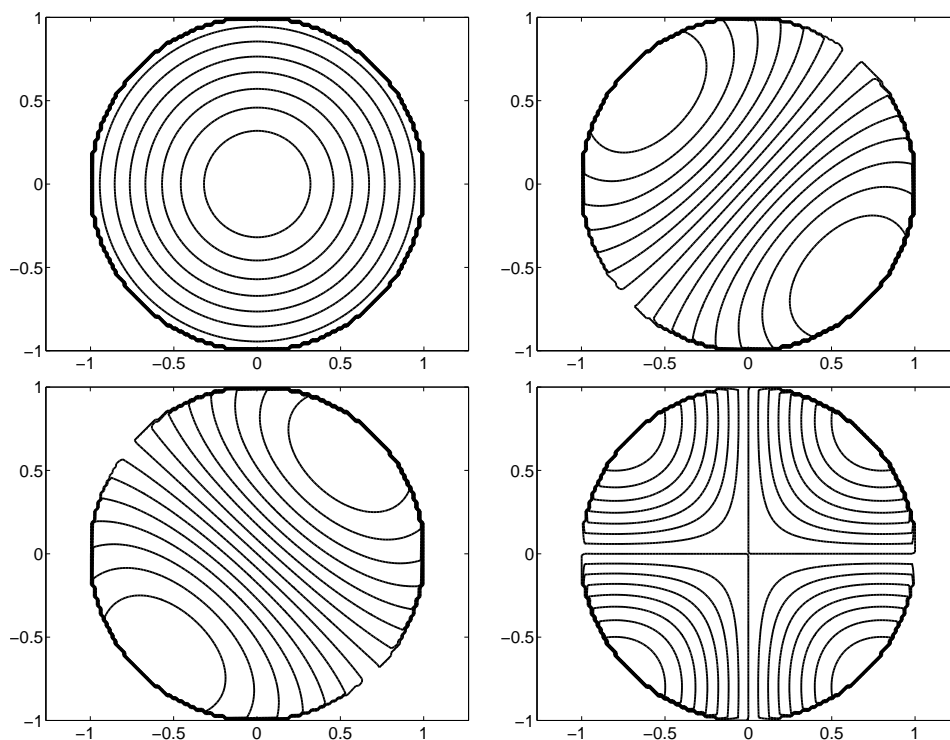
Figure 3.1: The first four basis elements of the discrete eigenbasis of $V(X_N)$ for the Gaussian kernel with $\varepsilon = 1$ on the unit disk, constructed with the algorithm of Section 3.2.1.

Finally, we remark that this algorithm could benefit from the generalization of Section 2.5.1. Indeed, if we know both the kernels K_λ and K_μ, for some $\{\mu_j\}_j$, we can use this algorithm without the need of computing $L_2(\Omega)$ inner products, but using the two kernel norms, which require only kernel evaluations.

3.3   Weighted Singular Value Decomposition basis

We present here a different algorithm, that was designed to construct a suitable approximation of the eigenbasis, and to use this approximation to stabilize the standard kernel-based interpolation process. We start by presenting its construction as it was intended at first, while we will give in Section 3.4 a different point of view on this algorithm, that will provide a bridge with the first algorithm and with the contents of the previous Chapter.

3.3.1 Motivation and definition

The main idea for the construction of the new basis is the discretization of the integral equation

$$\lambda_j \varphi_j = T \varphi_j \tag{3.2}$$

by means of the *symmetric Nyström method* (see e.g [2, §11.4]).

To this aim, consider on Ω a cubature rule composed by the points X_N and by a set of positive weights $W_N = \{w_j\}_{j=1}^N$, i.e., roughly speaking, a set of points and weights such that

$$\int_\Omega f(x)\, dx \approx \sum_{j=1}^N f(x_j) w_j, \quad f \in \mathcal{H}(\Omega)\,.$$

We are on purpose vague on the above assumption on the exactness of the formula. Indeed, we will see in the end that, at the stage involved in this algorithm, any cubature rule that reproduces constant functions is equivalent.

By means of this formula one can approximate the integral equation for each index $j > 0$ using the symmetric Nyström method. First, (3.2) can be evaluated on X_N as

$$\lambda_j \varphi_j(x_i) = \int_\Omega K(x_i, y)\varphi_j(y)dy, \quad 1 \le i \le N,\ j > 0,$$

and then discretized using the cubature rule by

$$\lambda_j \varphi_j(x_i) \approx \sum_{h=1}^N K(x_i, x_h)\varphi_j(x_h)w_h, \quad 1 \le i, j \le N. \tag{3.3}$$

Now, letting $W = \mathrm{diag}(w_j)$, to find the approximation of the eigenvalues and eigenfunctions it suffices to solve the matrix eigenvalue problem

$$\lambda v = (AW)v,$$

where each of the N solutions (λ, v) will be an approximated eigenvalue and a column vector containing the values of the approximated eigenfunctions on the points of X_N.

Unfortunately, this approach does not lead directly to the connection between the discretized version of the eigenbasis and a basis of the subspace $V(X_N)$. In fact it involves a scaled version of the kernel matrix, that is $A \cdot W$, which is no longer symmetric and that cannot be described as a factorization of A, as required by the theory presented in Section 3.1. A solution is to rewrite (3.3) using the fact that the weights are positive as

$$\lambda_j(\sqrt{w_i}\varphi_j(x_i)) = \sum_{h=1}^N (\sqrt{w_i}K(x_i, x_h)\sqrt{w_h})(\sqrt{w_h}\varphi_j(x_h)) \quad \forall 1 \le i, j \le N,$$

and then to consider the corresponding eigenvalue problem. Namely, the weighted eigenvalue problem

$$\lambda \left(\sqrt{W} v \right) = \left(\sqrt{W} A \sqrt{W} \right) \left(\sqrt{W} v \right)$$

which is equivalent to the previous one, but now involving the symmetric and positive definite matrix $A_W = \sqrt{W} A \sqrt{W}$. This matrix is normal, then a *singular value decomposition* of A_W is also a *unitary diagonalization*.

Motivated by this approach we can introduce the *weighted SVD basis* for $V(X_N)$, which is defined in the following in terms of the notation given in Theorem 3.1. Observe that the condition $\text{tr}(W) = |\Omega|$, $|\Omega|$ being the Lebesgue measure of Ω, is equivalent to require the exactness of the cubature formula for constant functions.

Definition 3.6. *The weighted SVD basis* $\{v_j\}_{j=1}^N$ *of* $V(X_N)$ *is characterized by the matrix*

$$C = \sqrt{W} Q \Sigma^{-1},$$

where

$$A_W = Q \Sigma^2 Q^T$$

is a singular value decomposition (and a unitary diagonalization) of the scaled kernel matrix $A_W = \sqrt{W} A \sqrt{W}$, *where* W *is a diagonal matrix with positive diagonal and* $\text{tr}(W) = |\Omega|$.

Figure 3.2 shows the first 4 elements of the WSVD basis, produced with the same setting as Figure 3.1, and with uniform weights $w_i = \pi/N$. The figures were produced with a Matlab implementation of the present algorithm, that can be downloaded from [49].
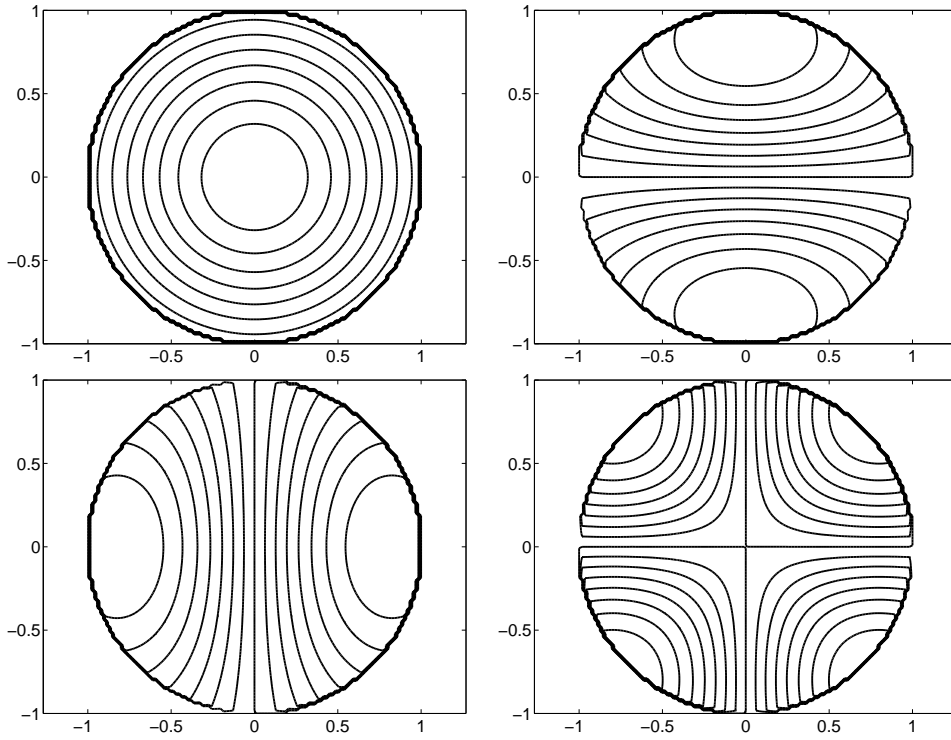
Figure 3.2: The first four WSVD basis elements for the Gaussian kernel with $\varepsilon = 1$ on the unit disk, constructed with uniform weights.

3.3.2 Properties

As expected, the WSVD basis preserves some interesting properties of the eigenbasis in a discrete version, as stated in the next Theorem. From now on, we denote as $\ell_{2,w}(X_N)$ the $\ell_2(X_N)$ inner product weighted with the weights $\{w_j\}_{j=1}^N$.

Theorem 3.7. *Let $\{v_j\}_{j=1}^N$ be a WSVD basis with weights $\{w_j\}_{j=1}^N$. We have the following*

(i) $v_j = \sigma_j^{-2} \sum_{i=1}^N w_i v_j(x_i) K(\cdot, x_i), \ 1 \leqslant j \leqslant N$;

(ii) *the basis is $\mathcal{H}(\Omega)$-orthonormal;*

(iii) *the basis is $\ell_{2,w}(X_N)$-orthogonal;*

(iv) $\|v_j\|_{\ell_{2,w}(X_N)}^2 = \sigma_j^2, \ 1 \leqslant j \leqslant N$;

(v) *if K is translational invariant, $\sum_{j=1}^N \sigma_j^2 = K(0,0) \, |\Omega|$.*

Proof. Properties (ii), (iii) and (iv) can be proved using the expression for the Gramians computed in [44] and observing that the entries of the $\ell_{2,w}(X_N)$-Gramian of the standard kernel basis are given by

$$G_{\ell_{2,w}(X_N)} = (K(\cdot,x_i), K(\cdot,x_j))_{\ell_{2,w}(X_N)} = \sum_{h=1}^{N} w_h K(x_h,x_i)K(x_h,x_j) = (AWA)_{ij}.$$

Indeed,

$$G(V)_{\mathcal{H}} = C^T G_{\mathcal{H}} C = C^T A C = \Sigma^{-1} Q^T \sqrt{W} A \sqrt{W} Q \Sigma^{-1} = I$$

and

$$
\begin{aligned}
G(V)_{\ell_{2,w}(X_N)} &= C^T G_{\ell_{2,w}(X_N)} C = C^T AWAC \\
&= \Sigma^{-1} Q^T \sqrt{W} A \sqrt{W} \sqrt{W} A \sqrt{W} Q \Sigma^{-1} \\
&= \Sigma^{-1} Q^T Q \Sigma^2 Q^T Q \Sigma^2 Q^T Q \Sigma^{-1} = \Sigma^2.
\end{aligned}
$$

To prove Property (i) it suffices to use the definition of C and $V = AC = \sqrt{W} Q \Sigma$, where we used the relation between C, V and A provided by Theorem 3.1. Indeed, denoting by a subscript j the j-th column of a matrix, we get

$$V = \sqrt{W^{-1}} Q \Sigma = \sqrt{W^{-1}}[Q_1 \sigma_1, \ldots, Q_N \sigma_N],$$

hence $V_j = (\sqrt{W})^{-1} Q_j \sigma_j$, we have

$$\sigma_j^{-1} Q_j = \sigma_j^{-2} \sqrt{W} V_j.$$

Using the last equality we can compute each component of C as

$$C_{ij} = (\sqrt{W} Q \Sigma^{-1})_{ij} = \sqrt{w_i}\,\frac{Q_{ij}}{\sigma_j} = \frac{w_i}{\sigma_j^2}\,v_j(x_i),$$

and then, by the definition of the WSVD basis,

$$
\begin{aligned}
v_j(x) &= \sum_{i=1}^{N} K(x,x_i)\,C_{ij} = \sum_{i=1}^{N} K(x,x_i)\,\frac{w_i}{\sigma_j^2}\,v_j(x_i) = \\
&= \frac{1}{\sigma_j^2}\sum_{i=1}^{N} w_i\,K(x,x_i)\,v_j(x_i)\,.
\end{aligned}
$$

Finally property v is proven by another linear algebra arguments. Recalling that

$$\sqrt{W} A \sqrt{W} = Q \Sigma^2 Q^T\,,$$

and the fact that the trace of a square matrix is equal to the sum of its eigenvalues, we get

$$\sum_{j=1}^{N} \sigma_j^2 = \sum_{j=1}^{N} w_j\,K(x_j,x_j) = K(0,0)\sum_{j=1}^{N} w_j = K(0,0)\,|\Omega|\,.$$

This concludes the proof.   □

It is important to notice that, $\{w_j\}_{j=1}^N$ being cubature weights, the inner product of $\ell_{2,w}(X_N)$ is a discretization of the $L_2(\Omega)$ inner product. Indeed, in view of this fact, property (i) is a discrete version of the eigenvalue problem (3.2). Moreover, for a translational invariant kernel property (v) tells us that the sum of the WSVD eigenvalues equals the sum of the true eigenvalues, i.e.,

$$\sum_{j=1}^N \sigma_j^2 = \int_\Omega K(x,x)\,dx = \sum_{j=1}^\infty \lambda_j.$$

This summation property allows to control the residual when truncating the WSVD basis, as will be explained in the next Section.

Apart from this interpretation of Theorem 3.7 as a discrete version of the properties of the eigenbasis, we remark once again that to construct the basis we require just that the weights $\{w_i\}_{i=1}^N$ are positive and able to reproduce constants. For what follows, it is possible to use weights not related to a cubature rule, e.g. by just taking $w_j = |\Omega|/N$. In this way no connection can be expected between the WSVD basis and the eigenbasis, while the approximation properties analyzed below remain unchanged.

3.3.3 Weighted discrete least-squares approximation

Now we can introduce a weighted discrete least-squares operator that turns out to be strictly related to the weighted basis just introduced. The goal is to avoid solving a full interpolation problem, and instead to project the unknown function $f \in \mathcal{H}(\Omega)$ into a properly chosen subspace of $V(X_N)$. This is done in order to obtain better results in terms of stability, without serious loss of convergence speed. This kind of approximation is meaningful when the data values are supposed to be affected by noise, or when the kernel matrix A is seriously ill-conditioned.

For a function $f \in \mathcal{H}(\Omega)$, the standard interpolation $s_N f$ is given by the projection into $V(X_N)$, i.e., thanks to the $\mathcal{H}(\Omega)$-orthonormality of the basis, $s_N f = \sum_{j=1}^N (f, v_j) v_j$. We can instead define a weighted least-squares approximation as follows.

Definition 3.8. *Consider a set of weights W_N and its associated WSVD basis $\{v_j\}_{j=1}^N$. For $f \in \mathcal{H}(\Omega)$ and $n < N$, the weighted discrete least-squares approximation of order n of f is the function $\Lambda_n f$ that satisfies the condition*

$$\Lambda_n f = \underset{g \in \mathrm{span}\{v_1,\dots,v_n\}}{\arg\min} \|f - g\|_{\ell_{2,w}(X_N)}.$$

Thanks to the construction of the WSVD basis, it is possible to view this approximation process in terms of the standard interpolation process in $\mathcal{H}(\Omega)$. We have the following.

Lemma 3.9. *The elements v_j of the WSVD basis satisfy the following relation, for $1 \leq j \leq N$:*

$$(f, v_j) = \frac{1}{\sigma_j^2} (f, v_j)_{\ell_{2,w}(X_N)} \quad \text{for all } f \in \mathcal{H}(\Omega).$$

Proof. Using property (i) of Theorem 3.7 and by direct calculations we get:

$$
\begin{aligned}
(f, v_j) &= \left(f, \frac{1}{\sigma_j^2} \sum_{i=1}^{N} w_i v_j(x_i) K(\cdot, x_i) \right) = \frac{1}{\sigma_j^2} \sum_{i=1}^{N} w_i v_j(x_i) \left(f, K(\cdot, x_i) \right), \\
&= \frac{1}{\sigma_j^2} \sum_{i=1}^{N} w_i v_j(x_i) f(x_i) = \frac{1}{\sigma_j^2} (f, v_j)_{\ell_{2,w}(X_N)},
\end{aligned}
$$

where, by property (iv) of Theorem 3.7, $\sigma_j^2 = (v_j, v_j)_{\ell_{2,w}(X_N)}$. $\qquad\square$

Using this Lemma with the notation of Definition 3.8, it comes easy to compute the weighted discrete least-squares approximation of a function $f \in \mathcal{H}(\Omega)$.

Theorem 3.10. *The weighted discrete least-squares approximation of a function $f \in \mathcal{H}(\Omega)$ is given by*

$$\Lambda_n f = \sum_{j=1}^{n} \sigma_j^{-2}(f, v_j)_{\ell_{2,w}(X_N)} v_j = \sum_{j=1}^{n} (f, v_j) v_j, \tag{3.4}$$

that is $\Lambda_n f$ is nothing else but a truncation to the first n terms of the interpolant $s_N f$.

The idea of constructing this weighted discrete least-squares approximant can be automated when dealing with very small singular values. In this case, in order to avoid numerical instability, we can leave out the basis corresponding to singular values less than a preassigned tolerance, skipping automatically from interpolation to discrete least-squares approximation. From a linear algebra point of view, this corresponds to solving the (weighted) linear system associated to the interpolation problem using the *total least-squares* method (see e.g. [31]).

Thanks to the last Theorem, we can give error bounds and stability bound on the approximation process through a WSVD basis. The following results is a simple consequence of the previous Theorem and of the results of Section 2.1, although there is no reason in general to expect such a relation if we consider a different kind of least-squares approximant in an arbitrary basis.

The following Corollary expresses the mentioned bounds in terms of truncations of the standard Power Function, where it is clear that by replacing an exact interpolant with a weighted discrete least-squares approximant, we can obtain better results in terms of stability.

Corollary 3.11. *Given a function $f \in \mathcal{H}(\Omega)$, consider its weighted discrete least-squares approximant of order $n \leq N$. The point-wise approximation error can be bounded in the following way*

$$|f(x) - \Lambda_n f(x)| \leq \sqrt{K(x,x) - \sum_{j=1}^{n} v_j(x)^2} \, \|f\| \quad \text{for all } x \in \Omega, \qquad (3.5)$$

and we have a stability bound

$$|\Lambda_n f(x)| \leq \sqrt{\sum_{j=1}^{n} v_j(x)^2} \, \|f\| \quad \text{for all } x \in \Omega.$$

Finally, we can control the approximation error also in terms of the $\ell_{2,w}(X_N)$-norm itself.

Proposition 3.12. *If K is translational invariant, for $n \leq N$ and $f \in \mathcal{H}(\Omega)$ we have*

$$\|f - \Lambda_n f\|_{\ell_{2,w}(X_N)} \leq \sqrt{\sum_{j=n+1}^{N} \sigma_j^2} \, \|f\| \qquad (3.6)$$

Proof. For $f \in \mathcal{H}(\Omega)$, inequality (3.5) gives

$$\|f - \Lambda_n f\|_{\ell_{2,w}(X_N)}^2 = \sum_{i=1}^{N} w_i (f(x_i) - \Lambda_n f(x_i))^2$$

$$\leq \sum_{i=1}^{N} w_i \left(K(x_i, x_i) - \sum_{j=1}^{n} v_j(x_i)^2 \right).$$

The rightmost term can be divided in the terms

$$\sum_{i=1}^{N} w_i K(x_i, x_i) = \sum_{j=1}^{N} \sigma_j^2,$$

where we used the relation between the trace and the eigenvalues of the scaled kernel matrix, and

$$\sum_{i=1}^{N} w_i \sum_{j=1}^{n} v_j(x_i)^2 = \sum_{j=1}^{n} \sum_{i=1}^{N} w_i v_j(x_i)^2 = \sum_{j=1}^{n} \sigma_j^2$$

where we used property iv. The difference of these two terms gives the bound of the statement. $\square$

This bound is of limited interest as it is, since an exact interpolant (i.e., $n = N$) always gives an exact reconstruction in this norm. Nevertheless, it is interesting for the connection it provides with the theory of the previous Chapter, as explained in the next Section.

3.4   Connections between the two algorithms

We conclude this Chapter by showing the relation between the WSVD basis and the methods developed in Chapter 2 and in Section 3.2.

The first connection is at an algorithmic level, and it proves that, in a particular case, the two algorithms of this Chapter are, in fact, the same. Indeed, in the direct construction of the eigenbasis of $V(X_N)$ of Section 3.2.1, we needed to consider the $L_2(\Omega)$ Gramian of the basis of kernel translates. To this end, in the likely situation when we do not know explicitly the entries of the $L_2(\Omega)$-Gramian (see (3.1)), we need to approximate them by means of another point set X_M, $M >> N$. If instead we decide to use the points X_N themselves, we make the approximation assumption $(\cdot, \cdot)_{L_2} \approx (\cdot, \cdot)_{\ell_2(X_N)}$. That is to say, the first algorithm becomes a particular instance of the second one, where the weights used to construct the WSVD basis are simply $w_i = |\Omega|/N$. In particular, as observed in Section 3.2, there is no need to compute a simultaneous diagonalization of two Gramians to construct the approximation of the discrete eigenbasis. In other terms, the use of a second Gramian (the one of $L_2(\Omega)$) does not provide more information than using simply the kernel matrix A, so we can not expect that the WSVD basis is as good an approximation of the true eigenbasis as the discrete eigenbasis $\{\varphi_{j,N}\}_{j=1}^N$ is. Nevertheless, as we will explain in the next Chapter, the approximated version of the WSVD basis provides a fast and stable way of solving approximation problems in $\mathcal{H}(\Omega)$.

The second connection is more theoretical, and it is related to the contents of the previous Chapter. Namely, the space spanned by the WSVD basis elements minimizes some widths defined as in Section 2.2 by replacing $L_2(\Omega)$ with $\ell_{2,w}(X_N)$.

First, observe that the discrete weighted least-squares approximation operator Λ_n is a particular instance of the $\ell_{2,w}(X_N)$-projector $\Pi_{\ell_{2,w}(X_N),V_n}$, where for $V_n \subset \mathcal{H}(\Omega)$

$$\Pi_{\ell_{2,w}(X_N),V_n} f = \sum_{j=1}^N (f, u_j)_{\ell_{2,w}(X_N)} u_j$$

and $\{u_j\}_{j=1}^N$ is any $\ell_{2,w}(X_N)$-orthonormal basis of V_n. Hence, we have $\Lambda_n = \Pi_{\ell_{2,w}(X_N),W_n}$, where $W_n \subset V(X_N)$ is the subspace spanned by the first n elements of the WSVD basis.

The following statement is a Corollary of the Theorems of Section 2.2, and it is in accordance with these results. Notice that, due to the notation used in the definition of the WSVD basis, the terms σ_j are squared in the following equalities.

Corollary 3.13. *For $n < N$, we have*

$$\inf_{\substack{V_n \subset \mathcal{H} \\ \dim(V_n)=n}} \sup_{f \in S(\mathcal{H})} \|f - \Pi_{\mathcal{H},V_n} f\|_{\ell_{2,w}(X_N)} = \sqrt{\sigma_{n+1}^2}$$

$$\inf_{\substack{V_n \subset \mathcal{H} \\ \dim(V_n)=n}} \left\| \sup_{f \in S(\mathcal{H})} |f(\cdot) - \Pi_{\mathcal{H},V_n} f(\cdot)| \right\|_{\ell_{2,w}(X_N)} = \sqrt{\sum_{j=n+1}^{N} \sigma_j^2}$$

and W_n is an optimal subspace.

Proof. The proof is the same of the ones of Theorems 2.6 and 2.8, where the $L_2(\Omega)$ norm is replaced by the $\ell_{2,w}(X_N)$ norm, and the eigenbasis is replaced by the WSVD basis. Furthermore, observe that the $\ell_{2,w}(X_N)$ norm only considers the values of a function f on X_N, so we can assume $f \in S(V(X_N))$. $\qquad\square$

Chapter 4

Computational aspects and numerical experiments

In this Chapter we present three algorithms that can be used for the numerical construction of an approximation of the eigenbasis, and to effectively solve approximation problems in $\mathcal{H}(\Omega)$.

The three algorithms are designed to provide a computationally feasible implementation of the two methods of the previous Chapter. To be more precise, they approximate the computations required by the two methods in a suitable way, so that we can obtain the same results without the computational limitations of the exact algorithms.

In particular, the two methods of Section 4.1 are a greedy approximation of the algorithm of Section 3.2.1. The two algorithms are different in the way they perform the greedy selection of a suitable point set in X_N, which is in both cases inspired by the results of [19]. Nevertheless, both of them are designed in order to avoid to compute a simultaneous diagonalization of two possibly large and ill-conditioned Gramians, but instead to perform the same operation on two properly chosen submatrices. This approximation allows to effectively construct the discrete eigenbasis, and we will present some numerical experiments that support the results of Chapter 2.

The one of Section 4.2, instead, is an approximation of the WSVD basis algorithm of Section 3.3. It replaces the exact computation of the singular value decomposition with a proper approximation given by the Lanczos algorithm (see e.g. [32]). The properties of the exact WSVD basis are in this way slightly modified, while the approximation capabilities of this new version of the basis remain untouched, with moreover a computational speed up in the approximation process. We will present some numerical experiments that test this algorithm.

The contents of this Chapter are published in the papers [52] and [17] and are partially taken from there. Moreover, it is possible to download a

Matlab implementation of the three algorithms from [49] and [51].

4.1 Greedy construction of the discrete eigenbasis

The algorithm of Section 3.2.1 requires to compute a simultaneous diagonalization of two $N \times N$ Gramians to construct the subspace $E_{n,N}$, and this approach easily becomes infeasible when N becomes large. Instead, here we first select n points in X_N such that working on these points is more stable than working with the full original matrix, and then solve the problem in $V(X_n)$. In this case, since we first operate a selection of n points in X_N, we can use X_N itself to approximate the $L_2(\Omega)$ inner product instead of choosing a different set X_M, $M > N$.

The selection of the set X_n is performed with a greedy construction of the Newton basis (see [41]). This basis is a particular $\mathcal{H}(\Omega)$-orthonormal basis of the native space, and it has the remarkable property of being recursively computable. Using this fact, we can select the points of X_n in a greedy way, compute a basis element at a time, and check at any step if a certain tolerance is reached or not.

Before analyzing the point selection strategies, we show how to construct the eigenspaces via the Newton basis. Assume that $v_1, \ldots, v_n$ is an $\mathcal{H}(\Omega)$-orthonormal set of functions (here, it will be the Newton basis for $V(X_n)$). Then

$$T_{V(X_n)}f(x) = \sum_{i=1}^{n} v_i(x) \sum_{j=1}^{n} (v_i, v_j)_{L_2}(f, v_j)$$

and if $\lambda_{j,n}$ is an eigenvalue with eigenfunction $\varphi_{j,n}$ then $T_{V(X_n)}\varphi_{j,n} = \lambda_{j,n}\varphi_{j,n}$ implies

$$\lambda_{j,n}(\varphi_{j,n}, v_i) = \sum_{k=1}^{n} (v_i, v_k)_{L_2}(\varphi_{j,n}, v_k).$$

Thus the coefficients of the eigenbasis with respect to this orthonormal basis are the eigenvectors of the $L_2(\Omega)$ Gramian of the $\{v_j\}_j$ basis.

Experimentally, the Newton basis is nearly $L_2(\Omega)$ orthogonal, thus the above procedure should have a nice Gramian matrix, provided that the $L_2(\Omega)$ inner products that are near zero can be calculated without loss of accuracy.

To select the points we use two similar greedy strategies, based on maximization of $L_\infty(\Omega)$ and $L_2(\Omega)$ norms of the Power Function. The first one has been introduced in [19], while the second one can be obtained by a simple modification. The first point is chosen as

$$x_1 = \arg\max_{x \in X_N} \left\| \frac{K(\cdot, x)}{\sqrt{K(x,x)}} \right\|_{L_\infty} \quad \text{or} \quad x_1 = \arg\max_{x \in X_N} \left\| \frac{K(\cdot, x)}{\sqrt{K(x,x)}} \right\|_{L_2}.$$

For $1 \leq i < n$, denoting by $X_i = \{x_1, \ldots, x_i\}$ the already chosen points, the $(i+1)$-th point is selected as

$$x_{i+1} = \arg\max_{x \in X_N \setminus X_i} \|v_{i+1}\|_{L_\infty}^2 \quad \text{or} \quad x_{i+1} = \arg\max_{x \in X_N \setminus X_i} \|v_{i+1}\|_{L_2}^2.$$

Both the selections are stopped when the relevant quantity, namely the $L_\infty(\Omega)$ or $L_2(\Omega)$ norm of the next Newton basis element, is below a certain tolerance.

The $L_2(\Omega)$ selection method is motivated by the observation that the eigenspaces are the $L_2(\Omega)$-maximizing orthonormal functions in $\mathcal{H}(\Omega)$. Nevertheless, numerical experiments suggest that the $L_\infty(\Omega)$ algorithm produces good enough points sets, and we will use the latter in the following, since is simpler and faster.

We remark that we do not yet have a satisfactory theoretical analysis of the convergence of this greedy algorithms to the direct solution, given by the eigenbasis.

4.1.1 Experiments

We consider the Matérn kernels of order $\beta = 0, 1, 2, 3$, whose native spaces on $\mathbb{R}^d$ are norm equivalent to the Sobolev spaces $H^{(\beta+d)/2}$. The kernels can be defined in radial form by the following formulas (see e.g. [22, Appendix D]),

$$
\begin{aligned}
\Phi_0(r) &= e^{-\varepsilon r} \\
\Phi_1(r) &= e^{-\varepsilon r}(1 + \varepsilon r) \\
\Phi_2(r) &= e^{-\varepsilon r}(3 + 3\varepsilon r + (\varepsilon r)^2) \\
\Phi_3(r) &= e^{-\varepsilon r}(15 + 15\varepsilon r + 6(\varepsilon r)^2 + (\varepsilon r)^3),
\end{aligned}
$$

where $K(x, y) = \Phi_\beta(\|x - y\|)$, $x, y \in \mathbb{R}^2$, for the corresponding β. For this experiment the value of the shape parameter is set to $\varepsilon = 1$.

In these spaces the asymptotic behavior of the Kolmogorov width, hence of the eigenvalues, is known as recalled in Section 2.4. We assume here that the same bounds hold in the unit disk, and we want to compare it with the discrete eigenvalues of $V(X_N)$.

All experiments have been performed with the Matlab code [51].

To perform the greedy selection we start from a grid of equally spaced points in $[-1, 1]^2$ restricted to the unit disk, so that the number of points inside the disk is $N \approx 10^4$. We use this grid both for point selection and to approximate the $L_2(\Omega)$ inner products as weighted $\ell_2(X_N)$ product, with weights $w_i = \pi/N$.

We then select $n = 200$ points by the greedy $L_\infty(\Omega)$ maximization of the Power Function in the unit disk. In Figure 4.1 we plot the set of the first 50 selected points for $\beta = 3$. As a comparison, we show also the first 50

points selected by the $L_2(\Omega)$ greedy algorithm, even if they are not used in the present experiments.
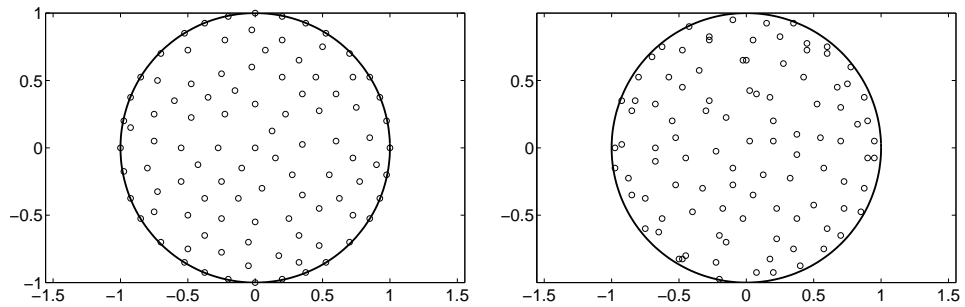

Figure 4.1: The first 50 points selected by the two greedy algorithms for the approximation of the eigenfunctions of the Matern kernel with $\beta = 3$. Maximization of the power function with respect the $L_\infty(\Omega)$ norm (left) and the $L_2(\Omega)$ norm (right).

The eigenvalues are then computed with the use of the Newton basis, as explained at the beginning of this Section, i.e., as eigenvalues of the $L_2(\Omega)$ Gramian of the Newton basis.

It is interesting to see (Figure 4.2) how the $L_2(\Omega)$ norms of the Newton basis and of the approximated eigenfunctions are related, after both the bases have been normalized with respect to the $\mathcal{H}(\Omega)$ norm. Although no information is known about the decay of the $L_2(\Omega)$ norms of the Newton basis, it can be clearly observed how they are strictly connected to the behavior of the eigenvalues.
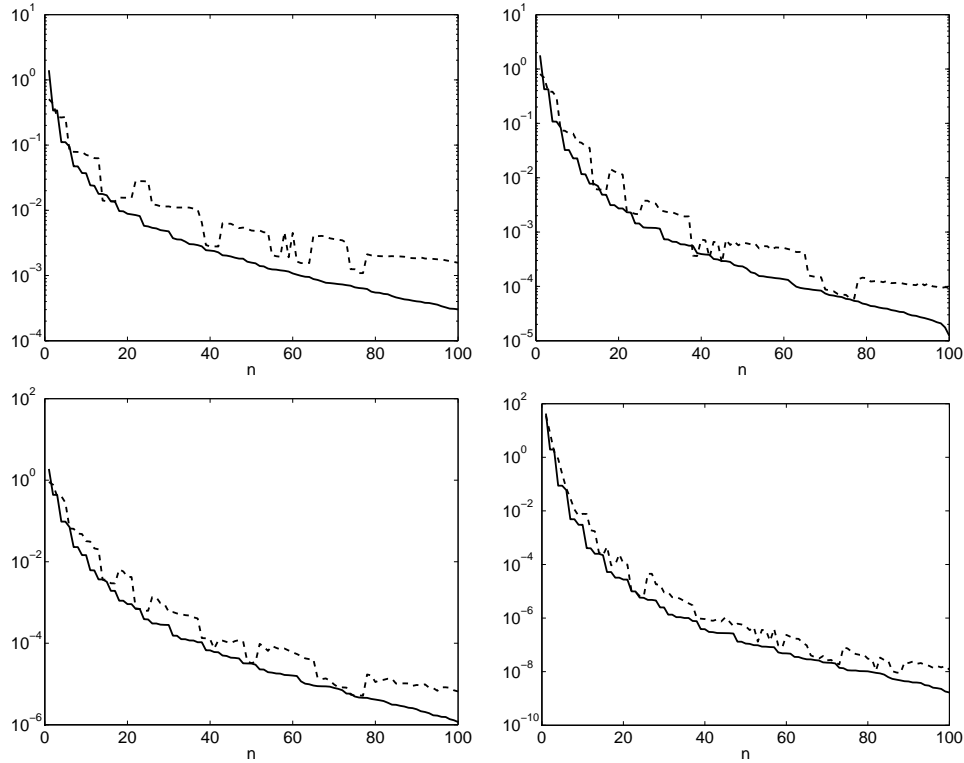
Figure 4.2: Comparison between the $L_2(\Omega)$ norms of the discrete eigenbasis (solid line) and the Newton basis (dotted line), both normalized in the $\mathcal{H}(\Omega)$ norm. From top left to bottom right: $\beta = 0, 1, 2, 3$.

Finally, we want to compare the theoretical behavior of the Kolmogorov width with the decay of the computed eigenvalues. The results are shown in Figure 4.3 and, as expected, for any order under consideration there exists a positive constant c such that the discrete eigenvalues decay with the same rate of the Sobolev best approximation.

Moreover, we expect that the discrete eigenvalues converge to the true ones with a rate that can also be controlled by β. Indeed, according to Corollary 2.15, we have

$$0 \le \lambda_j - \lambda_{j,V(X_n)} < c_1 n^{-\beta/d}, \quad 1 \le j \le n.$$

To verify this, we instead look at the decay of

$$\sum_{j=1}^{\infty} \lambda_j - \sum_{j=1}^{n} \lambda_{j,V(X_n)}.$$

since we can exactly compute the first term. Indeed, since the kernels are

radial, we have

$$\sum_{j=1}^{\infty} \lambda_j = \int_{\Omega} K(x,x)dx = \pi K(0,0).$$

Results are presented in Figure 4.4. From the experiments it seems that the actual convergence speed is somewhat faster than what expected, and in fact we obtain, in this particular case, a rate of order $(\beta + d/2)/d$ instead of β/d. This "gap of $d/2$" has been already observed in [55].
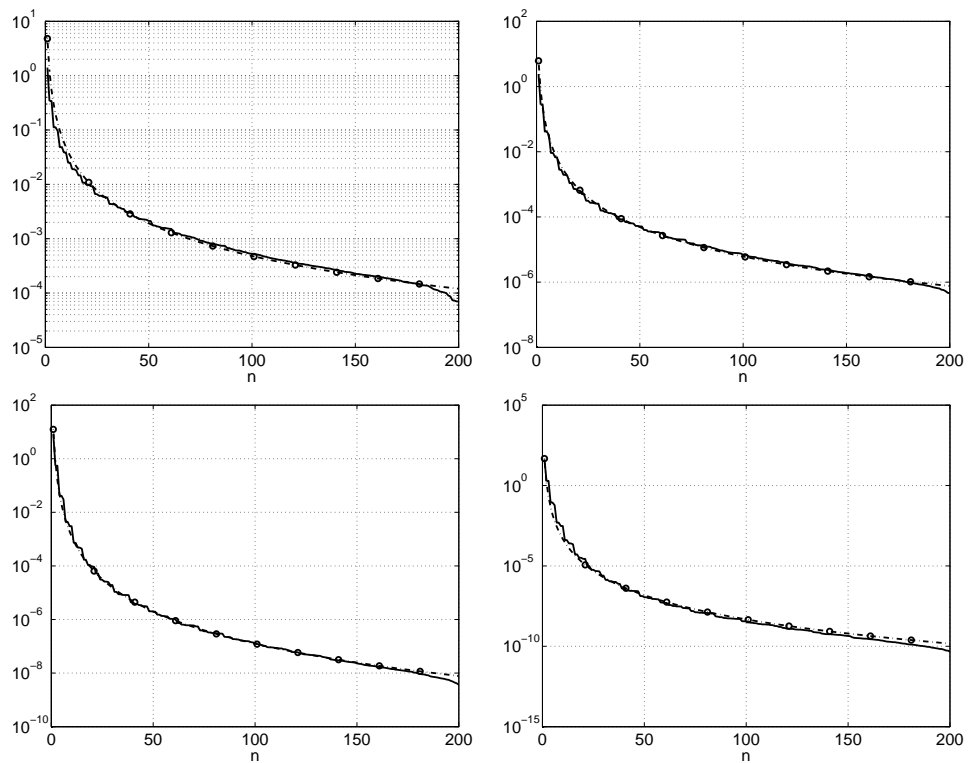

Figure 4.3: Decay of the discrete eigenvalues of the Matern kernels (solid line) compared with the theoretical decay rate $n^{-(\beta+d)/d}$ in the corresponding Sobolev spaces (circles). The theoretical bounds are scaled with a positive coefficient. From top left to bottom right: $\beta = 0, 1, 2, 3$.
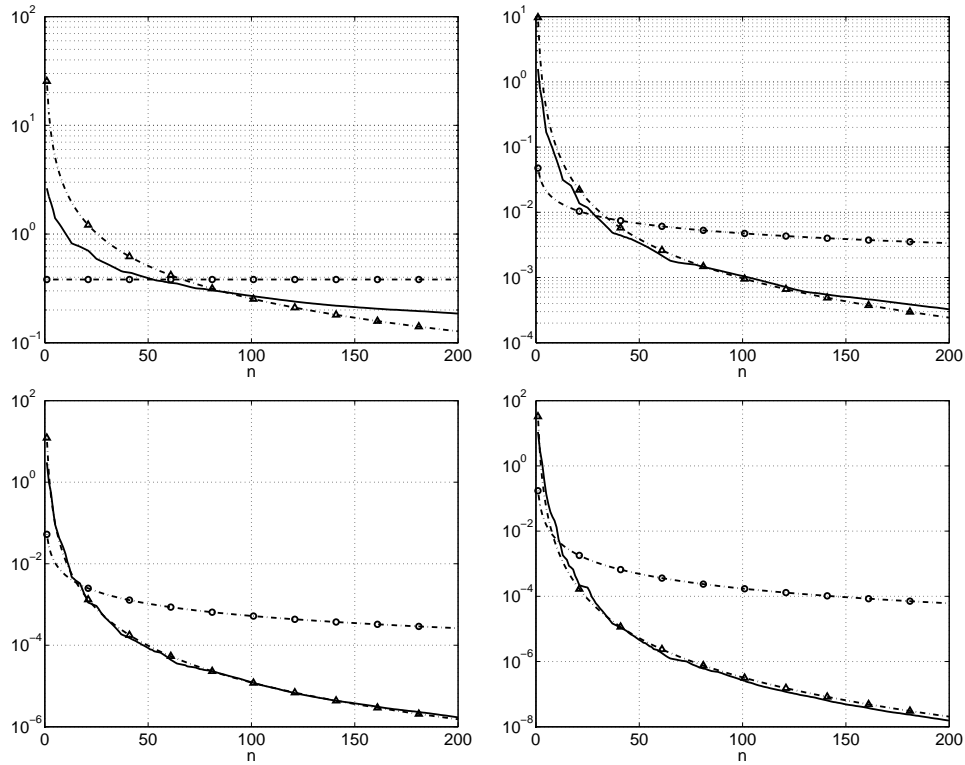
Figure 4.4: Difference between the sum of the real eigenalues and the discrete ones (solid line), for the Matern kernels, compared with the theoretical decay rate $n^{-\beta/d}$ in the corresponding Sobolev spaces (circles) and with $n^{-(\beta+d/2)/d}$ (triangles). The theoretical bounds are scaled with a positive coefficient. From top left to bottom right: $\beta = 0, 1, 2, 3$.

As an example, Figure 4.5 presents the first four elements of the eigenbasis computed by the algorithm in the same setting as above, and for $\beta = 3$.
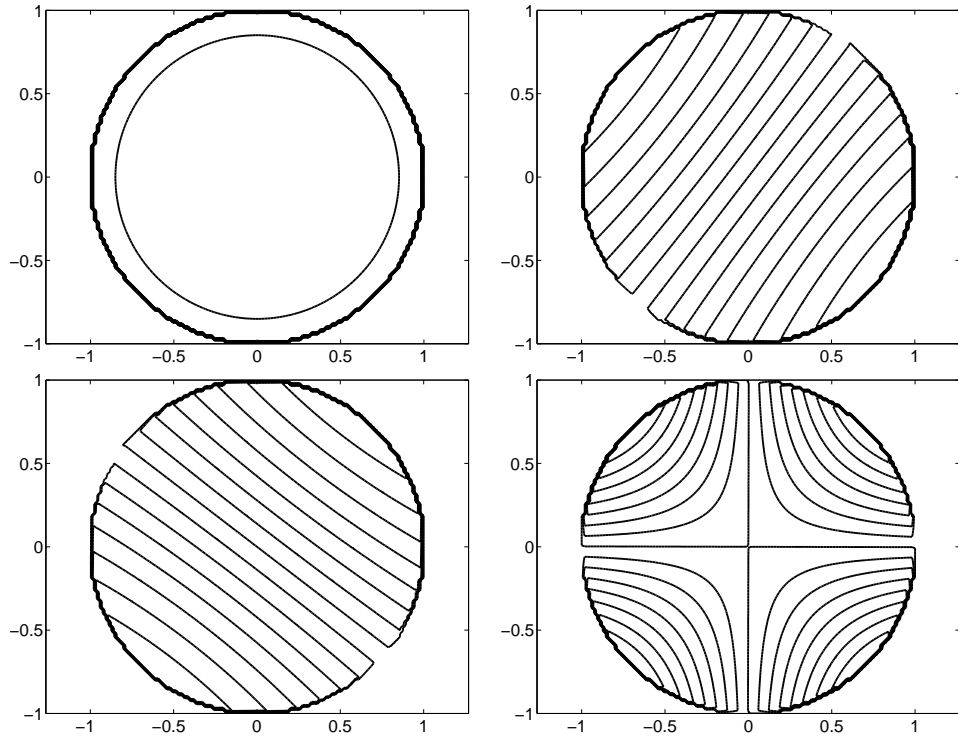
Figure 4.5: Contour plot of the first four approximated eigenfunctions (sorted from top left to bottom right) of the Matern kernel with $\beta = 3$.

Finally, we consider a situation where the Mercer decomposition is explicitly known. Namely, in the case of $\beta = 0$, the exact eigenbasis of the Matérn kernel on $\Omega = [-L, L]$ has been computed ([58, Section 3.4.1]), and it is given by the eigenvalues

$$\lambda_j = \begin{cases} \frac{2\varepsilon}{\varepsilon^2 + \alpha_j^2}, & j \text{ even} \\ \frac{2\varepsilon}{\varepsilon^2 + \beta_j^2}, & j \text{ odd} \end{cases}$$

and the eigenfunctions

$$\varphi_j = \begin{cases} \sqrt{\frac{2\alpha_j}{2L\alpha_j - \sin(L\alpha_j)}} \sin(\alpha_j \cdot), & j \text{ even} \\ \sqrt{\frac{2\beta_j}{2L\beta_j + \sin(L\beta_j)}} \cos(\beta_j \cdot), & j \text{ odd} \end{cases},$$

where $\{\alpha_j\}_j$, $\{\beta_j\}_j$ are defined as solution of two families of nonlinear equations (see Section 2.2 of [23, Appendix A] for further details). We consider a point set X_N a set of $N = 10^4$ equally spaced points in $[-L, L]$, with $L = 2$ and $\varepsilon = 0.5$. The values of $\{\alpha_j\}_j$, $\{\beta_j\}_j$ have been computed by a numerical solver up to $j = 5 \cdot 10^3$, which correspond to an average RMS error on X_N between the exact kernel and this truncated decomposition of

$9.61 \cdot 10^{-7}$. To compute our approximation we start from the closed form of the kernel, $K(x, y) = e^{-\varepsilon \|x-y\|_2}$, and we employ the $L_\infty(\Omega)$-greedy algorithm on the points X_N to compute E_n up to $n = 100$. The results for the approximation of the eigenvalues and the eigenfunctions are shown in Figure 4.6. As it is reasonable to expect, the approximation is good for the first eigencouples (RMSE of the order of 10^{-5} for the eigenfunctions and relative error of 10^{-6} for the eigenvalues), while it degrades up to 10^{-1} for bigger indexes j.
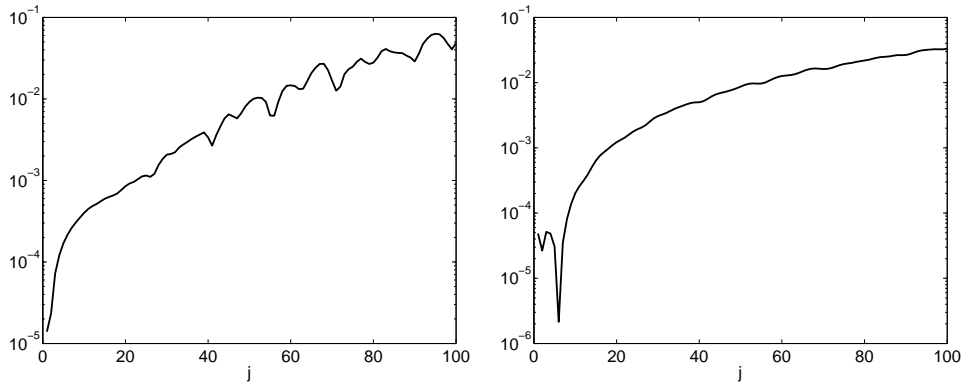

Figure 4.6: Approximation of the eigenfunctions (left, RMSE error) and eigenvalues (right, relative error) of the Matérn kernel with $\beta = 0$.

4.2 Fast computation of the WSVD basis

We concentrate here on the construction of an approximated version of the WSVD basis of Section 3.3. The main problem of the true WSVD basis is the efficiency of the computation. Namely, since the basis is found by a singular value decomposition, we have to compute all the elements v_j, $1 \leqslant j \leqslant N$, and then to leave out the last $N - n$ terms, the ones for which $\sigma_j^2 < \tau$, with τ a prescribed tolerance.

Here we present a way to slightly modify our basis in order to compute only its most significant part. The connection between the two bases will be discussed in details, but we stress in particular that the new algorithm has some characteristic features coming from its dependence on the data.

In what follows we will often omit the dependence on the weights $\{w_i\}_{i=1}^N$, which is equivalent to take all weights $w_i = |\Omega|/N$.

This method makes use of some tools from the theory of Krylov subspaces, recalled in the next Section.

4.2.1   The Lanczos method and the approximation of the SVD

We recall here some basic facts about the Lanczos method, as they will be used in the following. Further details can be retrieved in [32].

Given the set of points $X_N \subset \Omega$, the goal is to find the solution of the linear system $A\,c = b$ with A the kernel matrix and $b_i = f(x_i)$, $1 \le i \le N$ the vector of the evaluations of a function $f \in \mathcal{H}(\Omega)$. To this aim, for $n \le N$ let $\mathcal{K}_n(A, b) = \mathrm{span}\{b, Ab, \ldots, A^{n-1}b\}$ be the *Krylov subspace* of order n generated by A and b. The Lanczos method computes an orthonormal basis $\{p_1, \ldots, p_n\}$ of $\mathcal{K}_n(A, b)$ through a Gram-Schmidt orthonormalization. Namely, if we denote by $P_n \in \mathbb{R}^{N \times n}$ the matrix having the vectors p_i as columns, we can write the algorithm in matrix form as

$$AP_n = P_{n+1}\bar{H}_n, \ \bar{H}_n = \left[\begin{array}{c} H_n \\ \bar{h}e_n^T \end{array} \right], \tag{4.1}$$

where H_n is a $n \times n$ tridiagonal matrix (because of the symmetry of the kernel), $\bar{h} \in \mathbb{R}$ and $e_n \in \mathbb{R}^n$ is the n-th unit vector.

Once these matrices have been computed, the solution c can be approximated as $c = P_n\,y$, where $y \in \mathbb{R}^n$ is such that $\bar{H}_n y = \|b\|_2 e_1$. If A has a good low-rank approximation, we expect that a good approximation of x can be computed using n components with $n \ll N$.

This decomposition can be also used for other purposes. In particular, we want to relate the singular values of A and those of $\bar{H}_n$. One can start by considering a SVD decomposition $\bar{H}_n = U_n \bar{\Sigma}_n V_n^T$, where $U_n \in \mathbb{R}^{(n+1) \times (n+1)}$ and $V_n \in \mathbb{R}^{n \times n}$ are unitary matrices and

$$\bar{\Sigma}_n = \left[\begin{array}{c} \Sigma_n^2 \\ 0 \end{array} \right],$$

Σ_n^2 being the diagonal matrix having the singular values as its entries, and then use this decomposition to approximate a SVD A as $\left(P_{n+1}U_n, \bar{\Sigma}_n, P_n V_n\right)$. A detailed analysis of the relation between this approximation and the true singular value decomposition of A can be found in [43]. We will use this approximated decomposition to construct an approximated version of the WSVD basis.

For notational purposes, we observe that, since the last row of $\bar{\Sigma}_n$ is the zero vector, the decomposition does not change if we remove this row and the last column of U_n. To simplify our notation, from now onward we will denote by U_n the matrix without the last column, so that the decomposition becomes $\bar{H}_n = U_n \Sigma_n^2 V_n^T$. Moreover, the j-th diagonal element of Σ_n^2 will be denoted by σ_j^2 instead of $\sigma_{j,n}^2$ when no confusion arises.

4.2.2   Construction of the basis

By means of the Lanczos algorithm, we can construct a new set of functions $\{\bar{v}_j\}_{j=1}^n \in \mathcal{H}(\Omega)$ that are intended as an approximation of the WSVD basis.

The set will be orthogonal with respect to the $\ell_{2,w}(X_N)$-inner product and near-orthonormal in $\mathcal{H}(\Omega)$, in a way that we will specify later. Moreover, this set can be computed iteratively, i.e., we can choose to compute just the first n elements without computing the other $N-n$ basis. We stress that this new basis does not span $V(X_N)$ (unlike $n = N$, when the SVD of $\bar{H}_n$ equals to that of A). In this sense the term "basis" is not the most appropriate to identify these functions, even if we will use it with an abuse of language with the aim of avoiding a more complicated statement.

Using again the notation of Theorem 3.1, the basis can be defined as follows. Observe that also here we will use an imprecise notation. Indeed, the matrix C that defines the new basis is not invertible for $n < N$, since it maps the N-terms standard basis into the n-terms approximated WSVD basis. Nevertheless, the role of C is the same as in the case of a true change of basis.

Definition 4.1. *Let $n \leq N$ and let $\bar{H}_n$, P_n, V_n, U_n, Σ_n^2 be as introduced above. The approximated WSVD basis $\{\bar{v}_j\}_{j=1}^n$ is characterized by the matrix*

$$C_n = P_n V_n \Sigma_n^{-1}$$

or by the collocation matrix

$$V_n = P_{n+1} U_n \Sigma_n.$$

We remark that in this case the basis strongly depends on the particular function, say $\bar{f} \in \mathcal{H}(\Omega)$, used to construct the Krylov subspace. This dependence influences the behavior of the approximant, as it will be more clear from the following properties of the basis.

4.2.3 Properties

With the approximation introduced by the use of the Lanczos algorithm to compute an SVD of A, the properties of the WSVD basis are slightly modified. Before stating them in a precise way, we prove this technical Lemma.

Lemma 4.2. *Let $n \leq N$ and let $\tilde{U}_n$ be the square matrix obtained from U_n removing the last row u_n^T. Then $\tilde{U}_n$ and V_n coincide except for the last row, namely only the n-th row d_n^T of their difference is a non zero row vector.*

Proof. Using the SVD of $\bar{H}_n$, it is immediate to see that

$$
\left[
\begin{array}{c|c}
H_n^2 & \bar{h} H_n e_n \\
\hline
\bar{h} e_n^T H_m & \bar{h}^2
\end{array}
\right]
= \bar{H}_n \bar{H}_n^T =
\left[
\begin{array}{c|c}
\tilde{U}_n \Sigma_n^4 \tilde{U}_n^T & \tilde{U}_n \Sigma_n^4 u \\
\hline
u^T \Sigma_n^4 \tilde{U}_n & u^T \Sigma_n^4 u
\end{array}
\right],
$$

hence $H_n^2 = \tilde{U}_n \Sigma_n^4 \tilde{U}_n^T$. On the other hand

$$H_n^2 + \bar{h} e_n e_n^T = \bar{H}_n^T \bar{H}_n = V_n \Sigma_n^4 V_n.$$

From these two equalities easily follows

$$\left(V_n - \tilde{U}_n \right) \Sigma_n^4 \left(V_n - \tilde{U}_n \right)^T = \bar{h} e_n e_n^T.$$

We can conclude that each row of the difference matrix, $D_n = V_n - \tilde{U}_n$, is the zero vector except for the last one, say d_n^T, which satisfies $d_n^T \Sigma_n^4 d_n = \bar{h}$ □

We are now ready to prove the following Theorem, where the properties of the approximate WSVD basis are proven.

Theorem 4.3. *The approximated WSVD basis has the following properties:*

 i) *the basis is $\ell_{2,w}(X_N)$- orthogonal with $\|\bar{v}_j\|_{\ell_{2,w}(X_N)}^2 = \sigma_j^2$;*

 ii) *the basis is near-orthonormal on $\mathcal{H}(\Omega)$, meaning that $(\bar{v}_i, \bar{v}_j) = \delta_{ij} + r_{ij}^{(n)}$ where $(R^{(n)})_{ij} := r_{ij}^{(n)}$ is a rank one matrix for $1 \leq n < N$, and $r_{ij}^{(N)} = 0$, $1 \leq i, j \leq N$;*

 iii) *when $n = N$, $\bar{v}_j = v_j$, $1 \leq j \leq N$.*

Proof. It suffices to compute the Gramian matrix of the basis with respect to the $\ell_{2,w}(X_N)$ and $\mathcal{H}(\Omega)$ inner products, say $G(\bar{V}_n)_{\ell_{2,w}(X_N)}$ and $G(\bar{V}_n)_{\mathcal{H}}$, respectively. Using the formulas for the Gramians as in Theorem 3.7 we get

$$G(\bar{V}_n)_{\ell_{2,w}(X_N)} = C_n^T A W A C_n = V_n^T W V_n = \Sigma_n U_n^T P_{n+1}^T P_{n+1} U_n \Sigma_n = \Sigma_n^2,$$

and

$$\begin{aligned} G(\bar{V}_n)_{\mathcal{H}} =& C_n^T A C_n = \Sigma_n^{-1} V_n^T P_n^T P_{n+1} U_n \Sigma_n \\ =& \Sigma_n^{-1} V_n^T \left[I_n \mid 0 \right] U_n \Sigma_n = \Sigma_n^{-1} V_n^T \tilde{U}_n \Sigma_n. \end{aligned}$$

From the above Lemma 4.2, $\tilde{U}_n = V_n + e_n d_n^T$, hence if v_n^T is the last row of V_n,

$$V_n^T \tilde{U}_n = V_n^T \left(V_n + e_n d_n^T \right) = I_n + v_n d_n^T.$$

This allows to conclude that

$$G(\bar{V}_n)_{\mathcal{H}} = \Sigma_n^{-1} \left(I_n + v_n d_n^T \right) \Sigma_n = I_n + R_n$$

with $R_n = \Sigma_n^{-1} v_n d_n^T \Sigma_n$.

The last statement easily follows from the fact that the SVD of A is exactly computed by the Lanczos method when $n = N$.   □

Note that $r_{ij} = (R_n)_{ij}$ is in general non vanishing if $n < N$, but there is strong numerical evidence that the magnitude of r_{ij} is close to the machine precision except if both i and j are close to n. That is to say that the first elements of the basis are $\mathcal{H}(\Omega)$-orthonormal from a numerical point of view.

4.2.4 Approximation

It comes now easy to compute the approximant obtained by the approximated WSVD basis. If we consider the function $\bar{f} \in \mathcal{H}(\Omega)$ from which the basis is constructed, we get the approximant as a projection with respect to the $\ell_{2,w}(X_N)$ inner product. By denoting also in this case as $\Lambda_n \bar{f}$ the $\ell_{2,w}(X_N)$-projection of $\bar{f}$ on the subspace spanned by $\{\bar{v}_j\}_{j=1}^n$, we get

$$\Lambda_n \bar{f} = \sum_{j=1}^{n} \sigma_j^{-2} \left(\bar{f}, \bar{v}_j \right)_{\ell_{2,w}(X_N)} \bar{v}_j \tag{4.2}$$

since $\Sigma_m^{-2} V_n^T$ is a left inverse of V_n.

What is a bit surprising is that $\Lambda_n \bar{f}$ can be expressed also in this case in terms of the $\mathcal{H}(\Omega)$ inner product, as in the case of the true WSVD basis.

Theorem 4.4. *If the basis $\{\bar{v}_j\}_{j=1}^n$ is constructed from $\bar{f} \in \mathcal{H}(\Omega)$, for $1 \leq j \leq N$ we have*

$$\left(\bar{f}, \bar{v}_j \right)_{\ell_{2,w}(X_N)} = \sigma_j^2 \left(\bar{f}, \bar{v}_j \right). \tag{4.3}$$

In particular, the approximant Λ_n is a $\mathcal{H}(\Omega)$-projection, or, equivalently,

$$\Lambda_n(\bar{f}) = \sum_{j=1}^{n} \left(\bar{f}, \bar{v}_j \right) \bar{v}_j. \tag{4.4}$$

Proof. Observe that the coefficient $(f, \bar{v}_j)_{\ell_{2,w}(X_N)}$, for a general $f \in \mathcal{H}(\Omega)$, is the j-th column of $f_{X_N}^T V_n$, where f_{X_N} is the column vector of the evaluations of f at X_N. Using again Lemma 4.2 and denoting by u_n^T the last row of U_n, we get

$$V_n = P_n \tilde{U}_n \Sigma_n + p_{n+1} u_n^T \Sigma_n = P_n V_n \Sigma_n + p_n d_n^T \Sigma_n + p_{n+1} u_n^T \Sigma_n$$
$$= C_n \Sigma_n^2 + p_n d_n^T \Sigma_n + p_{n+1} u_n^T \Sigma_n.$$

If we take $f = \bar{f}$, since $\bar{f}_{X_N}$ is orthogonal to the vectors $\{p_2, \ldots, p_{n+1}\}$ by construction, we have $\bar{f}_{X_N}^T V_n = \bar{f}_{X_N}^T C_n \Sigma_n^2$, i.e., equality (4.3) holds. This implies in particular that the formula (4.4) is equivalent to (4.2). $\square$

We point out that if we take another function $f \in \mathcal{H}(\Omega)$, the equality (4.3) holds with a residual term on the right hand side. This is due to the fact that in the case $f \neq \bar{f}$ the terms depending on $p_2, \ldots, p_{n+1}$ are not deleted. On the other hand, equation (4.2) depends only on the relation between the left inverse of V_n and its transpose, not on the connection with C_n. This means that we can compute the approximant of a function $f \neq \bar{f}$ also using the basis constructed starting from $\bar{f}$. Although possible, this seems not an interesting approach, since the construction of the basis is fast and we can obviously expect a better result for a data-dependent basis.

To conclude, if the WSVD basis is replaced with the new one, the estimates of Corollary 3.11 still hold, with v_j replaced by $\bar{v}_j$. In the next Section we will show some numerical computations the quantities involved in those estimates.

4.2.5   Experiments

In this Section we show some examples in order to test the features of our approximation scheme. We show two examples. In the first we approximate a function from the native space of a Gaussian kernel with the aim of understanding the behavior of the method in a known setting. In the second example we approximate real data by an inverse multiquadric kernel.

All experiments have been performed with the Matlab code [49] on a Intel Celeron Dual-Core CPU T3100, 1.90GHz, with 3 Gb of RAM. The errors are measured by the Root Mean Squared Error (RMSE), i.e., by a weighted ℓ_2-norm computed on a grid of equally spaced points.

For the first experiment we consider the Gaussian kernel

$$K(x,y) = e^{-\varepsilon^2\,\|x-y\|_2^2} \;\; \forall x,y \in \Omega_1,$$

where Ω_1 is the unit disk in $\mathbb{R}^2$, and the shape parameter has the fixed value $\varepsilon = 1$. The function f_1 we want to approximate is defined for all $x \in \mathbb{R}^2$ as

$$f_1(x) = K(x,p_1) + 2\,K(x,p_2) - 2\,K(x,p_3) + 3\,K(x,p_4),$$
$$p_1 = (0,-1.2),\; p_2 = (-0.4,0.5),\; p_3 = (-0.4,1.1),\; p_4 = (1.2,1.3),$$

which is clearly a function in the native space. The set X_N is generated by the restriction to Ω_1 of a grid of equally spaced points in $[-1,1]^2$, with the grid size chosen so that the number of points which are inside Ω_1 is in fact N.

The first problem is to choose how to stop the Lanczos iteration. For sure this is a key point of this approximation and only few sophisticated stopping criteria are known. On the other hand, we would like to use a stopping rule which tells us something on the approximation process from a functional point of view. A reasonably good choice for stopping the Lanczos iteration is when, for a certain tolerance $\tau > 0$, we have

$$\left| \frac{1}{N} \sum_{j=1}^{n} (H_n)_{jj} - 1 \right| < \tau, \tag{4.5}$$

which is motivated by the Property (v) of Theorem 3.7. This is a rough criterion, but it seems good enough to control the iterations in the case of functions lying in the native space of the kernel. This behavior is shown in Figure 4.7 in the case of $\tau = 10^{-15}$. Observe that, in this example, the control on the residual (4.5) is in fact effective. Indeed, the RMSE decreases as

long as the residual decreases, while, shortly after the residual reaches the level τ, the RMSE rapidly increases. This is due to the numerical instability that emerges in the Lanczos algorithm if it is not stopped properly.
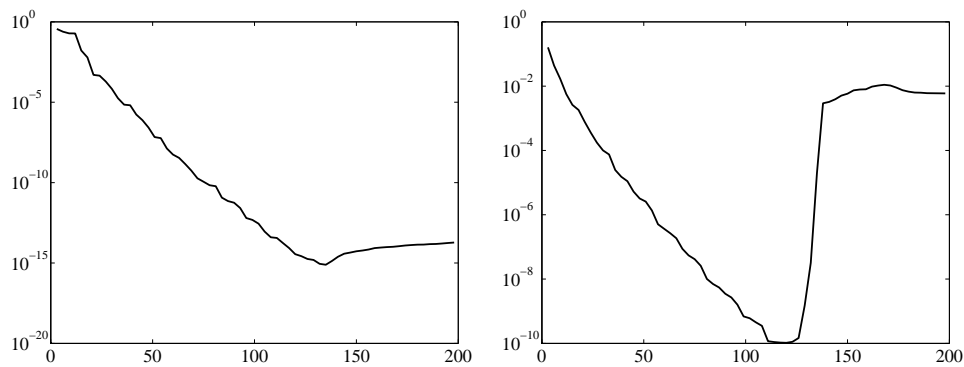

Figure 4.7: Decay of the residual described in (4.5) compared with the corresponding RMSE for the first example of this Section.

In Figure 4.8 we show the plots of the basis elements $\bar{v}_1$, $\bar{v}_{11}$, $\bar{v}_{21}$, $\bar{v}_{31}$. Notice, as we expect by an approximation of the eigenbasis, that the number of oscillations of the sign of the basis increases, while its maximum value decreases.
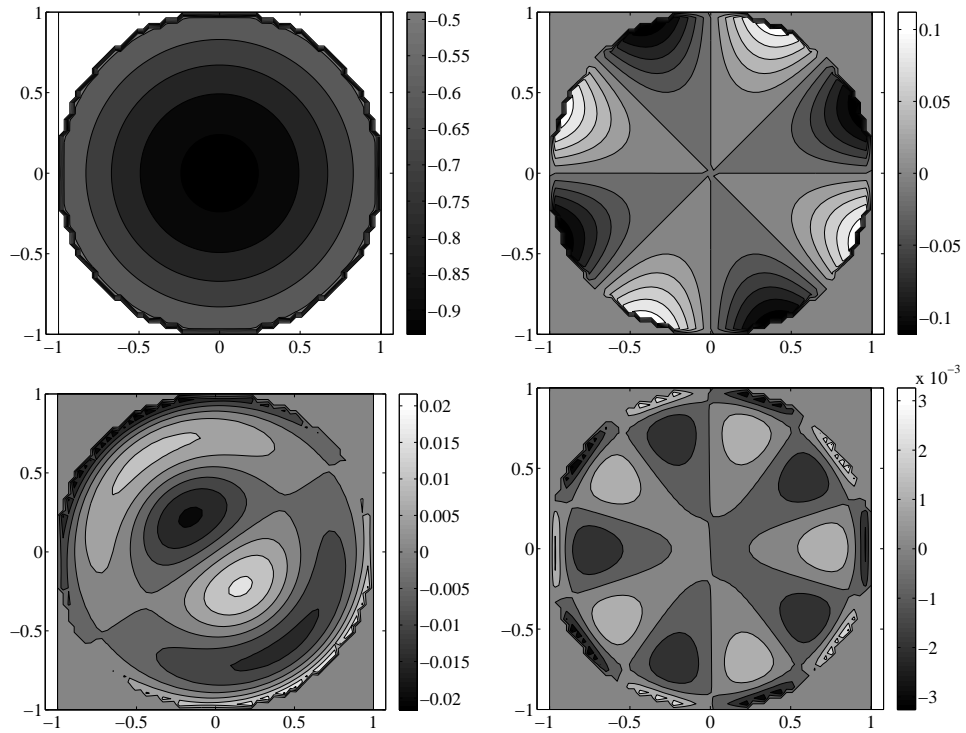
Figure 4.8: Basis functions computed in the first example of Section 4.2.5. From top left to bottom right $\bar{v}_1$, $\bar{v}_{11}$, $\bar{v}_{21}$, $\bar{v}_{31}$.

To analyze the computational time required to solve the approximation problem, we compared it with that required by the WSVD basis. The sets X_N of equally spaced points in Ω_1 are generated to have at least cardinality $N = 15^2, 23^2, 31^2, 39^2$ (since they are generated by restriction to Ω_1 of a grid in $[-1, 1]^2$, the actual values are $N = 253, 529, 1009, 1576$). The use of a quite small number of data is due to the slowness of the computation of the WSVD basis. For each N we compute the optimal n using the present algorithm with tolerance $\tau = 10^{-14}$. Then we use this value to select the corresponding number of WSVD basis elements to be used. In Table 4.1 we display this comparison, only for some values of N, in order to show the performance of the method. It is worth to mention that the computational times of Table 4.1 are those required by the truncated SVD, which are, in this case, computed from a full SVD. We can also see that for all choices of N the method takes less than 7 seconds to construct the approximant. Moreover, despite the approximation introduced by the new truncated basis, the computed errors are comparable to or even better than the one of the original basis. This behavior is due to the function-dependent construction of the basis.

Using these observations we can construct the approximation on the full data set. In this case, choosing $\tau = 1 \cdot 10^{-14}$ we obtain $n = 115$ basis el-

N	253	529	1009	1576
n	110	114	115	116
RMSE new	$3.4 \cdot 10^{-10}$	$6.7 \cdot 10^{-11}$	$5.5 \cdot 10^{-11}$	$3.4 \cdot 10^{-11}$
RMSE WSVD	$3.3 \cdot 10^{-9}$	$1.1 \cdot 10^{-9}$	$8.3 \cdot 10^{-10}$	$7.9 \cdot 10^{-10}$
Time new	$3.4 \cdot 10^{-1}$	$1.0 \cdot 10^{0}$	$2.6 \cdot 10^{0}$	$6.5 \cdot 10^{0}$
Time WSVD	$7.2 \cdot 10^{-1}$	$4.2 \cdot 10^{0}$	$2.5 \cdot 10^{1}$	$1.1 \cdot 10^{2}$

Table 4.1: Comparison of the WSVD basis and the new basis. Computational time in seconds and corresponding RMSE for the first example in Section 4.2.5, with roughly $N = 15^2$, 23^2, 31^2, 39^2 equally spaced points.

ements out of $3600 = 60^2$ points with RMSE $= 1.03 \cdot 10^{-10}$. The overall computation takes 45 seconds, with the first 40 seconds used to construct the basis, and the remaining 5 seconds required to compute the approximant.

The approximated Power Function computed through the approximated WSVD basis for this experiment is depicted in Figure 4.9. It is clear that this quantity is quite uniform in the domain Ω_1. Moreover, the maximum value of the approximated Power Function on the grid is of order 10^{-13}.
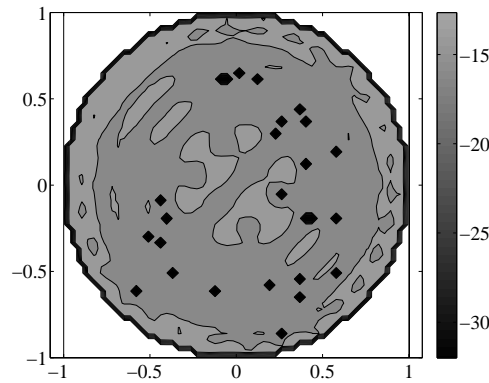

Figure 4.9: Logarithmic plot of the approximated Power Function associated to the first example of this Section

To conclude this experiment, we test the effect of the shape parameter ε. In particular, we use the example of this section, with roughly $N = 25^2$ (indeed, $N = 648$), to show how the number n of selected bases, and thus the RMSE, changes as $\varepsilon \to 0$. First, we show in Table 4.2 the results obtained by approximating the function f_1. It is clear that the number of selected bases decreases with ε, affecting also the corresponding RMS error.

By forcing instead the method to select a bigger number of bases, hence by prescribing a smaller tolerance, the error rapidly increases in the same way as presented in Figure 4.7 and this phenomenon is clearly affected by ϵ. Second, Table 4.3 presents the results obtained in the same setting, but modifying f_1 accordingly to the shape parameter, so that $f_1 \in \mathcal{H}(\Omega)$ for each choice of ε. The reduction of the number of selected basis elements as $\epsilon \to 0$ is analogous to the previous one but, in contrast, the fact that the test function is in the native space of the kernel allows the method to reach an error close to the machine precision.

ε	2	2^{-2}	2^{-4}	2^{-6}	2^{-7}
RMSE	$1.0 \cdot 10^{-8}$	$2.1 \cdot 10^{-6}$	$1.7 \cdot 10^{-3}$	$6.5 \cdot 10^{-2}$	$3.75 \cdot 10^{-1}$
n	249	64	26	16	12

Table 4.2: Number n of selected bases and the corresponding RMSE by decreasing ε for roughly $N = 25^2$ equally spaced points and fixed test function.

ε	2	2^{-2}	2^{-4}	2^{-6}	2^{-7}
RMSE	$4.9 \cdot 10^{-8}$	$1.4 \cdot 10^{-12}$	$9.9 \cdot 10^{-14}$	$1.6 \cdot 10^{-14}$	$8.7 \cdot 10^{-14}$
n	272	39	20	11	11

Table 4.3: Number n of selected bases and the corresponding RMSE by decreasing ε for roughly $N = 25^2$ equally spaced points and ε-dependent test functions.

The second experiment uses the inverse multiquadric kernel (IMQ)

$$K(x,y) = \left(\sqrt{1 + \varepsilon^2 \|x - y\|^2} \right)^{-1} \quad \forall x, y \in \Omega_2,$$

where $\varepsilon = 2$ and Ω_2 is a *lune*, namely, the set in $\mathbb{R}^2$ defined by the difference of two disks of radius 0.5 with centers in $(0,0)$ and $(0.5, 0.5)$.

We approximate the exponential Franke's test function (see [29]), say f_2, plotted in Figure 4.10, top left. To make the test as general as possible, for X_N we use a set of randomly distributed data sites with $N = 60^2$.

The results of the test are depicted in Figure 4.10. The method used $n = 176$ basis elements, giving $RMSE = 1.38 \cdot 10^{-6}$, and a pointwise error as in Figure 4.10 (bottom right). The whole computation required one minute of CPU.

In this case the tolerance of the stopping rule was set to $\tau = 10^{-10}$, since a smaller value led to an increase of the RMS error. The weakness of the stopping rule in this case is probably due to the fact that the function f_2 does not belong to the native space of the kernel.
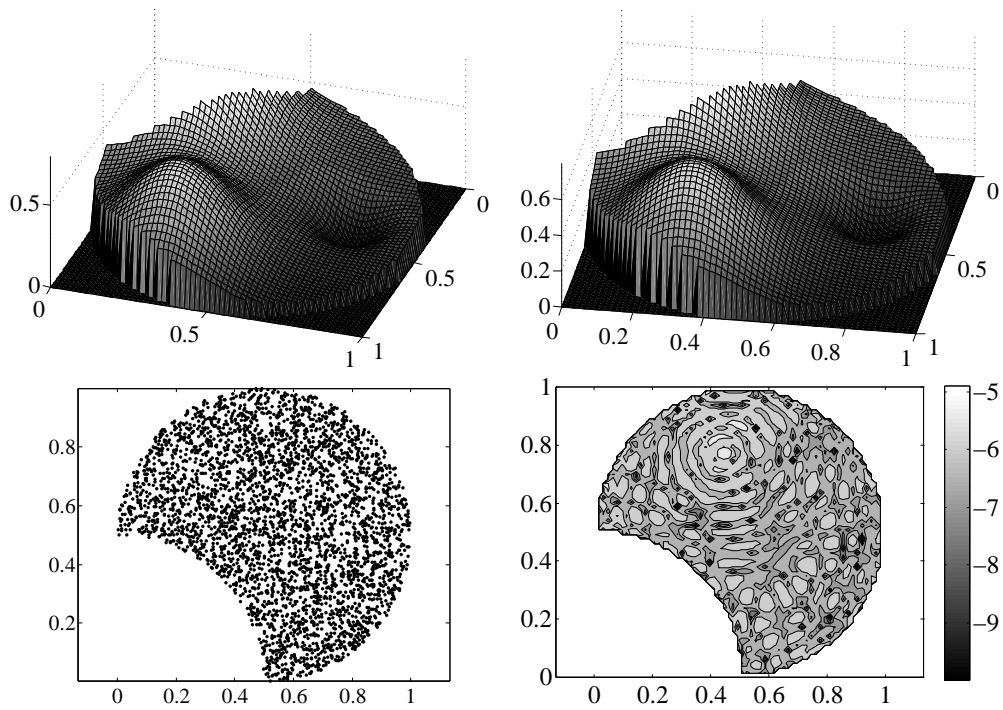
Figure 4.10: Exponential Franke's function (top left); approximation obtained with a IMQ kernel with $\varepsilon = 2$ (top right) using 60^2 randomly distributed data points (bottom left); pointwise error computed on a grid of 60×60 equally spaced points (bottom right, logarithmic scale).

Chapter 5

Applications

In this Chapter we present two applications of kernel-based methods to the solution of two different problems.

The first Section considers the solution of large interpolation problems, i.e., problems where the number N of points and function samples can be large. To face this kind of application, we couple the algorithms of Sections 3.3 and 4.2 with a domain decomposition technique. In this way we can preserve the stabilization properties of the mentioned method, with an enhanced computational efficiency. The method presented here is a starting point to test this coupling, where we use a standard implementation of the domain decomposition technique, based on *kd-trees* (see e.g. [3]). It is intended as the basis for a more efficient algorithm which will be published in [8], which uses instead the algorithms of [10, 11], but is perfectly equivalent to the present one from the side of the theory.

The second Section is devoted to the reconstruction of medical images from *CT scans* (see e.g. [25]). We will develop a method that employs kernel-based approximation for the solution of this reconstruction problem, and that allows to implement in the image recovery some techniques related to the extraction of optimal subspaces in $\mathcal{H}(\Omega)$. We present here only a preliminary analysis of this method, with application only to the Gaussian kernel. A detailed theoretical foundation for more general kernels will be published in the paper [14].

The content of Section 5.1 is published in [9], and it is partially taken from there. Moreover, it is possible to download a Matlab implementation of the method of Section 5.2 from [50].

5.1 Partition of unity with local stabilization

We present here an algorithm to approximate large datasets by kernel techniques. The method couples a fast domain decomposition procedure with a localized stabilization method. We will see in Section 5.1.1 how to deal

71

with large data sets by splitting the problem in smaller subproblems, and in Section 5.1.2 how to locally employ the change of basis to control the instability. We will work here with RBF kernels $K(x, y) = \Phi(\varepsilon \|x - y\|)$.

5.1.1   Background facts on the PUM

We want to avoid to solve a global problem involving a large and possibly dense matrix. To do so, we focus on the partition of unity (PU) method (see [59]). The idea is to decompose the problem, solve (many) small local approximation subproblems, and then blend together the results in a global approximant. We give here a general review of this method, while further implementation details are provided in Section 5.1.3.

We work with a set X_N of N points in the bounded set $\Omega \subset \mathbb{R}^d$ and with a function $f \in \mathcal{H}(\Omega)$. For a fixed natural number L we consider an open and finite covering $\{\Omega^{(k)}\}_{k=1}^L$ of the bounded set Ω, that is

$$\Omega \subset \bigcup_{k=1}^{L} \Omega^{(k)},$$

where we require mild overlapping of the patches. Associated to the covering one considers a partition of unity, i.e., a set of nonnegative and continuous real valued functions $\{P^{(k)}\}_{k=1}^L$ with

$$\mathrm{supp}(P^{(k)}) \subset \Omega^{(k)}, \quad \sum_{k=1}^{L} P^{(k)}(x) = 1 \text{ for all } x \in \Omega.$$

Since the kernel K is still positive definite when restricted to each $\Omega^{(k)}$, it is possible to uniquely solve the interpolation problem restricted to the each local domain. Namely, for each $\Omega^{(k)}$, $1 \le k \le L$, we consider the restricted set of $N^{(k)}$ data locations defined as $X_N^{(k)} = X_N \cap \Omega^{(k)}$, $N^{(k)} = \mathrm{card}(X_N^{(k)})$, and the corresponding values of f on $X_N^{(k)}$. The local approximant $s_{X_N^{(k)}}^{(k)} f$ is then computed as

$$s_{X_N^{(k)}}^{(k)} f = \sum_{i=1}^{N^{(k)}} c_i^{(k)} K(\cdot, x_i^{(k)}),$$

where now it suffices to solve a much smaller linear system to determine the coefficients in the above expansion. To simplify the notation we drop the subscript $X_N^{(k)}$ from $s_{X_N^{(k)}}^{(k)} f$, which will be denoted as $s^{(k)} f$.

We can now blend together the local approximants by means of the partition of unity $\{P^{(k)}\}_{k=1}^L$, i.e., we recover a global approximant sf on Ω as

$$sf(x) = \sum_{k=1}^{L} P^{(k)}(x) s^{(k)} f(x), \ x \in \Omega.$$

The latter satisfies the global interpolation conditions since $\sum_{k=1}^{L} P^{(k)} = 1$ on Ω.

It is important to recall that, under further assumptions on the covering and on the partition of unity, it is possible to prove that the global approximant keeps the same approximation order of the local ones (see [59]), so the method is an efficient and accurate way of constructing a kernel-based interpolant, even if different from the standard one.

We remark also that, besides the theoretical framework, the success of the PU method relies on a good point searching technique. Nevertheless, we stress that a change in the point-searching technique does not affect the present theoretical analysis. In this view, the numerical experiments presented in Section 5.1.3 should be indicative, as regards the error, of the behavior of the enhanced method of [8].

5.1.2 The algorithm

The partition of unity method allows to deal with large datasets in an efficient way, and partially reduces the ill-conditioning found in many cases when dealing with RBF problems, since at least it reduces the size of matrices involved. Nevertheless, using a domain decomposition technique can be not enough. To this end, we want to employ locally the algorithm of Section 4.2.

Before using the WSVD method, observe that for any subdomain $\Omega^{(k)}$ there is an associated space of functions $\mathcal{H}(\Omega^{(k)})$ given by the restriction of $\mathcal{H}(\Omega)$ to $\Omega^{(k)}$. As in the global setting, the interpolation operator is the projection into the finite subspace $V(X_N^{(k)})$ spanned by the kernel translates on $X_N^{(k)}$. To this subspace we apply the change of basis to obtain the WSVD basis $\{v_j^{(k)}\}_{j=1}^{N^{(k)}}$, where for any k we compute the local coefficient matrix $C^{(k)}$, coming from the local kernel matrix $A^{(k)}$.

Through the approximated version of the WSVD basis, we can select a truncation index $n^{(k)} \leq N^{(k)}$, and build up the local least-squares approximant (or truncated interpolant) that will be used to construct the global approximant through the technique of the previous Section.

Observe that this algorithm has some further advantage over the global (approximated) WSVD method. Namely, in the case of severe ill-conditioning the truncation approach may be over regularizing. Indeed, it has been observed that the WSVD method, when applied globally, is capable to deal with only a limited instability. In fact, when $\varepsilon \to 0$ or when the kernel is too smooth, the original algorithm needs to leave out too much elements of the basis, and the resulting approximant may be meaningless. Instead, when repeatedly applied to the small subdomains $\Omega^{(k)}$, the regularizing effect simply improves the approximation, as shown in the next Section.

5.1.3 Experiments

We present here two experiments to test our algorithm on two illustrative problems in dimension $d = 2$. For both examples we compare the behavior of two different RBF kernels with different smoothness, namely the Gaussian (which is a $\mathcal{C}^\infty$ kernel) and the $\mathcal{C}^2$ Matérn kernel. In both examples the test function is the bivariate exponential Franke's test function, approximated on the unit square $\Omega = [0, 1]^2$. The shape parameter ε varies on a log-spaced grid in $[10^{-4}, 10^2]$. The error is measured by means of the root mean squared error (RMSE) on a equally spaced grid of 40^d points in Ω.

In both examples the PU method uses a covering of Ω based on balls $\{\Omega^{(k)}\}_{k=1}^{L}$, where the centers $\{y^{(k)}\}_{k=1}^{L}$ of the balls lie on an equally spaced grid in Ω and the radii are the same for any $k = 1, \ldots, L$. The points are organized through a *kd-tree*, as in [7]. We use N Halton points, where $N = 65^d, 129^d, 257^d$, which correspond respectively to $L = 16^d, 32^d, 64^d$. The common radius of the balls is chosen to be $1/\sqrt{L}$, so that the boundary of every patch intersects the centers of the nearby patches.

As a partition of unity we use a superposition of $\mathcal{C}^2$ Wendland's functions, $P^{(k)}(x) = \psi(\nu \| x - y^{(k)} \|)$,

$$\psi(r) = (1 - \nu r)_+^4 (4\nu r + 1),$$

where the shape parameter $\nu > 0$ is scaled to control the support of the partition. We remark that these choices of $\Omega^{(k)}$ and $P^{(k)}$ fulfill the requirements on the covering and the partition in order to preserve globally the order of the local approximation. We use a tolerance $\tau = 10^{-14}$ for the Lanczos algorithm.

The first example compares the approximation obtained with the new method with the one obtained with a global application of the approximated WSVD basis described in Section 4.2. Since we use here a global approach, the experiment is limited to $N = 65^d$. The results are shown in Figure 5.1. For a not too small shape parameter, for both kernels, the global method and the new algorithm behaves in the same way. As $\varepsilon \to 0$ the global method loses accuracy, while the local one is still able to compute an accurate and stable approximant.

In the second example we compare the behavior of the PU method without local regularization with the new algorithm. Figure 5.2 presents the results for this test. As expected, there is a first phase where no instability is present and the two methods behave exactly in the same way. Instead, in the flat limit case the standard PU method starts to become unstable, while the stabilized method retains its accuracy and it is able to effectively compute the approximants. The effect is much more evident for the Gaussian (which is smoother) than for the Matérn kernel.
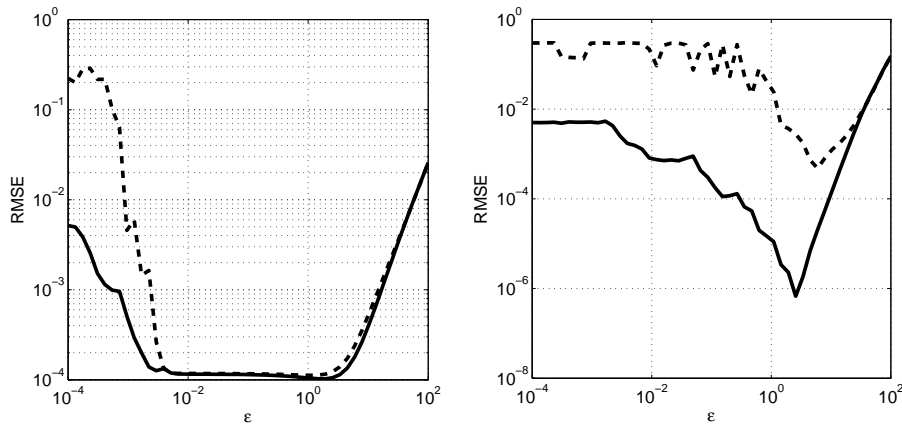
Figure 5.1: Comparison between the global approximant computed with the approximated WSVD basis (dotted lines) and the new method (solid lines), for the $\mathcal{C}^2$ Matérn kernel (left) and the Gaussian kernel (right).
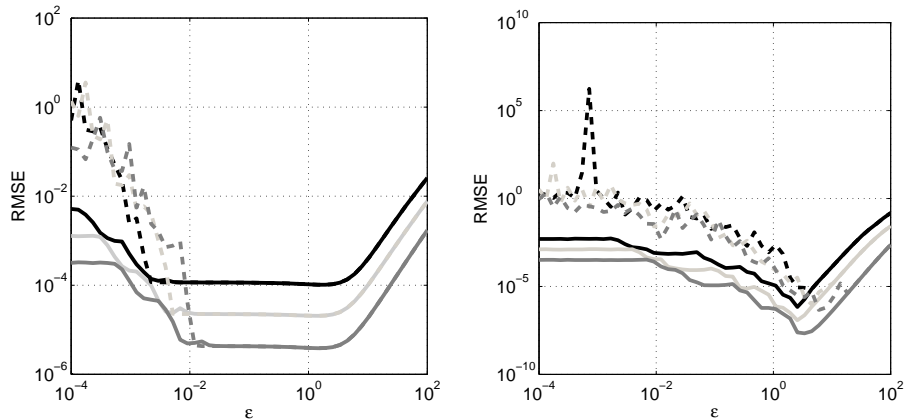

Figure 5.2: Comparison between the standard PU approximant (dotted lines) and the new method (solid lines), for the $\mathcal{C}^2$ Matérn kernel (left) and the Gaussian kernel (right).

5.2 Reconstruction of medical images from CT scans

This Section is devoted to present an application of kernel-based generalized interpolation to the problem of medical image reconstruction from CT scans (see [25]). From a mathematical point of view, this problem can be translated into the problem of recovering a function, which represents the unknown image, from the evaluation of a linear transform on certain points. This operator, which is called the Radon transform, represents the measurements collected by the medical machine. In our context, this problem can be interpreted as a standard generalized interpolation problem,

where the data are provided by the Radon transform itself. The use of kernel methods in this context has been proposed in [15], where radial kernels are used, and we present here a partial extension of this work.

Although different methods are known and used to solve this problem, the use of kernel based methods presents some advantage. Namely, while in standard algorithm one needs to consider a fixed geometry of the data (called the *parallel beam geometry*), this new approach allows to recover images from unorganized samples. This fact can be a starting point in the research of optimal geometries, which hopefully require less scans to recover a medical image. The content of this Section is a preliminary analysis of this application, and so far applies only to the Gaussian kernel.

5.2.1 Background facts on the Radon transform

Our goal is the recovery of a medical image, which is represented by an unknown function $f : \mathbb{R}^2 \to \mathbb{R}$ with compact support $\Omega \subset \mathbb{R}^2$. The compact set is usually assumed to be $\Omega = [0, 1]^2$, but this hypothesis is not restrictive.

To recover f, we will consider data coming from its Radon transform, which is defined, for any $t \in \mathbb{R}$, $\theta \in [0, \pi)$, as the line integral of f through the line

$$x_{t,\theta}(s) = (t \cos(\theta) - s \sin(\theta), t \sin(\theta) + s \cos(\theta)), \;\; s \in \mathbb{R}.$$

In other words, we have the following definition.

Definition 5.1. *For a function $f : \mathbb{R}^2 \to \mathbb{R}$ of compact support $\Omega \subset \mathbb{R}^2$, and $f \in L_1(\Omega)$, the Radon transform $\mathcal{R}(f)(t, \theta)$, $(t, \theta) \in \mathbb{R} \times [0, \pi)$ can be computed as*

$$\mathcal{R}(f)(t, \theta) = \int_{-\infty}^{\infty} f(x_{t,\theta}(s)) ds,$$

where

$$x_{t,\theta}(s) = (t \cos(\theta) - s \sin(\theta), t \sin(\theta) + s \cos(\theta)), \;\; s \in \mathbb{R}.$$

It is known that the Radon transform possesses an inverse transform, called the *filtered back projection*. In other words, the original function f can be exactly recovered if one knows the full set of samples $\{\mathcal{R}(f)(t, \theta)\}$ for $(t, \theta) \in \mathbb{R} \times [0, \pi)$.

In real-world applications, instead, one collects a set of discrete data $\{(\mathcal{R}f)(t_i, \theta_j)\}$, $1 \le i \le M$, $1 \le j \le N$, where $\{t_i\}_{i=1}^M \in \mathbb{R}$ and $\{\theta_j\}_{j=1}^N \in [0, \pi)$ are a set of discretization points. By the compactness of the support of f, we have $\mathcal{R}(f)(t, \theta) \neq 0$ only for some t, and we are not interested in collecting a set of vanishing samples. In particular, when $\Omega = [0, 1]^2$, it is customary to assume $\{t_i\}_{i=1}^M \in [-\sqrt{2}, \sqrt{2}]$.

The numerical implementation of the filtered back-projection leads to the so called *Fourier methods*, which are very effective but require the use of

samples organized in a grid of equally spaced samples. We will consider instead fully general samples $\{(t_i, \theta_i)\}_{i=1}^N \subset [-\sqrt{2}, \sqrt{2}] \times [0, \pi)$, which we denote again as X_N, where $x_i = (t_i, \theta_i)_i$.

5.2.2 Kernel-based reconstruction

The symmetric collocation method of Section 1.5 applied to this problem has been analyzed in [15]. We recall here this approach.

We look for a function $g \in \mathcal{H}(\Omega)$,

$$g = \sum_{j=1}^N c_j \mathcal{R}_j^y(K(\cdot, y)), \tag{5.1}$$

where $\mathcal{R}_j = \delta_{x_j} \circ \mathcal{R}$, that meets the interpolation conditions

$$\mathcal{R}_i(g) = \sum_{j=1}^N c_j \mathcal{R}_i^x \mathcal{R}_j^y(K(x, y)) = \mathcal{R}_i(f). \tag{5.2}$$

The coefficients $\{c_j\}_{j=1}^N$ in the above expansion are determined by the solution of the linear system involving the matrix

$$(A_{\mathcal{R}})_{ij} = \mathcal{R}_i^x \mathcal{R}_j^y(K(x, y)), \quad 1 \le i, j \le N.$$

Unfortunately, a direct application of this method leads to a singular problem.

Proposition 5.2. *For each radial kernel K, the matrix $A_{\mathcal{R}}$ is singular along the diagonal.*

The Authors propose to overcome the problem by using a kernel dependent weight function $w : \Omega \to \mathbb{R}$, used as a regularization term in the second Radon transform. By means of w, they define a regularized Radon transform

$$\mathcal{R}_w(f)(t, \theta) = \mathcal{R}(fw)(t, \theta) = \int_{\mathbb{R}} f(x_{t,\theta}(s)) w(x_{t,\theta}(s)) ds,$$

and consider the problem where the interpolation conditions imposed as

$$(\mathcal{R}_w)_i(g) = \sum_{j=1}^n c_j (\mathcal{R}_w)_i^x \mathcal{R}_j^y(K(x, y)) = \mathcal{R}_i(f).$$

Moreover, they prove that this reconstruction problem is well posed for a proper choice of w.

In the case of the Gaussian kernel, in particular, the idea works as follows. The weight they considered is in the form

$$w_\nu(x) = e^{-\nu\|x\|_2^2}, \quad x \in \mathbb{R}^2,$$

for a shape parameter $\nu > 0$. The entries of the regularized reconstruction matrix are given by

$$(\mathcal{R}_w)_i^x \mathcal{R}_j^y(K(x,y)) =$$

$$\frac{\pi}{\varepsilon\sqrt{\varepsilon^2 \sin(\theta_i - \theta_j)^2 + \nu^2}} \exp\left[-\nu^2\left(t_k^2 + \frac{\varepsilon^2(t_j - t_i\cos(\theta_i - \theta_j))^2}{\varepsilon^2 \sin(\theta_i - \theta_j)^2 + \nu^2}\right)\right]. \quad (5.3)$$

Observe that, in this way, the problem becomes well posed, but we lose some properties of the standard generalized interpolation approach. In particular, the kernel matrix is non symmetric, and it is hard to provide error bounds on the reconstruction error.

5.2.3   Symmetric formulation

Here instead we consider a symmetric weighting of the Radon transform, that is, both the transforms involved in the definition of the reconstruction matrix will be regularized by the same weight function. This is equivalent to apply the unweighted Radon transform to a modified version of the kernel K, i.e.,

$$\mathcal{R}_w^x \mathcal{R}_w^y(K(x,y)) = \mathcal{R}^x \mathcal{R}^y(w(x)K(x,y)w(y)).$$

This operation on the kernel is well understood, as the following Theorem shows (see e.g. Corollary 3 in [48, Chapter 2]).

Theorem 5.3. *If $w : \Omega \to \mathbb{R}$ is a non vanishing function,*

$$K_w(x,y) = w(x)w(y)K(x,y)$$

is a kernel on $\Omega \times \Omega$. If f_w is a function in its native space $\mathcal{H}_w(\Omega)$, there exist a function $f \in \mathcal{H}(\Omega)$ such that

$$f_w(x) = w(x)f(x), \quad x \in \Omega.$$

Moreover, for $f_w, g_w \in \mathcal{H}_w(\Omega)$, the inner product $(\cdot, \cdot)_{\mathcal{H}_w}$ of $\mathcal{H}_w(\Omega)$ is defined as

$$(f_w, g_w)_{\mathcal{H}_w} = \left(\frac{f_w}{w}, \frac{g_w}{w}\right).$$

This means that, by a proper modification of a RBF kernel, we can deal with the present reconstruction problem. Observe that we remove the radial symmetry by the kernel, but this is in general not enough to guarantee

that the functionals $\mathcal{R}_i$ are continuous in the native space of the new kernel. A complete analysis of this fact is the object of the paper [14], which is currently in preparation. Here we show only, through direct computations, how to treat the Gaussian kernel.

Before computing the Radon transforms of the modified Gaussian kernel, observe that the last Theorem allows us to understand this symmetric weighted method in the setting of standard kernel-based reconstruction. In particular, the construction of the Newton basis with a greedy selection of $n \leq N$ points is possible here, and we will use it in the following experiments. We use a plain Newton basis here, and not the algorithms developed in this Thesis, since the evaluations of the following formulas for the Radon transforms are computationally expensive, and the construction of our bases slows down too much the computations.

For the Gaussian kernel, we consider the same weight function $w_\nu(x) = \exp(-(\nu\|x\|)^2)$, $\nu > 0$, for a positive parameter $\nu > 0$. The symmetrically scaled kernel is defined by the formula:

$$K_w(x, y) = e^{-(\varepsilon\|x-y\|)^2} e^{-(\nu\|x\|)^2} e^{-(\nu\|y\|)^2}.$$

To compute the Radon transform of K_w, we will use the following lemma.

Lemma 5.4. *Let c_0, c_1, c_2 be constants, and let $c_2 < 0$. We have*

$$\int_{-\infty}^{\infty} e^{c_2 s + c_1 s + c_0} ds = \sqrt{-\frac{\pi}{c_2}}\, e^{c_0 - \frac{c_1^2}{4c_2}}.$$

In the following we will denote as n_θ the unit vector in $\mathbb{R}^2$ defined by the angle $\theta \in [0, \pi)$, i.e., $n_\theta = (\cos(\theta), \sin(\theta))$, and its orthogonal vector as $n_\theta^\perp = (-\sin(\theta), \cos(\theta))$. For any point $x = (x_1, x_2)$ in $\mathbb{R}^2$ we have

$$(x \cdot n_\theta)^2 + (x \cdot n_\theta)^2 = \|x\|^2. \tag{5.4}$$

The first Radon transform can be computed as follows.

Proposition 5.5. *The Radon transform $\mathcal{R}^y[K_{\varepsilon,\nu}(x, \cdot)](t, \theta)$ is given by*

$$\mathcal{R}^y[K_{\varepsilon,\nu}(x, \cdot)](t, \theta) = \sqrt{\frac{\pi}{\varepsilon^2 + \nu^2}}\, e^{\alpha_{\varepsilon,\nu}(x_1, x_2; t, \theta)}$$

where

$$\alpha_{\varepsilon,\nu}(x_1, x_2; t, \theta) = -(\varepsilon^2 + \nu^2)(\|x\|^2 + t^2) + 2\varepsilon^2 t x \cdot n_\theta + \frac{\varepsilon^4}{\varepsilon^2 + \nu^2}(x \cdot n_\theta^\perp)^2 \tag{5.5}$$

or, in a more compact form,

$$\alpha_{\varepsilon,\nu}(x_1, x_2; t, \theta) = -\frac{1}{\varepsilon^2 + \nu^2}\left(\nu^2(\nu^2 + 2\varepsilon^2)\|x\|^2 + \frac{\varepsilon^4}{\varepsilon^2 + \nu^2}\left(\frac{\nu^2 + \varepsilon^2}{\varepsilon^2}t - x \cdot n_\theta\right)^2\right)$$

Proof. We want to compute the integral

$$\int_{-\infty}^{\infty} K_{\varepsilon,\nu}(x, y_{t,\theta}(s))ds.$$

Since $\|y_{t,\theta}(s)\|^2 = t^2 + s^2$, the argument of the exponential in the integrand can be rewritten as follows:

$$-\nu^2(\|x\|^2 + \|y_{t,\theta}(s)\|^2) - \varepsilon^2 \|x - y_{t,\theta}(s)\|^2$$
$$= -(\nu^2 + \varepsilon^2)(\|x\|^2 + \|y_{t,\theta}(s)\|^2) + 2\varepsilon^2 x \cdot y_{t,\theta}(s)$$
$$= -(\nu^2 + \varepsilon^2)(\|x\|^2 + t^2 + s^2) + 2\varepsilon^2(s(-x_1 \sin(\theta) + x_2 \cos(\theta))$$
$$\qquad + t(x_1 \cos(\theta) + x_2 \sin(\theta)))$$
$$= -(\nu^2 + \varepsilon^2)(\|x\|^2 + t^2 + s^2) + 2\varepsilon^2(sx \cdot n_\theta^\perp + tx \cdot n_\theta).$$

To go back to the situation of Lemma 5.4, we define the coefficients

$$c_2 = -(\nu^2 + \varepsilon^2)$$
$$c_1 = 2\varepsilon^2 x \cdot n_\theta^\perp$$
$$c_0 = -(\nu^2 + \varepsilon^2)(\|x\|^2 + t^2) + 2\varepsilon^2 tx \cdot n_\theta$$

and, since $c_2 < 0$, we can conclude that

$$\mathcal{R}^y[K_{\varepsilon,\nu}(x, \cdot)](t, \theta) = \sqrt{-\frac{\pi}{c_2}}\, e^{c_0 - \frac{c_1^2}{4c_2}}.$$

In particular,

$$\sqrt{-\frac{\pi}{c_2}} = \sqrt{\frac{\pi}{\varepsilon^2 + \nu^2}}$$

and

$$c_0 - \frac{c_1^2}{4c_2} = -(\nu^2 + \varepsilon^2)(\|x\|^2 + t^2) + 2\varepsilon^2 tx \cdot n_\theta + \frac{\varepsilon^4 (x \cdot n_\theta^\perp)^2}{\nu^2 + \varepsilon^2}.$$

This term can be rewritten using (5.4) to eliminate $n_\theta^\perp$ and obtain the second formula of the statement:

$$c_0 - \frac{c_1^2}{4c_2} =$$

$$= -(\nu^2 + \varepsilon^2)(\|x\|^2 + t^2) + 2\varepsilon^2 tx \cdot n_\theta + \frac{\varepsilon^4}{\nu^2 + \varepsilon^2}(\|x\|^2 - (x \cdot n_\theta)^2)$$

$$= -\left(\nu^2 + \varepsilon^2 - \frac{\varepsilon^4}{\varepsilon^2 + \nu^2}\right)\|x\|^2 - (\nu^2 + \varepsilon^2)t^2 + x \cdot n_\theta \left(2\varepsilon^2 t - \frac{\varepsilon^4}{\nu^2 + \varepsilon^2} x \cdot n_\theta\right)$$

$$= -\frac{\nu^2(\nu^2 + 2\varepsilon^2)}{\varepsilon^2 + \nu^2}\|x\|^2 - \left(\sqrt{\nu^2 + \varepsilon^2}\, t - \frac{\varepsilon^2}{\sqrt{\nu^2 + \varepsilon^2}}(x \cdot n_\theta)\right)^2$$

$$= -\frac{1}{\varepsilon^2 + \nu^2}\left(\nu^2(\nu^2 + 2\varepsilon^2)\|x\|^2 + \frac{\varepsilon^4}{\varepsilon^2 + \nu^2}\left(\frac{\nu^2 + \varepsilon^2}{\varepsilon^2}t - x \cdot n_\theta\right)^2\right)$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$$

We can now apply a second Radon transform to the previous result, and we obtain the final kernel as follows.

Proposition 5.6. *The Radon transform $\mathcal{R}^x[\mathcal{R}^y[K_{\varepsilon,\nu}(\cdot,\cdot)](t,\theta)](r,\phi)$ is given by*

$$\mathcal{R}^x[\mathcal{R}^y[K_{\varepsilon,\nu}(\cdot,\cdot)](t,\theta)](r,\phi) = \frac{\pi}{\sqrt{\gamma_{\varepsilon,\nu}(\theta-\phi)}} \, e^{\beta_{\varepsilon\nu}(t,r,\theta-\phi)}$$

where

$$\beta_{\varepsilon\nu}(t,r,\eta) = -\frac{\nu^2(2\varepsilon^2+\nu^2)}{\gamma_{\varepsilon,\nu}(\eta)}\left((\varepsilon^2+\nu^2)(r^2+t^2)-2\varepsilon^2 rt\cos(\eta)\right)$$

$$\gamma_{\varepsilon,\nu}(\eta) = (\varepsilon^2+\nu^2)^2 - \varepsilon^4\cos^2(\eta).$$

Proof. We want to compute the integral

$$\int_{-\infty}^{\infty} \mathcal{R}^y[K_{\varepsilon,\nu}(x_{r,\phi}(s),\cdot)](t,\theta)ds,$$

and we will use the formula (5.5). First, observe that

$$x_{r,\phi}(s)\cdot n_\theta = \cos(\theta)(r\cos(\phi)-s\sin(\phi))+\sin(\theta)(r\sin(\phi)+s\cos(\phi))$$
$$= r\cos(\theta-\phi)+s\sin(\theta-\phi),$$

and similarly $x_{r,\phi}(s)\cdot n_\theta^\perp = s\cos(\theta-\phi)-r\sin(\theta-\phi)$. Thus, the argument of the exponential in the integral is

$$-(\nu^2+\varepsilon^2)(\|x_{r,\phi}(s)\|^2+t^2)+2\varepsilon^2 tx_{r,\phi}(s)\cdot n_\theta+\frac{\varepsilon^4(x_{r,\phi}(s)\cdot n_\theta^\perp)^2}{\nu^2+\varepsilon^2}$$
$$= -(\nu^2+\varepsilon^2)(s^2+r^2+t^2)+2\varepsilon^2 t(r\cos(\theta-\phi)+s\sin(\theta-\phi))+$$
$$+\frac{\varepsilon^4}{\varepsilon^2+\nu^2}(s\cos(\theta-\phi)-r\sin(\theta-\phi))^2.$$

Now we can repeat the same computation as in the previous Lemma and define

$$c_2 = -(\varepsilon^2+\nu^2)+\frac{\varepsilon^4}{\varepsilon^2+\nu^2}\cos(\theta-\phi)^2$$

$$c_1 = 2\varepsilon^2\sin(\theta-\phi)\left(t-\frac{\varepsilon^2}{\varepsilon^2+\nu^2}r\cos(\theta-\phi)\right)$$

$$c_0 = -(\varepsilon^2+\nu^2)(r^2+t^2)+2\varepsilon^2 rt\cos(\theta-\phi)+\frac{\varepsilon^4}{\varepsilon^2+\nu^2}r^2\sin(\theta-\phi)^2,$$

with

$$c_2 \le -(\varepsilon^2+\nu^2)+\frac{\varepsilon^4}{\varepsilon^2+\nu^2} = -\frac{\nu^2(\nu^2+2\varepsilon^2)}{\varepsilon^2+\nu^2} < 0.$$

Thanks to Lemma 5.4 and Lemma 5.5 we have

$$\mathcal{R}^x[\mathcal{R}^y[K_{\varepsilon,\nu}(\cdot,\cdot)](t,\theta)](r,\phi) = \sqrt{\frac{\pi}{\varepsilon^2+\nu^2}}\sqrt{-\frac{\pi}{c_2}}\,e^{c_0-\frac{c_1^2}{4c_2}}.$$

The coefficient can be rewritten as

$$\sqrt{\frac{\pi}{\varepsilon^2+\nu^2}}\sqrt{\frac{\pi}{-c_2}} = \frac{\pi}{\sqrt{(\varepsilon^2+\nu^2)^4-\varepsilon^4\cos(\theta-\phi)^2}} = \frac{\pi}{\sqrt{\gamma_{\varepsilon,\nu}(\theta-\phi)}},$$

while the argument of the above exponential is

$$c_0 - \frac{c_1^2}{4c_2} = -(\varepsilon^2+\nu^2)(r^2+t^2) + 2\varepsilon^2 rt\cos(\theta-\phi) + \frac{\varepsilon^4}{\varepsilon^2+\nu^2}r^2\sin(\theta-\phi)^2$$

$$+ \frac{\varepsilon^4\sin(\theta-\phi)^2\left(t-\frac{\varepsilon^2}{\varepsilon^2+\nu^2}r\cos(\theta-\phi)\right)^2}{(\varepsilon^2+\nu^2)-\frac{\varepsilon^4}{\varepsilon^2+\nu^2}\cos(\theta-\phi)^2},$$

i.e.,

$$\gamma_{\varepsilon,\nu}(\theta-\phi)\left(c_0 - \frac{c_1^2}{4c_2}\right) =$$

$$= \gamma_{\varepsilon,\nu}(\theta-\phi)\left(-(\varepsilon^2+\nu^2)(r^2+t^2) + 2\varepsilon^2 rt\cos(\theta-\phi) + \frac{\varepsilon^4}{\varepsilon^2+\nu^2}r^2\sin(\theta-\phi)^2\right)$$

$$+ \varepsilon^4(\varepsilon^2+\nu^2)\sin(\theta-\phi)^2\left(t-\frac{\varepsilon^2}{\varepsilon^2+\nu^2}r\cos(\theta-\phi)\right)^2$$

$$= \gamma_{\varepsilon,\nu}(\theta-\phi)\left(-(\varepsilon^2+\nu^2)(r^2+t^2) + 2\varepsilon^2 rt\cos(\theta-\phi) + \frac{\varepsilon^4}{\varepsilon^2+\nu^2}r^2\sin(\theta-\phi)^2\right)$$

$$+ \varepsilon^4(\varepsilon^2+\nu^2)\sin(\theta-\phi)^2\left(t^2 - 2\frac{\varepsilon^2}{\varepsilon^2+\nu^2}rt\cos(\theta-\phi) + \frac{\varepsilon^4}{(\varepsilon^2+\nu^2)^2}r^2\cos(\theta-\phi)^2\right).$$

Now $c_0 - \frac{c_1^2}{4c_2}$ can be written in compact form by collecting the coefficients of the terms corresponding to r^2, t^2 and rt, say c_r, c_t and c_{rt}, respectively. Their values are

$$c_r = -\frac{\nu^2(\varepsilon^2+\nu^2)(2\varepsilon^2+\nu^2)}{\gamma_{\varepsilon,\nu}(\theta-\phi)}$$

$$c_t = c_r$$

$$c_{rt} = \frac{2\varepsilon^2\nu^2(2\varepsilon^2+\nu^2)\cos(\theta-\phi)}{\gamma_{\varepsilon,\nu}(\theta-\phi)},$$

and they give the symmetric formula of the statement.

Indeed,

$$\gamma_{\varepsilon,\nu}(\theta - \phi)c_r$$
$$= \gamma_{\varepsilon,\nu}(\theta - \phi)\left(-(\varepsilon^2 + \nu^2) + \frac{\varepsilon^4}{\varepsilon^2 + \nu^2}\sin(\theta - \phi)^2\right)$$
$$\quad + \frac{\varepsilon^8}{(\varepsilon^2 + \nu^2)}\sin(\theta - \phi)^2\cos(\theta - \phi)^2$$
$$= ((\varepsilon^2 + \nu^2)^2 - \varepsilon^4\cos^2(\theta - \phi))\left(-(\varepsilon^2 + \nu^2) + \frac{\varepsilon^4}{\varepsilon^2 + \nu^2}\sin(\theta - \phi)^2\right)$$
$$\quad + \frac{\varepsilon^8}{(\varepsilon^2 + \nu^2)}\sin(\theta - \phi)^2\cos(\theta - \phi)^2$$
$$= -(\varepsilon^2 + \nu^2)^3 + \varepsilon^4(\varepsilon^2 + \nu^2)\cos^2(\theta - \phi) + \varepsilon^4(\varepsilon^2 + \nu^2)^2\sin(\theta - \phi)^2$$
$$\quad - \frac{\varepsilon^8}{\varepsilon^2 + \nu^2}\sin(\theta - \phi)^2\cos(\theta - \phi)^2 + \frac{\varepsilon^8}{(\varepsilon^2 + \nu^2)}\sin(\theta - \phi)^2\cos(\theta - \phi)^2$$
$$= -(\varepsilon^2 + \nu^2)^3 + \varepsilon^4(\varepsilon^2 + \nu^2)$$
$$= (\varepsilon^2 + \nu^2)(\varepsilon^4 - \varepsilon^4 - \nu^4 - 2\varepsilon^2\nu^2)$$
$$= -\nu^2(\varepsilon^2 + \nu^2)(\nu^2 - 2\varepsilon^2),$$

and

$$\gamma_{\varepsilon,\nu}(\theta - \phi)c_t$$
$$= -\gamma_{\varepsilon,\nu}(\theta - \phi)(\varepsilon^2 + \nu^2) + \varepsilon^4(\varepsilon^2 + \nu^2)\sin(\theta - \phi)^2$$
$$= -((\varepsilon^2 + \nu^2)^2 - \varepsilon^4\cos^2(\theta - \phi))(\varepsilon^2 + \nu^2) + \varepsilon^4(\varepsilon^2 + \nu^2)\sin(\theta - \phi)^2$$
$$= -(\varepsilon^2 + \nu^2)^3 + \varepsilon^4(\varepsilon^2 + \nu^2)$$
$$= -\nu^2(\varepsilon^2 + \nu^2)(\nu^2 - 2\varepsilon^2),$$

and

$$\gamma_{\varepsilon,\nu}(\theta - \phi)c_{rt}$$
$$= \gamma_{\varepsilon,\nu}(\theta - \phi)2\varepsilon^2\cos(\theta - \phi) - 2\varepsilon^6\sin(\theta - \phi)^2\cos(\theta - \phi)$$
$$= 2\varepsilon^2\cos(\theta - \phi)\left((\varepsilon^2 + \nu^2)^2 - \varepsilon^4\cos^2(\theta - \phi) - \varepsilon^4\sin(\theta - \phi)^2\right)$$
$$= 2\varepsilon^2\cos(\theta - \phi)\left(\nu^2(2\varepsilon^2 + \nu^2)\right).$$

$\square$

5.2.4 Experiments

As an example, we use here our method to reconstruct some *phantom*, i.e., some reference images that are often used to test image reconstruction algorithms. At this point, these experiments are intended only to show examples of reconstructions in a fixed setting, while further and more accurate

tests are still under development. The three phantoms are depicted in the left of Figure 5.3, and they are called the *bulls-eye*, *crescent-shape* and *Shepp-Logan* (see [25]). We consider phantoms defined over a grid of K^2 pixels, $K = 256$, and we reconstruct them with our method with $N = M = 64$ samples in the parallel-beam geometry, and $n = 1500$. The shape parameters are kept fixed, with values $\varepsilon = 26$ and $\nu = 0.3$. The reconstruction is shown on the right column of Figure 5.3.
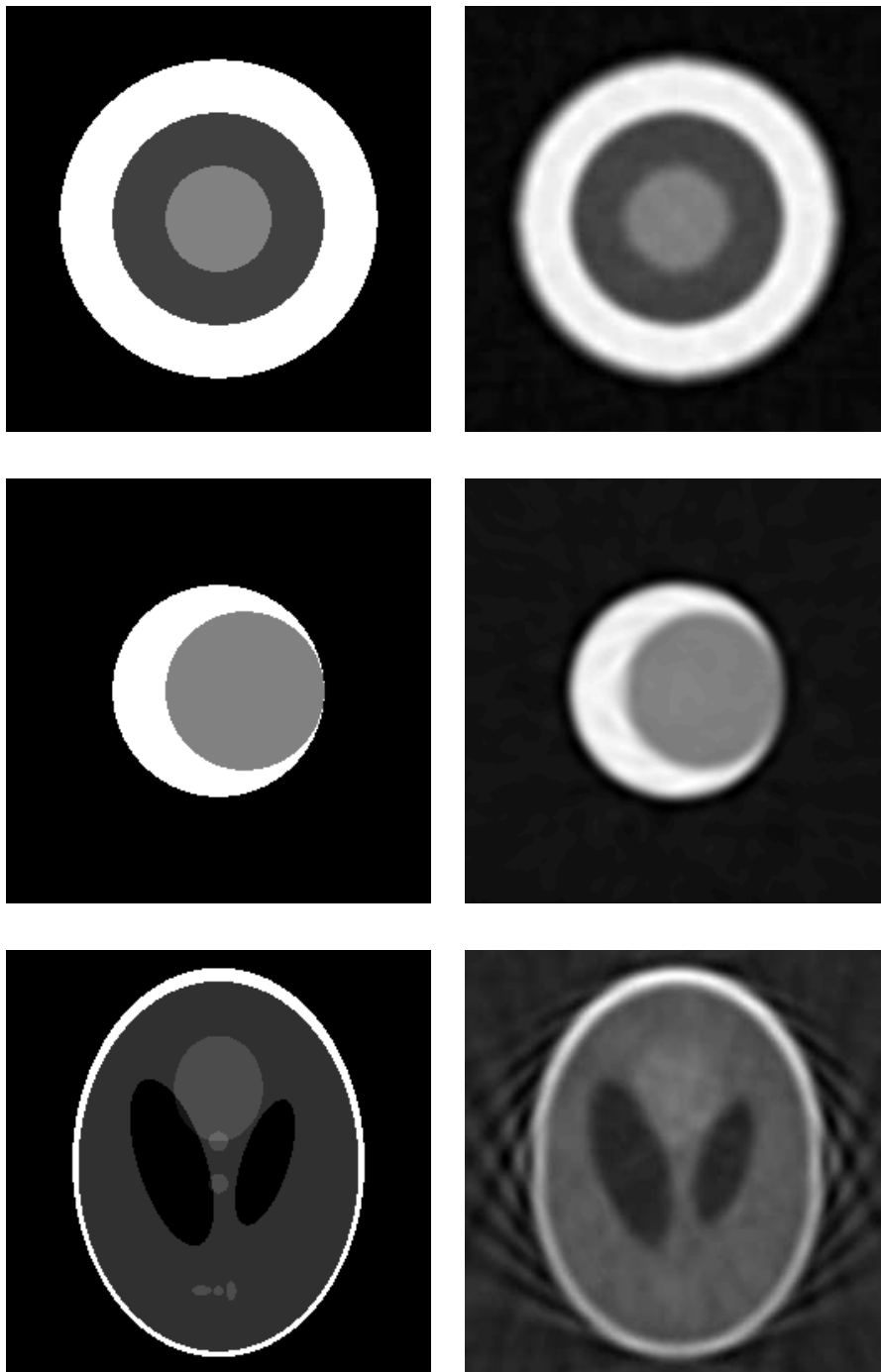
Figure 5.3: Examples of reconstruction from Radon data (left: true phantoms, right: reconstructed phantoms). From top to bottom: bull-eye phantom, crescent-shape phantom, Sheep-Logan phantom.

Appendices

Appendix A

Construction of native spaces from sequences

This Appendix is devoted to present a different way of constructing the native space $\mathcal{H}(\Omega)$.

We will at first review some result first presented in [54]. Namely, instead of starting from a kernel K to build up a Hilbert space as in Chapter 1, we will consider first a sequence of functions in $L_2(\Omega)$ and a sequence of weights, such that they satisfy a summability assumption that will be introduced. This decomposition will then give rise to a kernel through a Mercer series, and to its native space.

This construction allows us to analyze to some extent the relation between the native spaces coming from the same set of eigenfunctions, but with different weights, or eigenvalues. Under certain assumptions that will be precisely stated, the two native spaces satisfy an embedding condition, and relations between the corresponding norms are derived.

Once two different native spaces are constructed, it is possible to retrieve some of the results of Section 2.1 by replacing $L_2(\Omega)$ with a kernel-based space. In particular, the relation between projectors and subspaces is analyzed.

With these tools, it comes easy to partially extend the results of Section 2.2, as is explained in Section 2.5.

We will slightly change the notation used so far in order to take into account the dependence of a kernel on its expansion.

A.1 Native space from sequences

We present here the construction introduced in [54] to build up a kernel from a particular expansion.

Given a compact set $\Omega \subset \mathbb{R}^d$, we consider an orthonormal system $\{\varphi_j\}_j$ in $L_2(\Omega)$ and a set of positive and non increasing weights $\{\lambda_j\}_j$. The se-

quence of the weights is said to be *admissible* with respect to the orthonormal system if they satisfy

$$\sum_j \lambda_j \varphi_j^2(x) < \infty, \quad \text{for all } x \in \Omega. \tag{A.1}$$

For admissible weights, it is well defined on Ω the kernel

$$K_\lambda(x,y) = \sum_j \lambda_j \varphi_j(x) \varphi_j(y).$$

Moreover, if the space spanned by $\{\varphi_j\}_j$ separates points in Ω, the kernel is strictly positive definite (see [54]). We denote by $\mathcal{H}_\lambda(\Omega)$ and $(\cdot,\cdot)_\lambda$ the corresponding native space and inner product.

   With such a kernel we can repeat the construction of Chapter 1. Namely, for any admissible sequence $\{\lambda_j\}_j$, since $\{\varphi_j\}_j \subset L_2(\Omega)$ the native space $\mathcal{H}_\lambda(\Omega)$ is continuously embedded into $L_2(\Omega)$ (see [54, Theorem 7]) and the kernel K_λ defines the integral operator

$$\begin{aligned} T_\lambda: \quad & L_2(\Omega) \to L_2(\Omega) \\ T_\lambda f(x) \quad &= \int_\Omega K_\lambda(x,y) f(y) dy, \end{aligned}$$

having eigenfunctions $\{\varphi_j\}_j$ and eigenvalues $\{\lambda_j\}_j$.

   The operator T_λ is the adjoint of the embedding of $\mathcal{H}_\lambda(\Omega)$ in $L_2(\Omega)$, i.e.,

$$(T_\lambda f, g)_{\mathcal{H}_\lambda} = (f,g)_{L_2} \text{ for all } f \in L_2(\Omega), g \in \mathcal{H}_\lambda(\Omega),$$

and in particular it is injective (i.e., $T_\lambda f = 0$ if and only if $f = 0$ a.e. in Ω). The sequence of functions $\{\sqrt{\lambda_j}\varphi_j\}_j$ is an orthonormal basis of $\mathcal{H}_\lambda(\Omega)$, and the native spaces is

$$\mathcal{H}_\lambda(\Omega) = \left\{ f \in L_2(\Omega) : \sum_j \frac{(f,\varphi_j)_{L_2}}{\lambda_j} < \infty \right\}. \tag{A.2}$$

A.2   Relation between different native spaces

Given another sequence $\{\mu_k\}_k$ that satisfies the summability condition (A.1) with the same orthonormal system, we can define also the kernel K_μ.

   Moreover, under further assumptions on the weights sequences, some relation between the corresponding native spaces can be established.

   First, if there is an index $\bar{j}$ such that $\lambda_j \leq \mu_j$ for all $j > \bar{j}$, by (A.2) we have $\mathcal{H}_\lambda(\Omega) \subset \mathcal{H}_\mu(\Omega)$, and the embedding is continuous. A particular case is when $\lambda_j \leq \mu_j$ for all j, and from now on we will assume to be in this situation for simplicity of exposition.

Second, the embedding can be compact. Namely, from the two weights sequences we can define the further sequence $\{\nu_j\}_j := \{\lambda_j/\mu_j\}_j$. If also this new sequence is admissible with respect to $\{\varphi_j\}_j$, the kernel $K_{\lambda/\mu}$ is well defined and we can consider the operator $T_{\lambda/\mu}$ having eigenvalues given by $\{\nu_j\}_j$ and the same eigenfunctions as before. Under this condition we have

$$T_{\lambda/\mu}[K_\mu(\cdot, y)](x) = K_\lambda(x, y). \tag{A.3}$$

To simplify the discussion we make the further assumption that

$$\frac{\lambda_{j+1}}{\mu_{j+1}} \leq \frac{\lambda_j}{\mu_j} \quad \text{for all } j, \tag{A.4}$$

so that the eigenfunctions of the operator $T_{\lambda/\mu}$ are sorted as $\{\lambda_j\}_j$ and $\{\mu_j\}_j$. If this condition is not satisfied, it suffices to sort $\{\nu_j\}_j$, without changing the order of $\{\lambda_j\}_j$ and $\{\mu_j\}_j$.

As before, the operator $T_{\lambda/\mu}$ is the adjoint of the embedding of $\mathcal{H}_\lambda(\Omega)$ into $\mathcal{H}_\mu(\Omega)$, i.e.,

$$(f, g)_\mu = (T_{\lambda/\mu}f, g)_\lambda \quad \text{for all } f \in \mathcal{H}_\mu(\Omega), g \in \mathcal{H}_\lambda(\Omega),$$

and in particular the norms satisfy

$$\|f\|_\mu \leq \frac{\lambda_1}{\mu_1}\|f\|_\lambda \quad \text{for all } f \in \mathcal{H}_\lambda(\Omega).$$

with equality for $f = \varphi_1$.

Observe that the space $L_2(\Omega)$ is the limiting case corresponding to unit weights, which are not admissible since $\sum_j \varphi_j^2(x)$ is not bounded on Ω (otherwise $L_2(\Omega)$ itself would be a reproducing kernel space, see Lemma 2.2). But the last inequality can be understood in this case as the standard embedding inequality, in the sense that

$$\|f\|_{L_2} \leq \sqrt{\lambda_1}\,\|f\|_\lambda, \quad \text{for all } f \in \mathcal{H}_\lambda(\Omega),$$

again with equality for $f = \varphi_1$.

A.3   Projectors and subspaces

Under the above conditions on $\{\lambda_j\}_j$, $\{\mu_j\}_j$ and $\{\lambda_j/\mu_j\}_j$, we can prove some of the results of the paper, with $\mathcal{H}_\mu(\Omega)$ playing the role of the $L_2(\Omega)$ space.

First, we have the following version of Lemma 2.9.

Lemma A.1. *Any closed subspace $V \subset \mathcal{H}_\lambda(\Omega)$ has a unique basis that is $\mathcal{H}_\mu(\Omega)$-orthonormal and $\mathcal{H}_\lambda(\Omega)$-orthogonal.*

Proof. The proof is the same of the one of Lemma 2.9, and works by restriction of $T_{\lambda/\mu}$ on V. □

Moreover, for any n dimensional subspace $V_n \subset \mathcal{H}_\lambda$ we can define on $\mathcal{H}_\lambda(\Omega)$ the projectors

$$
\begin{aligned}
\Pi_{L_2,V_n} f &= \sum_{j=1}^n (f, w_j)_{L_2} w_j, & f \in \mathcal{H}_\lambda(\Omega), \\
\Pi_{\mathcal{H}_\mu,V_n} f &= \sum_{j=1}^n (f, u_j)_\mu u_j, & f \in \mathcal{H}_\lambda(\Omega), \\
\Pi_{\mathcal{H}_\lambda,V_n} f &= \sum_{j=1}^n (f, v_j)_\lambda v_j, & f \in \mathcal{H}_\lambda(\Omega),
\end{aligned}
$$

where the w_j, u_j and v_j are $L_2(\Omega)$-, $\mathcal{H}_\mu(\Omega)$- and $\mathcal{H}_\lambda(\Omega)$- orthonormal basis functions of V_n, respectively. As in Lemma 2.5, a subspace for which two of the above projectors coincide is necessarily an eigenspace.

Lemma A.2. *If two of the projectors coincide on V_n, then V_n is spanned by n eigenfunctions, and the three projectors coincide on $\mathcal{H}_\lambda(\Omega)$ if V_n is spanned by n eigenfunctions.*

Proof. The proof is the same of the the one of Lemma 2.5, and it works by proving that the double orthogonal (in $\mathcal{H}_\mu(\Omega)$ and $\mathcal{H}_\lambda(\Omega)$) basis of V_n can be extended to be a basis of all of $\mathcal{H}_\lambda(\Omega)$ with the same orthogonality properties if and only if V_n is spanned by eigenfunctions of $T_{\lambda/\mu}$. □

Bibliography

[1] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

[2] K. Atkinson and W. Han. *Theoretical Numerical Analysis*, volume 39 of *Texts in Applied Mathematics*. Springer, Dordrecht, third edition, 2009. A functional analysis framework.

[3] J. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 1975.

[4] M. D. Buhmann. Radial basis functions. In *Acta numerica, 2000*, volume 9 of *Acta Numer.*, pages 1–38. Cambridge Univ. Press, Cambridge, 2000.

[5] M. D. Buhmann. *Radial basis functions: theory and implementations*, volume 12 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2003.

[6] M. D. Buhmann, S. Dinew, and E. Larsson. A note on radial basis function interpolant limits. *IMA J. Numer. Anal.*, 30(2):543–554, 2010.

[7] R. Cavoretto. A numerical algorithm for multidimensional modeling of scattered data points. *Comput. Appl. Math.*, 34(1):65–80, 2015.

[8] R. Cavoretto, S. De Marchi, A. De Rossi, E. Perracchione, and G. Santin. Partition of unity interpolation using stable kernel-based techniques. *in preparation*, 2015.

[9] R. Cavoretto, S. De Marchi, A. De Rossi, E. Perracchione, and G. Santin. Partition of unity interpolation using stable kernel-based techniques. In J. Vigo-Aguiar, editor, *CMMSE 2015 : Proceedings of the 15th International Conference on Mathematical Methods in Science and Engineering*, volume I, pages 317–326, 2015.

[10] R. Cavoretto and A. De Rossi. A meshless interpolation algorithm using a cell-based searching procedure. *Comput. Math. Appl.*, 67(5):1024–1038, 2014.

[11] R. Cavoretto and A. De Rossi. A trivariate interpolation algorithm using a cube-partition searching procedure. *SIAM J. Sci. Comput.*, 37(4):A1891–A1908, 2015.

[12] R. Cavoretto, G. Fasshauer, and M. McCourt. An introduction to the Hilbert-Schmidt SVD using iterated Brownian bridge kernels. *Numerical Algorithms*, pages 1–30, 2014.

[13] P. C. Curtis, Jr. n-parameter families and best approximation. *Pacific J. Math.*, 9:1013–1027, 1959.

[14] S. De Marchi, A. Iske, and G. Santin. Kernel-based image reconstruction from scattered Radon data by anisotropic positive definite functions. *in preparation*, 2015.

[15] S. De Marchi, A. Iske, and A. Sironi. Kernel-based image reconstruction from scattered Radon data by positive definite functions. under revision, 2012.

[16] S. De Marchi and G. Santin. A new stable basis for radial basis function interpolation. *Journal of Computational and Applied Mathematics*, 253:113, Dec. 2013.

[17] S. De Marchi and G. Santin. Fast computation of orthonormal basis for RBF spaces through Krylov space methods. *BIT Numerical Mathematics*, 55(4):949–966, 2014.

[18] S. De Marchi and R. Schaback. Stability of kernel-based interpolation. *Advances in Computational Mathematics*, 32(2):155161, 2010.

[19] S. De Marchi, R. Schaback, and H. Wendland. Near-optimal data-independent point locations for radial basis function interpolation. *Adv. Comput. Math.*, 23(3):317–330, 2005.

[20] T. A. Driscoll and B. Fornberg. Interpolation in the limit of increasingly flat radial basis functions. *Comput. Math. Appl.*, 43(3-5):413–422, 2002. Radial basis functions and partial differential equations.

[21] G. E. Fasshauer. Solving partial differential equations by collocation with radial basis functions. In *Surface Fitting and Multiresolution Methods A. Le Mèhautè, C. Rabut and L.L. Schumaker (eds.), Vanderbilt*, pages 131–138. University Press, 1997.

[22] G. E. Fasshauer. *Meshfree Approximation Methods with MATLAB*, volume 6 of *Interdisciplinary Mathematical Sciences*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007. With 1 CD-ROM (Windows, Macintosh and UNIX).

[23] G. E. Fasshauer and M. McCourt. *Kernel-Based Approximation Methods Using MATLAB*, volume 19 of *Interdisciplinary Mathematical Sciences*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2015.

[24] G. E. Fasshauer and M. J. McCourt. Stable evaluation of Gaussian radial basis function interpolants. *SIAM J. Sci. Comput.*, 34(2):A737–A762, 2012.

[25] T. G. Feeman. *The mathematics of medical imaging*. Springer Undergraduate Texts in Mathematics and Technology. Springer, New York, 2010. A beginner's guide.

[26] B. Fornberg, E. Larsson, and N. Flyer. Stable computations with Gaussian radial basis functions. *SIAM J. Sci. Comput.*, 33(2):869–892, 2011.

[27] B. Fornberg and G. Wright. Stable computation of multiquadric interpolants for all values of the shape parameter. *Comput. Math. Appl.*, 48(5-6):853–867, 2004.

[28] B. Fornberg, G. Wright, and E. Larsson. Some observations regarding interpolants in the limit of flat radial basis functions. *Comput. Math. Appl.*, 47(1):37–55, 2004.

[29] R. Franke. *A critical comparison of some methods for interpolation of scattered data*, volume 253. Naval Postgraduate School Tech.Rep., Mar. 1979.

[30] I. C. Gohberg and M. G. Kreĭn. *Introduction to the theory of linear non-selfadjoint operators*. Translated from the Russian by A. Feinstein. Translations of Mathematical Monographs, Vol. 18. American Mathematical Society, Providence, R.I., 1969.

[31] G. H. Golub and C. F. Van Loan. An analysis of the total least squares problem. *SIAM J. Numer. Anal.*, 17(6):883–893, 1980.

[32] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[33] Y. C. Hon and R. Schaback. On unsymmetric collocation by radial basis functions. *Appl. Math. Comput.*, 119(2-3):177–186, Apr. 2001.

[34] R. Ismagilov. On n-dimensional diameters of compacts in a Hilbert space. *Functional Analysis and Its Applications*, 2(2):125–132, 1968.

[35] J. W. Jerome. On n-widths in Sobolev spaces and applications to elliptic boundary value problems. *J. Math. Anal. Appl.*, 29:201–215, 1970.

[36] M. A. Krasnosel′skiĭ, G. M. Vaĭnikko, P. P. Zabreĭko, Y. B. Rutitskii, and V. Y. Stetsenko. *Approximate Solution of Operator Equations*. Wolters-Noordhoff Publishing, Groningen, 1972. Translated from the Russian by D. Louvish.

[37] E. Larsson and B. Fornberg. Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions. *Comput. Math. Appl.*, 49(1):103–130, 2005.

[38] J. C. Mairhuber. On Haar′s theorem concerning Chebychev approximation problems having unique solutions. *Proc. Amer. Math. Soc.*, 7:609–615, 1956.

[39] 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 A: Mathematical, Physical and Engineering Sciences*, 209(441-458):415–446, 1909.

[40] M. Mouattamid and R. Schaback. Recursive kernels. *Anal. Theory Appl.*, 25(4):301–316, 2009.

[41] S. Müller and R. Schaback. A Newton basis for kernel spaces. *J. Approx. Theory*, 161(2):645–655, 2009.

[42] E. Novak and H. Woźniakowski. *Tractability of multivariate problems. Vol. 1: Linear information*, volume 6 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2008.

[43] P. Novati and M. R. Russo. A GCV based Arnoldi-Tikhonov regularization method. *BIT Numerical mathematics*, 2013.

[44] M. Pazouki and R. Schaback. Bases for kernel-based spaces. *Journal of Computational and Applied Mathematics*, 236(4):575588, 2011.

[45] M. Pazouki and R. Schaback. Bases for conditionally positive definite kernels. *J. Comput. Appl. Math.*, 243:152–163, 2013.

[46] A. Pinkus. n-*Widths in Approximation Theory*, volume 7 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1985.

[47] W. Pogorzelski. *Integral Equations and Their Applications. Vol. I*. Translated from the Polish by Jacques J. Schorr-Con, A. Kacner and Z. Olesiak. International Series of Monographs in Pure and Applied Mathematics, Vol. 88. Pergamon Press, Oxford, 1966.

[48] S. Saitoh. *Integral transforms, reproducing kernels and their applications*, volume 369 of *Pitman Research Notes in Mathematics Series*. Longman, Harlow, 1997.

[49] G. Santin. WSVD and FastWSVD, Matlab code. `http://www.math.unipd.it/˜gsantin/file/WSVD_and_FCoOB.zip`, 2013.

[50] G. Santin. Kernel-based medical image reconstruction, Matlab code. `http://www.math.unipd.it/˜gsantin/file/KBMIR.zip`, 2014.

[51] G. Santin. Direct and greedy eigenbasis construction, Matlab code. `http://www.math.unipd.it/˜gsantin/file/EigenApprox.zip`, 2015.

[52] G. Santin and R. Schaback. Approximation of eigenfunctions in kernel-based spaces. *Adv. Comput. Math.*, 2015.

[53] R. Schaback. Error estimates and condition numbers for radial basis function interpolation. *Adv. Comput. Math.*, 3(3):251–264, 1995.

[54] R. Schaback and H. Wendland. Approximation by positive definite kernels. In M. Buhmann and D. Mache, editors, *Advanced Problems in Constructive Approximation*, volume 142 of *International Series in Numerical Mathematics*, pages 203–221, 2002.

[55] R. Schaback and H. Wendland. Inverse and saturation theorems for radial basis function interpolation. *Math. Comp.*, 71(238):669–681 (electronic), 2002.

[56] R. Schaback and H. Wendland. Kernel techniques: from machine learning to meshless methods. *Acta Numer.*, 15:543–639, 2006.

[57] H. Sun and Q. Wu. Application of integral operator for regularized least-square regression. *Math. Comput. Modelling*, 49(1-2):276–285, 2009.

[58] H. Van Trees. *Detection, Estimation, and Modulation Theory*. Number pt. 1 in Detection, Estimation, and Modulation Theory. Wiley, 2004.

[59] H. Wendland. Fast evaluation of radial basis functions: methods based on partition of unity. In *Approximation theory, X (St. Louis, MO, 2001)*, Innov. Appl. Math., pages 473–483. Vanderbilt Univ. Press, Nashville, TN, 2002.

[60] H. Wendland. *Scattered Data Approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2005.