

# Stable Bases for Kernel Based Methods

## Dissertation

zur Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Georg-August-Universität Göttingen

vorgelegt von  
Maryam Pazouki  
aus  
Teheran (Iran)

Göttingen 2012

*Referent* : Prof. Dr. R. Schaback  
*Koreferent* : Prof. Dr. G. Lube  
*Tag der mündlichen Prüfung:*

## Acknowledgments

I would like to use this chance to thank everybody who helped me complete this thesis.

This work could not have been possible without Prof. Robert Schaback, who encouraged and challenged me throughout my academic program. I wish to express my sincere gratitude to his great support and patience while supervising my studies. His enthusiasm, precision and criticism have enormously influenced my development as a mathematician.

Furthermore, my thanks go to Prof. Gert Lube who gave his valuable advice and agreed to be my co-examiner.

Most especially thanks go to my family. Words alone cannot express what I owe them for their encouragement and support which has always been my source of strength and inspiration. I am grateful to my friend Trutz Thelemann for keeping my spirit up during my thesis and all his great assistance. I am obliged to many of my colleagues and friends who supported me in this three years.



dedicated to my parents

**Mohammad Pazouki and Fakhro Saadat Amini**



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Native Hilbert Spaces for Kernels</b>	<b>5</b>
2.1	Unconditionally Positive Definite Kernels . . . . .	5
2.2	Data-independent Bases . . . . .	10
2.3	Conditionally Positive Definite Kernels . . . . .	11
2.3.1	Conditional Lagrange Basis . . . . .	14
2.3.2	Reduced Kernel . . . . .	14
<b>3</b>	<b>Bases for Kernel Based Spaces</b>	<b>17</b>
3.1	General Data-dependent Bases . . . . .	17
3.2	Dual Basis . . . . .	21
3.3	Power Function . . . . .	24
3.4	$\mathcal{K}$ -Orthonormal Bases . . . . .	24
3.5	Discretely Orthonormal Bases . . . . .	27
3.6	SVD Bases . . . . .	29
3.7	Newton Basis . . . . .	30
3.8	Generalized Interpolation . . . . .	33
3.9	Notes and comments . . . . .	34
<b>4</b>	<b>Adaptive Calculation of Newton Bases</b>	<b>35</b>
4.1	$P$ -dependent Point Selection Strategy . . . . .	35
4.2	$f$ -dependent Point Selection Strategy . . . . .	38
4.3	Numerical Tests . . . . .	39
<b>5</b>	<b>The Conditionally Positive Definite Bases</b>	<b>43</b>
5.1	Notation and Basic Facts . . . . .	43
5.2	General Data-Dependent Bases . . . . .	45
5.3	Polynomial Projectors . . . . .	49
5.3.1	Partially Orthonormal Bases . . . . .	50
5.4	Eliminating Moment Conditions . . . . .	52
5.5	Points and Polynomials First . . . . .	53
5.6	Standard Basis . . . . .	58
5.7	Lagrange Basis . . . . .	59
5.8	Newton Basis, Iterative Construction . . . . .	61
5.9	Dual Bases . . . . .	63
5.10	Numerical Tests . . . . .	64
5.11	Notes and Comments . . . . .	66
<b>6</b>	<b>Conclusion</b>	<b>69</b>



# Introduction

---

In practical applications over a wide field of study such as surface reconstruction, fluid-structure interaction, and kernel learning one often deals with the problem of reconstructing an unknown function from irregular “scattered” data sites, which in some applications lie in a very high-dimensional space. One possible way to construct a function from scattered data is interpolation. However, the interpolation can be done by polynomials in the univariate case, but working in space dimension  $d \geq 2$ , by the well known theorem of Mairhuber [15], it is impossible to fix a multivariate function space for interpolation on all sets of scattered data. But there is a remarkable situation, where all space dimensions can be handled. Define the multivariate *Radial basis function* by  $\phi(\|x - y\|_2)$  on  $\mathbb{R}^d \times \mathbb{R}^d$ , where  $x, y \in \mathbb{R}^d$  with a smooth univariate function  $\phi : [0, \infty] \rightarrow \mathbb{R}$ . This generates a data-dependent space of functions. In particular, the span of *translates*  $K(\cdot, x_j) := \phi(\|x - x_j\|_2)$  for scattered *centers*  $x_j \in \Omega$  forms a very useful trial space. Furthermore, the theory of the radial basis functions leads the theory of the *reproducing kernel Hilbert spaces*  $\mathcal{K}$  of functions on domains  $\Omega$ , where the kernel  $K$  reproduces functions  $f \in \mathcal{K}$  with the inner products  $(\cdot, \cdot)_{\mathcal{K}}$  as

$$(f, K(x, \cdot))_{\mathcal{K}} = f(x), \text{ for all } x \in \Omega, f \in \mathcal{K}. \quad (1.1)$$

A reproducing kernel Hilbert space can be associated to any positive definite kernel  $K : \Omega \times \Omega \rightarrow \mathbb{R}$ . A symmetric kernel  $K$  is a positive (semi-)definite on  $\Omega$ , if for all sets  $X = \{x_1, \dots, x_N\}$  of  $\Omega$  the *kernel matrix*

$$A := (K(x_j, x_k))_{1 \leq j, k \leq N},$$

is symmetric and positive (semi-)definite. This correspondence is the basis of many useful algorithms.

It is well-known that the standard basis of translates leads to ill-conditioned kernel matrices, but results of [8] show that interpolants to data on  $X$ , when viewed as functions, are rather stable. This leads us to consider different data-dependent bases.

Chapter 3 will be concerned with suitable bases for subspaces of  $\mathcal{K}$ , if  $K$  is positive definite. We concentrate to introduce a new technique for finding bases in kernel spaces for multivariate interpolation on a set of scattered data. These will come in different forms, and we shall explore their variety

and prove results about their connection.

It is evident that if a square matrix  $A$  happens to be symmetric and positive definite, then it has some special efficient decomposition. While symmetric positive definite matrices seem to be rather special, they occur quite frequently in some applications, so their special factorization provides a variety of different bases based on factorizations of the kernel matrix. These bases differ in their stability, orthogonality, adaptivity, duality, and computational efficiency properties. Special emphasis is given to the “Newton” basis appeared in [18] arising from a pivoted Cholesky factorization. It provides stable and recursively computable sets of basis functions while being orthonormal in the “native” Hilbert space of the kernel. There are efficient adaptive algorithms for calculating the Newton basis along the lines of Orthogonal Matching Pursuit. We shall describe this in Chapter 4.

We should also consider “data-independent” bases. The most natural is the *eigenfunction basis* coming from the Mercer theorem. We shall describe it in Section 2.2 and compare it later to data-dependent bases. Of course, data-independent bases are much more “natural” than data-dependent bases, and they provide intrinsic information about the kernel. But when starting with a data-independent-basis, problems may arise when working on a set  $X$  of centers later, because the Mairhuber–Curtis theorem [15] shows that matrices of values of a data-independent basis on points of  $X$  have a data-dependent rank, if one is in a truly multivariate and nontrivial situation. This is why we shall use the data-independent eigenfunction basis only for theoretical comparisons.

Chapter 5 will be concerned with suitable bases for subspaces of  $\mathcal{K}$ . We will define different bases, and explore their connection. But Chapter 3 will only deal with “unconditionally” positive definite kernels. The interpolation problem does not work for all choices of radial basis functions. In special cases of basis functions, for instance the thin-plate spline  $\phi(x) = \|x\|_2^2 \log(\|x\|_2)$ ,  $x \in \mathbb{R}^d$ , let  $N = d + 1$  and let the centers be the vertices of a regular simplex whose edges are all of unit length. Then all entries  $\phi(x_j - x_k)$  of the interpolation matrix are zero, so then interpolation is not guaranteed for multivariate scattered data interpolation, unless the data sites are in special locations. In this case, the kernel is not unconditionally positive definite. It belongs to the important class of conditionally positive definite kernels. We shall consider data-dependent bases for these kernels. The notation of conditionally positive kernels requires the space  $\mathcal{P}_m^d$  of  $d$ -variate polynomials of order (= degree + 1)  $m$ , and in some sense words modulo these polynomials. We leave details to Chapter 5, but we should call  $m$  the order of conditional positive definiteness of a kernel  $K$ , and we need the notation  $Q := \dim \mathcal{P}_m^d = \binom{m-1+d}{d}$ , throughout.

To deal with bases for conditionally positive definite kernels of order  $m$  that provide interpolation on sets  $X = \{x_1, \dots, x_N\} \in \mathbb{R}^d$ , we need “unisolvency”

of  $X$ , meaning that zero is only polynomial in  $\mathcal{P}_m^d$  that vanishes on  $X$ . Then there are various way to construct data-dependent bases. One can preselect a minimal unisolvent subset of  $Q$  points first and then use a “reduced” kernel. Also, one can try to reduce the conditional positive definite case to the unconditionally positive definite case by a suitable change of the kernel, but even this change can be done in different ways. These different approaches sometimes lead to the same result, all of this will be treated in Chapter 5. The bases we construct will in some cases be orthogonal or orthonormal. They can be based on Cholesky or SVD factorization of certain matrices, or just by requiring that the basis should have orthonormal vectors of value on  $X$ . In the end of Chapter, we provide numerical examples for illustration.

### Brief History

Initially, the main purpose of basic radial basis functions method comes from application in geodesy, geophysics, mapping, or meteorology, and recently in many contexts, including approximation, surface reconstruction, and numerical analysis of fluid-structure interactions. Radial basis functions have become a popular mathematical discipline which started with the practical work of Hardy [13] in 1971 and theoretical work of Duchon [9], lately it developed some domain of mathematics and physics by Dyn [10], Powell [19], Schaback and Wendland [22, 20, 25], Buhmann [4],[5]. Conditionally positive definite kernels arise in many contexts including approximation function algorithms, surface reconstruction [27], [23], numerical analysis of fluid-structure interactions [28], computer experiments [1] and geostatistics [6], [26]. They generalize the well known positive definite kernel case, that later applications were found in many other areas such as numerical solutions PDEs, computer graphics, statistical learning theory, signal and image processing, finance and optimization. The current theoretical framework in conditionally positive definite kernel context is the native spaces theory which recently is developed by H. Wendland [27].



# Native Hilbert Spaces for Kernels

---

This chapter serves as an introduction for readers working in applications. It describes some properties of kernel functions in Hilbert space and it provides some guidelines for error estimation and stability of interpolation in Hilbert spaces of functions. We do this mainly by standard methods of linear algebra.

## 2.1 Unconditionally Positive Definite Kernels

In practice, users have a finite set  $X := \{x_1, \dots, x_N\} \subseteq \Omega \subseteq \mathbb{R}^d$  of *centers* or *data locations* and work in a subspace

$$\mathcal{K}_X := \text{span} \{K(\cdot, x_j) : x_j \in X\}, \quad (2.1)$$

spanned by the basis of (generalized) *translates*  $K(\cdot, x_j)$ ,  $1 \leq j \leq N$ .

**Definition 2.1.1.** *A kernel  $K : \Omega \times \Omega \rightarrow \mathbb{R}$  is a positive (semi-)definite symmetric kernel on  $\Omega$ , if for all finite sets  $X := \{x_1, \dots, x_N\} \subseteq \Omega$ , the kernel matrix*

$$A := (K(x_j, x_k))_{1 \leq j, k \leq N}, \quad (2.2)$$

*is symmetric and positive (semi-)definite.*

We now assume  $K$  to be a symmetric real-valued positive semidefinite kernel on  $\Omega$ . Note that general elements from  $\mathcal{K}_X$  with  $a \in \mathbb{R}^N$  take the form

$$f_{a,X}(x) := \sum_{j=1}^N a_j K(x_j, x),$$

i.e. all finite linear combinations of generalized translates of the kernel. We define the point evaluation functionals  $\delta_x : f \mapsto f(x)$ , for  $f$  at  $x \in \Omega$  by  $\delta_x(f) = f(x)$ . The linear space of all linear combinations of point evaluation functionals is defined as

$$L := \text{span}\{\delta_x : x \in \Omega\},$$

while  $X = \{x_1, \dots, x_N\} \subseteq \Omega$ , then

$$L_X := \text{span}\{\delta_x : x \in X\},$$

and

$$\lambda_{a,X} := \sum_{j=1}^N a_j \delta_{x_j}.$$

For another functional

$$\mu_{b,Y} := \sum_{k=1}^M b_k \delta_{y_k}, \quad b \in \mathbb{R}^M$$

we have

$$\mu_{b,Y}(f_{a,X}) = \lambda_{a,X}(f_{b,Y}),$$

and can define bilinear forms

$$\begin{aligned} (f_{a,X}, f_{b,Y}) &:= \lambda_{a,X}(f_{b,Y}) = \mu_{b,Y}(f_{a,X}) \\ &=: (\mu_{a,X}, \mu_{b,Y}). \end{aligned} \tag{2.3}$$

By standard arguments, there are an inner products on

$$\mathcal{K}_\Omega := \text{span}\{K(\cdot, x) : x \in \Omega\},$$

and  $\mathcal{L}_\Omega$ , turning them into pre-Hilbert spaces. Then one can define

$$\begin{aligned} \mathcal{K} &:= \text{close}_{(\cdot, \cdot)_{\mathcal{K}_\Omega}} \mathcal{K}_\Omega \\ \mathcal{L} &:= \text{close}_{(\cdot, \cdot)_{\mathcal{L}_\Omega}} \mathcal{L}_\Omega, \end{aligned}$$

using the respective inner products that we now denote by  $(\cdot, \cdot)_{\mathcal{K}}$  and  $(\cdot, \cdot)_{\mathcal{L}}$ , with  $\mathcal{L} = \mathcal{K}^*$  being the dual of  $\mathcal{K}$  with the Riesz map that is the closure of the map

$$R : \lambda_{a,X} \mapsto f_{a,X}, \quad \mathcal{L}_\Omega \rightarrow \mathcal{K}_\Omega.$$

The formula (2.3) then implies the reproduction formula

$$f(x) = (f, K(x, \cdot))_{\mathcal{K}},$$

for all  $f \in \mathcal{K}$ ,  $x \in \Omega$ , and in particular

$$(K(x, \cdot), K(y, \cdot))_{\mathcal{K}} = K(x, y) = (\delta_x, \delta_y)_{\mathcal{L}}$$

for all  $x, y \in \Omega$ . We call  $\mathcal{K}$  the “native” Hilbert space for  $K$ . It is unique up to isometry, and we sometimes use the notation  $N_K$  for  $\mathcal{K}$ , if we start from a kernel  $K$ . It is well-known (see the books [3, 27, 11] for a full account of this research area)

**Theorem 2.1.1.** *If the point evaluation functionals in a Hilbert space  $\mathcal{K}$  of functions on some domain  $\Omega$  are continuous, then has a reproducing kernel function  $K : \Omega \times \Omega \mapsto \mathbb{R}$  with*

$$(f, K(x, \cdot))_{\mathcal{K}} = f(x) \text{ for all } x \in \Omega, f \in \mathcal{K}, \tag{2.4}$$

*and  $K$  is a positive semidefinite kernel. Furthermore,  $\mathcal{K}$  can be identified with the native Hilbert space for  $K$ .*

A Hilbert space  $\mathcal{K}$  is called *reproducing kernel Hilbert space* if (2.4) holds with a positive semidefinite kernel. Thus the point evaluation functionals  $\delta_x$  have the functions  $K(x, \cdot)$  as Riesz representers in the dual  $\mathcal{K}^*$  of  $\mathcal{K}$ , and satisfy

$$|\delta_x(f)| = |f(x)| \leq \|\delta_x\|_{\mathcal{K}^*} \|f\|_{\mathcal{K}}, \text{ for all } x \in \Omega, f \in \mathcal{K}$$

and the kernel can be expressed as

$$K(x, y) = (K(x, \cdot), K(y, \cdot))_{\mathcal{K}} = (\delta_x, \delta_y)_{\mathcal{K}^*}, \text{ for all } x, y \in \Omega. \quad (2.5)$$

The values of such kernels can always be bounded by

$$|K(x, y)|^2 \leq K(x, x)K(y, y). \text{ for all } x, y \in \Omega.$$

**Corollary 2.1.1.** *Let  $\mathcal{K}$  be a Hilbert space of real- or complex- valued functions on  $\Omega$ , with continuous point evaluation functionals, then  $\mathcal{K}$  is a reproducing kernel Hilbert space, and its kernel  $K : \Omega \times \Omega$  is a positive semidefinite function.*

**Corollary 2.1.2.** *If in addition, point evaluation functionals for different points are linearly independent, then the kernel is positive definite.*

Scattered data interpolation needs a comprehensive mathematical treatment to reconstruct the value of function  $f$  by a linear combination. Now fix a finite set  $X = \{x_1, \dots, x_N\} \subseteq \Omega$ , to interpolate values  $f(x_1), \dots, f(x_N)$  of a function  $f$  on  $X$ , let us introduce the vectors

$$f_X := (f(x_1), \dots, f(x_N))^T, \quad \alpha_X = (\alpha_1, \dots, \alpha_N)^T$$

satisfying the linear system

$$A_{X,X} \alpha_X = f_X,$$

where  $A_{X,X}$  is the kernel matrix on  $X$ , then the function

$$s_{X,f}(x) = \alpha_X^T K_X(x),$$

where  $K : \Omega \times \Omega \rightarrow \mathbb{R}$ , a symmetric positive definite kernel, interpolates  $f$  on  $X$  and we have a basis of translates as components of the vector

$$K_X(x) := (K(x, x_1), \dots, K(x, x_N))^T.$$

**Theorem 2.1.2.** *Given a function  $f$  from the native space  $\mathcal{K}$ , then the solution  $s = f_{\alpha^*, X}$  of the finite-dimensional approximation problem*

$$\min_{s \in \mathcal{K}_X} \|f - s\|_{\mathcal{K}} = \min_{a \in \mathbb{R}^N} \|f - f_{a,X}\|_{\mathcal{K}},$$

is an interpolant of  $f$  on  $X$ , and the coefficients satisfy the interpolation conditions

$$f_{a^*,X}(x_k) = \sum_{j=1}^N \alpha_j^* K(x_j, x_k) = f(x_k), \quad 1 \leq k \leq N.$$

Also, the interpolating functions  $f_{a^*,X}$  minimize the native space norm for all  $g \in N_X$ , i.e.

$$\min_{g=f \text{ on } X} \|g\|_{\mathcal{K}} = \|f_{a^*,X}\|_{\mathcal{K}}.$$

Note that the linear system with  $f \in \mathcal{K}$ , and the kernel matrix  $A_{X,X}$  is always solvable, though the matrix may be singular. The matrix is only positive semidefinite, but the right-hand side of interpolation conditions, being a set of values on  $X$  of a function from the native space, always lies in the span of the columns. The coefficients  $a_j^*$  and the resulting function  $f_{a^*,X}$  may not be unique unless the kernel is positive definite.

Define the function  $f_X(\cdot) := K(x, \cdot) \in \mathcal{K}$  and carry the interpolation construction for  $x \in \Omega$ . Then define and solve the system

$$\sum_{j=1}^N u_j(x) K(x_j, x_k) = K(x_k, x), \quad 1 \leq k \leq N, x \in \Omega.$$

Since  $f_{x_k} = K(x_k, \cdot)$  is in the native space, this system is solvable, and its solution yields the Lagrange basis with characteristic equation  $u_j(x_k) = \delta_{j,k}$ ,  $1 \leq j, k \leq N$ . Define the linear Quasi-interpolation operator  $Q_X(f)$  on the native space  $\mathcal{K}$ , which reproduces all functions on  $X$  from  $\mathcal{K}$ , i.e

$$Q_X(f) := \sum_{j=1}^N u_j(\cdot) f(x_j),$$

**Definition 2.1.2.** Given the Lagrange basis  $u_j, 1 \leq j \leq N$  on  $X = \{x_1, \dots, x_N\}$ , and a positive definite kernel  $K$ , define the **Power Kernel**

$$\begin{aligned} K_N(x, y) &= K(x, y) - \sum_{j=1}^N u_j(x) K(x_j, y) - \sum_{k=1}^N u_k(y) K(x, x_k) \\ &+ \sum_{j,k=1}^N u_j(x) u_k(y) K(x_j, x_k), \end{aligned}$$

where  $K_0(x, y) = K(x, y)$ .

Then the power kernel has the following important properties

$$\begin{aligned} K_N(x_j, x) &= K_N(x, x_j) = 0, \quad 1 \leq j \leq N \\ K_N(\cdot, x_j) &\in \text{span}\{K(\cdot, x_i) : 1 \leq i \leq j\}, \end{aligned}$$

$$K_{j+1}(x, y) = K_j(x, y) - \frac{K_j(x, x_{j+1})K_j(x_{j+1}, y)}{K_j(x_{j+1}, x_{j+1})}, \quad 1 \leq j \leq N-1$$

**Definition 2.1.3.** Let  $K$  be a positive semidefinite kernel on  $\Omega$ , and let  $X = \{x_1, \dots, x_N\} \subseteq \Omega$  be a fixed set. Define the **power function**

$$\begin{aligned} P_N^2(x) &:= \|K_N(x, \cdot)\|_{\mathcal{K}}^2 \\ &= K(x, x) - 2 \sum_{j=1}^N u_j(x) K(x_j, x) + \sum_{j,k=1}^N u_j(x) u_k(x) K(x_j, x_k). \end{aligned}$$

For all points  $x \in \Omega$  the coefficient functions  $u(x)$  define the pointwise error functional

$$f \mapsto \delta_x(f) - s_f(x) = \delta_x(f) - \sum_{j=1}^N u_j(x) \delta_{x_j}(f),$$

that has the norm

$$P_X(x) := \left\| \delta_x - \sum_{j=1}^N u_j(x) \delta_{x_j} \right\|.$$

It allows pointwise error bounds

$$|f(x) - s_f(x)| \leq P(x) \|f\|_{\mathcal{K}}, \quad \text{for all } x \in \Omega, f \in \mathcal{K}. \quad (2.6)$$

By well-known optimality arguments [27], it satisfies

$$P_X^2(x) = \min_{a \in \mathbb{R}^N} \left\| \delta_x - \sum_{j=1}^N a_j \delta_{x_j} \right\|^2, \quad (2.7)$$

the optimal coefficients  $a_j^*(x)$  being given by the Lagrange basis functions  $u_j^*(x)$ . Using this optimality, the formula (2.5) and some standard algebraic manipulations within our formalism implies that the power function is independent of the basis chosen in  $\mathcal{K}_X$ . The power function is practically important because for fixed  $X$ , it can be calculated explicitly everywhere in  $\Omega$ , and the error bound for interpolant allows to estimate the error of all possible interpolants based on the data locations in  $X$ . The power function has some more properties see [23], following from (2.7)

$$\begin{aligned} P_\emptyset^2(x) &= K(x, x), & \text{for all } x \in \Omega \\ P_Y(x) \leq P_X(x) &\leq P_\emptyset(x), & \text{for all } x \in \Omega, X \subseteq Y \subseteq \Omega \\ P_X(x) &= \inf_{g \in \mathcal{K}_X} \|K(x, \cdot) - g\|_{\mathcal{K}}, & \text{for all } x \in \Omega \end{aligned}$$

By an additional duality argument based on (2.6), we obtain

$$\begin{aligned} P_X(x) &= \sup_{\substack{f \in \mathcal{K} \\ \|f\|_{\mathcal{K}} \leq 1 \\ f(X) = 0}} f(x), & \text{for all } x \in \Omega \end{aligned}$$

In addition, due to (2.7) the power function  $P_X(x)$  always vanishes at the points of  $X$ .

## 2.2 Data-independent Bases

This framework of data-independent bases appeared in [14]. Generally, the natural basis for Hilbert spaces with reproducing kernels comes from

**Theorem 2.2.1.** (*Mercer*)

*Continuous positive definite kernels  $K$  on bounded domains  $\Omega$  have an eigenfunction expansion*

$$K(x, y) = \sum_{n=0}^{\infty} \lambda_n \varphi_n(x) \varphi_n(y) \text{ for all } x, y \in \Omega$$

*which is absolutely and uniformly convergent. Furthermore,*

$$\begin{aligned} \lambda_n \varphi_n(x) &= \int_{\Omega} K(x, y) \varphi_n(y) dy \text{ for all } x \in \Omega, n \geq 0, \\ \{\varphi_n\}_n &\text{ orthonormal in } \mathcal{K}, \\ \{\varphi_n\}_n &\text{ orthogonal in } L_2(\Omega), \\ \|\varphi_n\|_2^2 &= \lambda_n \rightarrow 0, n \rightarrow \infty. \end{aligned}$$

This basis directly describes the action of the kernel as an integral operator performing a generalized convolution. In many cases, the eigenvalues decrease very rapidly towards zero, and this implies that there is a very good low-rank approximation to the kernel. This observation has serious consequences for kernel-based algorithms, because one has to encounter rank loss in linear systems that are based on values of  $K$ . The systems will show very bad condition, but on the other hand one knows that the rank-reduced system will be very close to the exact system. Thus, in many cases, linear systems arising from kernels have bad condition, but they also have a low-rank subsystem that performs like the full system, i.e. allows to approximate the right-hand side very much the same way as the full matrix does. This has been observed in many applications from Machine Learning to PDE solving, and it cannot be eliminated by other choices of bases, only by proper regularization or by adaptive techniques that find and use good subproblems. We shall come back to this issue later in Chapter 5

The numerical calculation of approximations to the eigenfunction basis can be based on a sufficiently large point set  $X = \{x_1, \dots, x_N\}$  which allows numerical integration

$$\int_{\Omega} f(y) dy \approx \sum_{j=1}^N w_j f(x_j)$$

for functions  $f \in \mathcal{K}$  using certain positive weights  $w_1, \dots, w_N$ . Then the discretization of the eigenfunction equation

$$\lambda_n \varphi_n(x_k) = \int_{\Omega} K(x_k, y) \varphi_n(y) dy$$

leads to

$$\lambda_n \underbrace{\sqrt{w_k} \varphi_n(x_k)}_{v_k^{(n)}} \approx \sum_{j=1}^N \underbrace{\sqrt{w_k} K(x_k, x_j) \sqrt{w_j}}_{b_{jk}} \underbrace{\sqrt{w_j} \varphi_n(x_j)}_{v_j^{(n)}}$$

and which is an approximation to the discrete eigenvalue problem

$$\lambda_n v_k^{(n)} = \sum_{j=1}^N b_{jk} v_j^{(n)}, \quad 1 \leq j, n \leq N \quad (2.8)$$

involving a scaled version of the kernel matrix (2.2). The obtained values  $v_j^{(n)}$  lead to functions  $v^{(n)} \in \mathcal{K}_X$  by solving the interpolation problems

$$v^{(n)}(x_j) = v_j^{(n)}, \quad 1 \leq j, n \leq N,$$

in the space  $\mathcal{K}_X$ , and we see that we have constructed a data-dependent basis as an approximation to a data-independent basis. Thus this case falls into the next section, we shall come back to it later.

## 2.3 Conditionally Positive Definite Kernels

The previous discussion covers kernels like

- the Gaussian  $K(x, y) = \exp(-\|x - y\|_2^2)$ ,
- inverse multiquadrics  $K(x, y) = (1 + \|x - y\|_2^2)^{-n}$ ,  $n > 0$ ,
- and Wendland's compactly supported kernels like  $K(x, y) = (1 - \|x - y\|_2)_+^4 (1 + 4\|x - y\|_2)$ .

But there are other kernels, e.g.

- multiquadrics  $K(x, y) = (-1)^{\lceil \beta/2 \rceil} (1 + \|x - y\|_2^2)^{\beta/2}$ ,  $\beta \in (0, \infty) \setminus 2\mathbb{Z}$ ,
- powers  $K(x, y) = (-1)^{\lceil \beta/2 \rceil} \|x - y\|_2^\beta$ ,  $\beta \in (0, \infty) \setminus 2\mathbb{Z}$ , or
- thin-plate splines  $K(x, y) = (-1)^{1+\beta/2} \|x - y\|_2^\beta \log \|x - y\|_2$ ,  $\beta \in 2\mathbb{Z}$

that are not covered so far, because the kernels are not positive definite, but only *conditionally* positive definite of some positive order  $m = \lceil \beta/2 \rceil$ ,  $\lceil \beta/2 \rceil$ , and  $m = 1 + \beta/2$  in the above three cases, respectively.

To define the general notation of conditional positive semidefiniteness, we fix the finite dimensional space  $\mathcal{P}_m^d$  of  $d$ -variate real-valued polynomials of order ( $=$  degree  $+ 1$ ) at most  $m$ , which does not necessarily consist of monomials, with

$$Q := \binom{m-1+d}{d} = \dim \mathcal{P}_m^d.$$

Then we have to restrict the admissible point sets  $X \subset \mathbb{R}^d$  by

**Definition 2.3.1.** A subset  $X = \{x_1, \dots, x_N\} \subset \Omega \subset \mathbb{R}^d$  is called  $\mathcal{P}_m^d$ -unisolvent, if the only function in  $\mathcal{P}_m^d$  that vanishes on  $X$  is zero.

We consider  $\mathcal{P}_m^d$ -unisolvent sets  $X = \{x_1, \dots, x_N\} \subset \Omega \subset \mathbb{R}^d$  of  $N \geq Q$  points, where unisolvency means that the  $N \times Q$  matrix

$$P_X := (p_i(x_j))_{1 \leq j \leq N, 1 \leq i \leq Q}, \quad (2.9)$$

has full rank  $Q$ , if  $p_1, \dots, p_Q$  is an arbitrary basis of  $\mathcal{P}_m^d$ .

**Definition 2.3.2.** A kernel  $K : \Omega \times \Omega \rightarrow \mathbb{R}^d$  on a set  $\Omega \subseteq \mathbb{R}^d$  is conditionally positive (semi-) definite of order  $m \geq 0$ , if for all point sets  $X = \{x_1, \dots, x_N\} \subset \Omega$  that are  $\mathcal{P}_m^d$ -unisolvent, the quadratic forms

$$Q_X(c) := \sum_{j,k=1}^N c_j c_k K(x_j, x_k), \quad c \in \mathbb{R}^N, \quad (2.10)$$

are positive (semi-) definite on the subspaces

$$M_{X,m} := \left\{ c \in \mathbb{R}^N : \sum_{j=1}^N c_j p(x_j) = 0 \text{ for all } p \in \mathcal{P}_m^d \right\}, \quad (2.11)$$

of coefficients satisfying discrete moment conditions of order  $m$  on  $X$ .

We shall also use the terms “ $\mathcal{P}_m^d$ -conditionally positive (semi-)definite” with a fixed  $m$ , always in the sense of “conditionally positive (semi-)definite of order  $m$ ”. We know assume that the kernel is assumed to be  $\mathcal{P}_m^d$ -conditionally positive definite. Then on all  $\mathcal{P}_m^d$ -unisolvent sets  $X$ , we can define the symmetric *kernel matrix*

$$A_{X,X} := (K(x_i, x_j))_{1 \leq i, j \leq N}, \quad (2.12)$$

which is not necessarily positive semidefinite. Then

$$(c_X, c_Y) := c_X^T A_{X,Y} c_Y \quad (2.13)$$

is an inner product on vectors  $c_X \in \mathbb{R}^{|X|,m}$  with  $P_X^T c_X = 0$  for a unisolvent set  $X$  and vectors  $c_Y \in \mathbb{R}^{|Y|,m}$  with  $P_Y c_Y = 0$  for a unisolvent set  $Y$ , if we define

$$A_{X,Y} := (K(x, y))_{x \in X, y \in Y},$$

and also the space (2.1) must now be replaced by

$$\mathcal{K}_X := \mathcal{P}_m^d + \left\{ \sum_{x_j \in X} c_j K(x_j, \cdot) : X \text{ is } \mathcal{P}_m^d \text{- unisolvent, } c \in M_{X,m} \right\}, \quad (2.14)$$

and this coincides with (2.1) for  $m = 0$ . The sum is direct if the kernel  $K$  is  $\mathcal{P}_m^d$ -conditionally positive semidefinite.

Note that these kernels are not reproducing kernels of a certain Hilbert space, because otherwise they would be unconditionally positive semidefinite. They do not have any direct connection to Hilbert theory, thus we have to start with kernels and then construct a related Hilbert space. Note that (unconditionally)positive (semi-)definite kernels are  $\mathcal{P}_m^d$ -conditionally positive (semi-)definite for all finite-dimensional spaces  $\mathcal{P}_m^d$ .

Now we can modify the assumption of the interpolation by scattered data interpolation by adding certain polynomials to the space  $\mathcal{K}_X$ . Let us again get a fixed  $\mathcal{P}_m^d$ -unisolvent set  $X = \{x_1, \dots, x_N\}$  in  $\Omega$  for a  $\mathcal{P}_m^d$ -conditionally positive kernel  $K$ . Thus  $s_f(y)$ , the interpolant on  $y \in \Omega$  is now assumed to be of the form

$$s_f(y) = \sum_{j=1}^N c_j K(x_j, y) + \sum_{k=1}^Q b_k p_k(y), \quad \forall y \in \Omega$$

while enforcing the interpolation condition  $s_f(x_j) = f_j$ ,  $j = 1, \dots, N$ . This leads to a system of  $N$  linear equations in  $N + Q$  unknowns. One usually just adds the moment conditions, thus leading to a well-posed interpolation system for value data  $f_1, \dots, f_N$  in  $X$  as

$$\begin{cases} s_f(x_k) = \sum_{j=1}^N c_j K(x_j, x_k) + \sum_{m=1}^Q b_m p_m(x_k) = f(x_k), & 1 \leq k \leq N, \\ \sum_{j=1}^N c_j p_n(x_j) + 0 = 0, & 1 \leq n \leq Q. \end{cases} \quad (2.15)$$

By standard and simple arguments, the matrix

$$\mathcal{A}_X := \begin{pmatrix} A_{X,X} & P_X \\ P_X^T & 0 \end{pmatrix}, \quad (2.16)$$

of this system. If  $K$  is  $\mathcal{P}_m^d$ -conditionally positive definite, and  $X$  is a  $\mathcal{P}_m^d$ -unisolvent set this system is uniquely solvable. However, it should be noted that if  $K$  is  $\mathcal{P}_m^d$ -conditionally positive semi-definite, the system is solvable with  $c_X = 0$ , if the  $f(x_j)$  come from functions in  $\mathcal{P}_m^d$  see proof in [23]. Using coefficient vectors  $c_X = (c_1 \dots, c_N)^T \in \mathbb{K}^N$  and  $b_X = (b_1 \dots, b_Q)^T \in \mathbb{K}^Q$ , the system of linear equations for interpolation of data  $y_X := (y_1, \dots, y_N)^T \in \mathbb{R}^N$  on  $X$  can be carried out by solving

$$\begin{pmatrix} A_{XX} & P_X \\ P_X^T & 0 \end{pmatrix} \cdot \begin{pmatrix} c_X \\ b_X \end{pmatrix} = \begin{pmatrix} y_X \\ 0 \end{pmatrix},$$

where  $c_X$  automatically satisfies the moment conditions. The interpolant is

$$s_{X,c_X,b_X}(x) := K_X(x)c_X + p_X(x)b_X, \quad (2.17)$$

with the row-vector notation

$$\begin{aligned} K_X(x) &:= (K(x, x_1), \dots, K(x, x_N)) \in \mathbb{R}^{|X|}, \\ p_X(x) &:= (p_1(x), \dots, p_Q(x)) \in \mathbb{R}^Q, \end{aligned}$$

for all  $x \in \Omega$ . We denote the span of these functions as  $\mathcal{K}_X$ , but note that the coefficient vectors  $c_X$  have to satisfy moment conditions.

### 2.3.1 Conditional Lagrange Basis

We proceed toward the conditional Lagrange basis, on a unisolvent set  $X$  by solving the system

$$\begin{cases} \sum_{j=1}^N K(x_k, x_j) u_j(x) + \sum_{m=1}^Q b_m(x) p_m(x_k) = K(x_k, x), & 1 \leq k \leq N \\ \sum_{j=1}^Q u_j(x) p_n(x_j) + 0 = p_n(x), & 1 \leq n \leq Q. \end{cases}$$

The coefficient matrix is the same as in (2.15). Then the system is also solvable, if the kernel is definite, and then the functions  $u_j$  are a Lagrange basis on  $X_Q$  with  $u_j(x_k) = \delta_{j,k}$ ,  $1 \leq j, k \leq N$ . Consequently, the interpolation for a  $f$  from  $\mathcal{K}_X$  has a solution of the form

$$s(y) = \sum_{j=1}^N u_j(y) f(x_j).$$

The conditional Lagrange basis will occur later in different circumstances.

### 2.3.2 Reduced Kernel

We describe a way to transit from a  $\mathcal{P}_m^d$ -conditionally positive (semi-)definite Kernel to an unconditionally positive (semi-)definite kernel. Define the Lagrange basis  $\pi_1, \dots, \pi_Q$  on a minimal  $\mathcal{P}_m^d$ -unisolvent set  $\Xi \subset \Omega$  with  $|\Xi| = Q$ . Then we can reproduce all  $p \in \mathcal{P}_m^d$ , via

$$p(x) = \sum_{m=1}^Q p(\xi_m) \pi_m(x), \quad \forall x \in \Omega, \quad p \in \mathcal{P}_m^d.$$

This defines a projector

$$\Pi_{\Xi}(f)(x) := \sum_{m=1}^Q f(\xi_m) \pi_m(x), \quad \forall x \in \Omega. \quad (2.18)$$

**Definition 2.3.3.** Assume a fixed  $\mathcal{P}_m^d$ -unisolvent set  $\Xi$ , define the **Reduced Kernel**

$$\begin{aligned} \tilde{K}(x, y) &:= K(x, y) - (\Pi_{\Xi} K(x, \cdot))(y) - (\Pi_{\Xi} K(\cdot, y))(x) \\ &+ ((\Pi_{\Xi} K(x, \cdot))(y) \cdot (\Pi_{\Xi} K(\cdot, y))(x)), \quad \forall x, y \in \Omega. \end{aligned}$$

**Theorem 2.3.1.** *The reduced kernel  $\tilde{K}$  is a symmetric and unconditionally positive semi-definite kernel on  $\Omega$ , which vanishes on  $\Omega \setminus \Xi$ , which vanishes on  $\Xi$ .*

The quadratic forms for  $K$  and  $\tilde{K}$  with moment conditions are the same

$$\sum_{j,k=1}^N a_j a_k \tilde{K}(x_j, x_k) = \sum_{j,k=1}^N a_j a_k K(x_j, x_k).$$

Also  $\tilde{K}$  is a reproducing kernel of the Hilbert space  $\tilde{\mathcal{K}}$  of functions  $f$  vanishing on  $\Xi$ . Each function  $\tilde{f}(x) \in \tilde{\mathcal{K}}$  has the reproduction formula i.e.

$$\tilde{f}(x) = (\tilde{f}, \tilde{K}(x, \cdot))_{\tilde{\mathcal{K}}}, \quad \text{for all } x \in \Omega, \tilde{f} \in \tilde{\mathcal{K}}.$$

Since all elements of  $\mathcal{P}_m^d$  are represented by their value on  $\Xi$ , and also we know that the functions of  $\tilde{\mathcal{K}}$  vanish on  $\Xi$ , it is reasonable to define the direct sum

$$\mathcal{K} := \mathcal{P}_m^d + \tilde{\mathcal{K}}$$

to be the native space for a  $\mathcal{P}_m^d$ -conditionally positive semidefinite kernel  $K$ . For all  $f, g \in \mathcal{K}$  we can define

$$(f, g)_{\mathcal{K}} := (f - \Pi_{\Xi} f, g - \Pi_{\Xi} g)_{\tilde{\mathcal{K}}},$$

to get a semi-inner product on  $\mathcal{K}$ . Then the interpolation on  $X$  can be written as

$$s := (\Pi_{\Xi})(f) + s_0,$$

$$s_0(x) := \sum_{j=1}^N \alpha_j \tilde{K}(x_j, x),$$

where  $s_0$  is the interpolant on data  $(f - \Pi_{\Xi} f)(x_i), 1 \leq i \leq N$ , with

$$s_0(\xi_j) = 0, \quad j = 1, \dots, Q.$$

We just use the reduced kernel on the point set  $Y := X \setminus \Xi$  for data  $f - \Pi_{\Xi}(f)$ , then add the interpolant in  $\mathcal{P}_m^d$  to  $\Pi_{\Xi}(f)$ . Using the reduced kernel this way means to use a kernel matrix

$$\tilde{A} := (\tilde{K}(x_j, x_k))_{Q+1 \leq j, k \leq N},$$

and to solve

$$\tilde{A} \tilde{c} = \tilde{f},$$

with  $Y$  and  $\tilde{f} = (f - \Pi_{\Xi} f)|_Y \in \mathbb{R}^{N-Q}$ . Another way to proceed is to split the matrix of (2.15) into

$$\begin{pmatrix} A_{\Xi, \Xi} & A_{\Xi, Y} & I_{Q \times Q} \\ A_{\Xi, Y}^T & A_{Y, Y} & B \\ I & B^T & 0_{Q \times Q} \end{pmatrix} \cdot \begin{pmatrix} c_{\Xi} \\ c_Y \\ b \end{pmatrix} = \begin{pmatrix} f_{\Xi} \\ f_Y \\ 0 \end{pmatrix},$$

in block form  $(Q, N - Q, Q)$ , since the value matrix of the Lagrange basis consists of  $I_{Q \times Q}$  for the point  $\Xi$  and a matrix  $B := (p_j(x_k))_{Q+1 \leq k \leq N, 1 \leq j \leq Q}$  for the points  $Y = X \setminus \Xi$ . Finally, this way of blocking can be helpful to investigate bases which partly obey a Lagrange rule.

In line with the previously discussed structure of the reduced kernel, one can add polynomials to the reduced kernel to get a particular unconditionally positive definite kernel. If we fix a unisolvent set  $\Xi$ , and  $K$  is a  $\mathcal{P}_m^d$ -conditionally positive semidefinite kernel on  $\Omega$ , then

**Definition 2.3.4.** *Given the Lagrange basis  $\pi_1, \dots, \pi_Q$  of  $\mathcal{P}_m^d$ , for all  $x, y \in \Omega$ , define the **Extended Kernel***

$$K^\dagger(x, y) := \tilde{K}(x, y) + \sum_{m=1}^Q \pi_m(x)\pi_m(y).$$

**Theorem 2.3.2.** *The kernel  $K^\dagger$  is a symmetric and unconditionally positive semi-definite kernel on  $\Omega$ , whose native space  $\mathcal{K}^\dagger$  coincides with the native space  $\mathcal{K}$  as a vector space.  $K^\dagger$  is definite if  $K$  is definite.*

The new inner product on  $\mathcal{K}^\dagger$ , for all  $f, g \in \mathcal{K} = \mathcal{P}_m^d + \tilde{\mathcal{K}}$  and the projector defined in (2.18), is

$$(f, g)_{\mathcal{K}^\dagger} := \sum_{m=1}^Q f(\xi_m)\overline{g(\xi_m)} + (f - \Pi_\Xi(f), g - \Pi_\Xi(g))_{\tilde{\mathcal{K}}}.$$

If satisfies the reproduction equation

$$f(x) = (f, K^\dagger(x, \cdot))_{\mathcal{K}^\dagger}, \quad \text{for all } f \in \mathcal{K}, x \in \Omega.$$

We extend the reduced kernel in order to get an extended kernel with respect to an inner product on the whole space. The extended kernel turns  $\mathcal{P}_m^d$  into a subspace of  $\mathcal{K}$  that is orthogonal to  $\tilde{\mathcal{K}}$  in the new inner product. We refer to [23] for more details.

# Bases for Kernel Based Spaces

---

This chapter is based on the paper [14]. It is observed in [8] that standard bases of kernel translates are badly conditioned while the interpolation itself is not unstable in function space. The first sections of chapter come up with definitions of useful bases. All data-dependent bases turn out to be defined via a factorization of the kernel matrix defined by these data, hence a discussion of various matrix factorizations (e.g. Cholesky,  $QR$ , SVD) provides a variety of different bases with different properties. Special emphasis is given to duality, stability, orthogonality, and computational efficiency.

Then we describe the “Newton” basis arising from a pivoted Cholesky factorization, it turns out to be stable and recursively computable while being orthonormal in the “native” Hilbert space of the kernel. Throughout this Chapter kernels will be unconditionally positive definite.

## 3.1 General Data-dependent Bases

Let fix a set  $X = \{x_1, \dots, x_N\} \subset \Omega$ , the space  $\mathcal{K}_X$  of (2.1) and the kernel matrix  $A = (K(x_j, x_k))_{1 \leq j, k \leq N}$ . Any basis  $u_1, \dots, u_N$  of  $\mathcal{K}_X$  has a general expansion

$$u_k(x) = \sum_{j=1}^N K(x, x_j) c_{jk}, \quad 1 \leq k \leq N \quad (3.1)$$

in terms of the basis of translates. In vector-matrix notation, it can be arranged into a *row* vector

$$U(x) := (u_1(x), \dots, u_N(x)) \in \mathbb{R}^N,$$

and it can be expressed by the basis of translates

$$T(x) := (K(x, x_1), \dots, K(x, x_N))$$

by a *coefficient* or *construction* matrix  $C_U$  via

$$U(x) = T(X) \cdot C_U, \quad (3.2)$$

where  $C_U$  is the coefficient matrix

$$C_U = (c_{k,m})_{k,m=1}^N,$$

with the “Value Matrix” on  $X$

$$V_U = \begin{pmatrix} u_1(x_1) & \dots & u_N(x_1) \\ \vdots & \ddots & \vdots \\ u_1(x_N) & \dots & u_N(x_N) \end{pmatrix}. \quad (3.3)$$

The set  $\mathcal{B}$  of all possible data-dependent bases is the bijective image of the group  $GL(n, \mathbb{R})$  of all nonsingular real  $N \times N$  matrices under the map  $C \mapsto T(x) \cdot C$ . This means that one has a composition  $\circ$  of two bases  $U$  and  $V$  via

$$(U \circ V)(x) := T(x) \cdot C_U \cdot C_V = T(x) \cdot C_{U \circ V}$$

and thus one can define the *inverse* of a basis. This concept, but with a “coordinate space” being fixed instead of a fixed basis  $T$ , was introduced and exploited by C. de Boor in [7]. Here, we just note that the full set of possible bases  $U$  can be parametrized by arbitrary matrices  $C_U \in GL(n, \mathbb{R})$ . Thus we shall express formulae for features of bases  $U$  mainly in terms of  $C_U$ , but there are other parametrizations as well, as we shall see.

The “Evaluation” operator  $E$  based on the set  $X$  will map functions  $f$  into columns

$$E(f) := (f(x_1), \dots, f(x_N))^T \in \mathbb{R}^N,$$

for which then we can rewrite kernel matrix  $A$  by

$$E(T) = (K(x_i, x_j))_{1 \leq j, k \leq n} = A.$$

By equation (3.3), for a general basis  $U$  we can form the *value* matrix

$$V_U = E(U) = (u_k(x_j))_{1 \leq j, k \leq N} = E(T)C_U = A \cdot C_U. \quad (3.4)$$

A very similar way to use columns and rows, and the connection to duality we shall use later are nicely described already in [7]. From the identity

$$V_U = E(U) = E(T) \cdot C_U = A \cdot C_U,$$

we immediately get that the basis  $U(x)$  is connected to factorizations of  $A$ . In particular, the kernel matrix can be decomposed to the value matrix  $V_U$  and inverse of the coefficient matrix.

**Theorem 3.1.1.** *Any data-dependent basis  $U$  arises from a factorization*

$$A = V_U \cdot C_U^{-1} \quad (3.5)$$

*of the kernel matrix  $A$  into the value matrix  $V_U = A \cdot C_U$  and the inverse construction matrix  $C_U^{-1}$  of the basis.*

The *Gramian*  $G_U$  of a general basis  $U$  comes out to be

$$\begin{aligned}
G_U &:= ((u_j, u_k)_{\mathcal{K}})_{1 \leq j, k \leq N} \\
&= (\sum_k c_{k,j} K(\cdot, x_k), \sum_l c_{l,m} K(\cdot, x_l))_{\mathcal{K}} \\
&= \sum_k \sum_l c_{k,j} c_{l,m} K(x_k, x_l) \\
&= C_U^T \cdot A \cdot C_U.
\end{aligned} \tag{3.6}$$

These are the  $\mathcal{K}$ -inner products, but we also have discrete  $\ell_2(X)$  inner products forming a Gramian  $\Gamma_U$  via

$$\begin{aligned}
\Gamma_U &:= ((u_j, u_k)_{\ell_2(X)})_{1 \leq j, k \leq N} \\
&= \left( \sum_{n=1}^N u_j(x_n) u_k(x_n) \right)_{1 \leq j, k \leq N} \\
&= V_U^T \cdot V_U \\
&= C_U^T \cdot A^2 \cdot C_U
\end{aligned}$$

using (3.5).

Particularly, the factorization in (3.5) for the basis of translates

$$T(x) := (K(x, x_1), \dots, K(x, x_N))$$

is  $A = A \cdot I$ , where we use  $I$  to stand for the  $N \times N$  identity matrix, which is clearly the coefficient matrix  $C_T$  for the basis of translates.

We now look at the standard *Lagrange* basis  $L(x) := (L_1(x), \dots, L_N(x))$ , whose elements are data-dependent and satisfy

$$L_j(x_k) = \delta_{jk}, \quad 1 \leq j \leq k.$$

Clearly,  $V_L = I = \Gamma_L$  and by (3.5) we get  $C_L = A^{-1}$ . The  $\mathcal{K}$ -Gramian is  $G_L = A^{-1}$  by (3.6). Consequently, via the basis of translates and the Lagrange basis, we have  $L(x)A = T(x)$ , then

**Theorem 3.1.2.** *Assume  $L(x)$  to be the Lagrange basis, then for all data-dependent bases  $U(x)$*

$$U(x) = L(x)V_U,$$

where  $V_U$  is the value matrix.

**Proof:** By the definition of the Lagrange basis and  $U(x) = T(x)C_U$ , we have

$$U(x) = \underbrace{T(x)A^{-1}}_{L(x)} \underbrace{AC_U}_{V_U} = L(x)V_U. \quad \square$$

The value matrix  $V_U = (u_k(x_i))_{i,k}$  of values of the general basis expresses the correspondence between the general basis and the Lagrange basis.

Note that (3.5) also shows that we could as well parametrize the set of bases  $U$  via the value matrices  $V_U$  using  $C_U = A^{-1} \cdot V_U$  to come back to the parametrization via  $C_U$ . But for sorting out special classes of bases we need more. Multivariate interpolation on a set of  $X$  of scattered data locations for values  $E(f)$  of functions  $f$  with a symmetric positive definite and continuous kernel  $K$  is calculated via the linear system

$$A\alpha_X = E(f), \quad (3.7)$$

where  $\alpha_X := (\alpha_1, \dots, \alpha_N)^T$  and  $A = (K(x_j, x_k))_{1 \leq j, k \leq N}$ . The interpolant  $s_f \in \mathcal{K}_X$  can be written as

$$s_f(x) = \sum_{j=1}^N \alpha_j K(x, x_j),$$

this means with  $T(x) := (K(x, x_1), \dots, K(x, x_N))^T$ , and the system (3.7)

$$s_f(x) = T(x)\alpha.$$

This is well-known, but also follows immediately from

$$E(s_f) = E(T)\alpha = A\alpha = E(f),$$

using our notation. For general bases, the interpolant takes the form

$$\begin{aligned} s_f(x) &:= T(x) \cdot A^{-1} \cdot E(f) \\ &= U(x) \cdot \underbrace{C_U^{-1} \cdot A^{-1} \cdot E(f)}_{=:\Lambda_U(f)} \\ &= \sum_{j=1}^N u_j(x) \lambda_j(f) \end{aligned} \quad (3.8)$$

with a column vector

$$\Lambda_U(f) := (\lambda_1(f), \dots, \lambda_N(f))^T = C_U^{-1} \cdot A^{-1} \cdot E(f) \quad (3.9)$$

of linear *functionals*. They are from the span of the point evaluation functionals  $\delta_{x_1}(f), \dots, \delta_{x_N}(f)$ , and they can be composed from them by

$$\Lambda_U(f) = \underbrace{C_U^{-1} \cdot A^{-1}}_{=:\Delta_U} \cdot E(f) = \Delta_U \cdot E(f)$$

as a matrix operation with

$$\Delta_U = C_U^{-1} \cdot A^{-1} = V_U^{-1}.$$

We have chosen the notation  $\Delta_U$  here, because the action of the matrix is like forming divided differences from function values. For the basis  $T$  of

translates, we have  $\Delta_T = V_T^{-1} = A^{-1}$ . Note that we could parametrize bases  $U$  also via  $\Delta_U$ .

The evaluation of an interpolant  $s_f$  at some point  $x$  via (3.8) can be unstable, if either the  $u_j(x)$  or the  $\lambda_j(f)$  or both are large, cancelling finally when forming the result  $s_f(x)$ . This regularly occurs for the basis  $T$  of translates, because the coefficient vector  $A^{-1}E(f)$  tends to have huge absolute values of opposite sign. A measure for the stability of the evaluation of  $s_f(x)$  thus is the Hölder–Minkowski bound

$$|s_f(x)| \leq \|U(x)\|_p \|\Lambda_U(f)\|_q \text{ for all } f \in \mathcal{K}, x \in \Omega \quad (3.10)$$

with  $1/p + 1/q = 1$ . We shall have to look at both factors in what follows.

In [8], there is an analysis of the stability of the Lagrange basis along this line, proving that the right-hand side is well-behaving. This implies that interpolation is not unstable in function space, though the calculation of coefficients in the basis of translates is unstable. Consequently, we have to look for other bases which allow good bounds in (3.10).

## 3.2 Dual Basis

The linear functionals in  $\Lambda_U$  are in some sense *dual* to the basis  $U$ , but we define duality slightly differently:

**Definition 3.2.1.** *The dual basis to a basis  $U$  is the basis  $U^*$  of the Riesz representers of the functionals of  $\Lambda_U$ .*

**Theorem 3.2.1.** *There is correspondence between the dual basis  $U^*$  and bases of translates  $T(x)$ ,*

$$U^*(x) = T(x)(V_U^{-1})^T, \quad (3.11)$$

where  $V_U$  is the value matrix of the general basis  $U$ .

**Proof:** The dual basis functions  $u_j^*$  are defined by

$$\lambda_j(f) = (u_j^*, f)_{\mathcal{K}}, \text{ for all } f \in \mathcal{K}, 1 \leq j \leq N. \quad (3.12)$$

being the Riesz representers of the  $\lambda_j$ . Thus

$$\begin{aligned} u_j^*(x) &= (u_j^*, K(x, \cdot))_{\mathcal{K}} = \lambda_j(K(x, \cdot)), \quad 1 \leq j \leq N, \text{ for all } x \in \Omega, \\ U^*(x) &= \Lambda_U^T(K(x, \cdot)) \\ &= C_U^{-1} \cdot A^{-1} \cdot E(K(x, \cdot))^T \\ &= V_U^{-1} \cdot E(K(x, \cdot))^T \\ &= T(x)(V_U^{-1})^T. \quad \square \end{aligned} \quad (3.13)$$

Given some basis  $U$ , we now have to find the value matrix  $V_{U^*}$  and the construction matrix  $C_{U^*}$  of the dual basis  $U^*$ .

**Theorem 3.2.2.** *The dual basis  $U^*$  to a data-dependent basis  $U$  satisfies*

$$\begin{aligned} V_{U^*} &= (C_U^T)^{-1}, \\ C_{U^*} &= (V_U^T)^{-1}, \\ A &= V_U \cdot V_{U^*}^T, \\ K(x_j, x_k) &= \sum_{m=1}^N u_m(x_k) u_m^*(x_j), \quad 1 \leq j, k \leq N, \\ (u_j, u_k^*)_{\mathcal{K}} &= \delta_{jk}, \quad 1 \leq j, k \leq N. \end{aligned}$$

**Proof:** (3.13) prove  $C_{U^*} = (V_U^T)^{-1}$ , for all  $x \in \Omega$ . Then (3.5) yields

$$\begin{aligned} V_{U^*} &= A \cdot C_{U^*} \\ &= A \cdot (V_U^T)^{-1} \\ &= A \cdot ((A \cdot C_U)^T)^{-1} \\ &= (C_U^T)^{-1}, \end{aligned}$$

and the next three relations are easy consequences. Finally,

$$\begin{aligned} (u_j, u_k^*)_{\mathcal{K}} &= \lambda_k(u_j) \\ &= e_k^T \Lambda_U(u_j) \\ &= e_k^T \Delta_U \cdot E(u_j) \\ &= e_k^T \Delta_U \cdot E(Ue_j) \\ &= e_k^T \Delta_U \cdot E(U)e_j \\ &= e_k^T C_U^{-1} A^{-1} V_U e_j \\ &= e_k^T e_j, \quad 1 \leq j, k \leq N \end{aligned}$$

proves the last assertion and shows that the functionals of  $\Lambda_U$  always are a biorthogonal basis with respect to  $U$ .  $\square$

**Corollary 3.2.1.** *Assume  $U^{**}$  is double dual to  $U$  then*

$$U^{**} = U.$$

**Proof:** By the above properties for the dual basis it is just necessary to show that

$$C_{U^{**}} = C_U.$$

By the definition of the coefficient matrix

$$C_{U^{**}} = (V_{U^*}^{-1})^T = A^{-1} V_U = C_U. \quad \square$$

Therefore, the Gramian matrix for the dual basis is

$$G_{U^*} = (u_j^*, u_k^*)_{\mathcal{K}} = C_{U^*}^T A C_{U^*} = V_U^{-1} A (V_U^{-1})^T.$$

The transition from a basis  $U$  to its dual follows a simple rule-of-thumb. Starting with (3.5), we take the transpose and re-interpret this in the sense

of Theorem (3.1.1) as the product of a dual value matrix times the inverse of a dual construction matrix:

$$\begin{aligned} A &= V_U \cdot C_U^{-1} \\ = A^T &= (C_U^{-1})^T \cdot V_U^T \\ &= V_{U^*} \cdot C_{U^*}^{-1}. \end{aligned}$$

For later use in stability considerations along the lines of (3.10), we use (3.8) to get

$$\|s_f\|_{\mathcal{K}}^2 = \Lambda^T(f) G_U \Lambda(f)$$

with the  $\mathcal{K}$ -Gramian  $G_U$ . By standard eigenvalue manipulations and the inequality  $\|s_f\|_{\mathcal{K}} \leq \|f\|_{\mathcal{K}}$ , this implies

**Theorem 3.2.3.** *For all  $f \in \mathcal{K}$  and all data-dependent bases  $U$ ,*

$$\|\Lambda(f)\|_2^2 \leq \|f\|_{\mathcal{K}}^2 \rho(G_U^{-1})$$

with  $\rho$  being the spectral radius.

Note that this bound is particularly bad for the basis  $T$  of translates with  $G_T = A$ . By Theorem (3.2.2) or by the above rule-of-thumb in the form

$$A = A \cdot I = V_T \cdot C_T^{-1} = A^T = I \cdot A^T = V_{T^*} \cdot C_{T^*}^{-1}$$

we get

**Theorem 3.2.4.** *The Lagrange basis  $L$  and the basis  $T$  of translates are a dual pair.*

**Proof:** The value matrix of the Lagrange basis is the identity matrix  $V_U = I$ , so by the dual property, simply

$$U^*(x) = T(x)(V_U^{-1})^T = T(x).$$

For the value matrix of the basis of translates  $V_K(x) = T(x)$ , compute

$$T^*(x) = T(x)(V_K^{-1})^T = L(x). \quad \square$$

Another way to see this duality is by noting that the functionals  $\lambda_j(f) = \delta_{x_j}(f)$  of the Lagrange basis are the Riesz representers of the kernel translates  $K(\cdot, x_j)$ .

### 3.3 Power Function

We use the well-known *power function* in (2.7) and express it for general bases. By the formula (2.5) and some standard algebraic manipulations within our formalism, we can express it in terms of a general basis as

$$\begin{aligned}
P^2(x) &= K(x, x) - 2 \sum_{j=1}^N K(x, x_j) L_j(x) + \sum_{j,k=1}^N K(x_k, x_j) L_j(x) L_k(x) \\
&= K(x, x) - \sum_{j=1}^N K(x, x_j) L_j(x) \\
&= K(x, x) - T(x) \cdot L^T(x) \\
&= K(x, x) - T(x) \cdot A^{-1} \cdot T^T(x) \\
&= K(x, x) - U(x) \cdot C_U^{-1} \cdot A^{-1} \cdot (C_U^{-1})^T \cdot U^T(x) \\
&= K(x, x) - U(x) \cdot G_U^{-1} \cdot U^T(x).
\end{aligned} \tag{3.14}$$

This identity has several consequences. First, due to positive definiteness of the  $\mathcal{K}$ -Gramian  $G_U$ , this yields bounds

$$0 \leq U(x) \cdot G_U^{-1} \cdot U^T(x) = K(x, x) - P^2(x) \leq K(x, x)$$

for the pointwise behavior of the general basis  $U$ . By standard eigenvalue bounds, this implies a bound that is useful for (3.10).

**Theorem 3.3.1.** *For arbitrary data-dependent bases  $U$ , we have*

$$\|U(x)\|_2^2 \leq K(x, x) \rho(G_U) \text{ for all } x \in \Omega$$

with the spectral radius  $\rho(G_U)$  of the  $\mathcal{K}$ -Gramian. Furthermore, the stability bound (3.10) for  $p = q = 2$  is

$$|s_f(x)|^2 \leq \|f\|_{\mathcal{K}}^2 K(x, x) \text{ cond}_2(G_U) \text{ for all } f \in \mathcal{K}, x \in \Omega, \tag{3.15}$$

where  $\text{cond}_2(G_U)$  is the condition number of the  $\mathcal{K}$ -Gramian with respect to the Euclidean norm.  $\square$

### 3.4 $\mathcal{K}$ -Orthonormal Bases

We now focus on constructing the  $\mathcal{K}$ -orthonormal bases. Using the factorization technique of the kernel matrix  $A = V_U C_U^{-1}$  to produce the value matrix and the coefficient matrix of all data-dependent bases, we look for bases  $U$  with the  $\mathcal{K}$ -Gramian matrix

$$G_U = C_U^T A C_U = I,$$

i.e. for  $\mathcal{K}$ -orthonormal bases. This obviously is equivalent to

$$A = (C_U^{-1})^T C_U^{-1} = I.$$

**Theorem 3.4.1.** *Each data-dependent  $\mathcal{K}$ -orthonormal basis  $U$  arises from a decomposition*

$$A = B^T \cdot B \text{ with } B = C_U^{-1}, V_U = B^T = (C_U^{-1})^T. \quad (3.16)$$

*Among all data-dependent bases, the  $\mathcal{K}$ -orthonormal bases are exactly those which are self-dual.*

**Proof:** Clearly, above discussion proved the first assertion. By Theorem (3.2.2), all  $\mathcal{K}$ -orthonormal bases are self-dual. Conversely, if  $U$  is a self-dual basis, then

$$\begin{aligned} V_U^{-1} &= C_U^T, \\ A &= V_U \cdot (V_U^{-1})^T, \end{aligned}$$

and the second assertion follows from  $C_U = (V_U^{-1})^T$  to get an orthonormal basis in the native space. By above theorem, for any  $\mathcal{K}$ -orthonormal basis  $U$ ,

$$U^*(x) = T(x)C_{U^*} = T(x)(V_U^{-1})^T = T(x)C_U = U(x). \quad \square$$

There are two important special cases.

The Cholesky decomposition  $A = L \cdot L^T$  with a nonsingular lower triangular matrix  $L$  leads to the *Newton* basis  $N$  treated in [18] with a different normalization. It can be recursively calculated and has the property  $N_j(x_k) = 0$ ,  $1 \leq k < j \leq N$  like the basis of functions

$$N_j(x) = \prod_{1 \leq k < j} (x - x_k), \quad 1 \leq j \leq N$$

in Newton's formula for polynomial interpolation.

The other case is induced by *singular value decomposition* (SVD) in the form  $A = Q^T \cdot \Sigma^2 \cdot Q$  with an orthogonal matrix  $Q$  and a diagonal matrix  $\Sigma$  having the eigenvalues of  $A$  on its diagonal. This *SVD basis*  $S$  satisfies

$$B = \Sigma \cdot Q, \quad C_S = Q^T \cdot \Sigma^{-1}, \quad V_S = Q^T \cdot \Sigma.$$

Before we analyze these special cases further, we prove slightly more than we had for Theorem (3.3.1).

**Theorem 3.4.2.** *For all  $\mathcal{K}$ -orthonormal bases  $U$ , the value of  $\|U(x)\|_2$  for fixed  $x \in \Omega$  is the same and bounded above by  $K(x, x)$  independent of the placement and number of data points. Dually, the value of  $\|\Lambda(f)\|_2$  for fixed  $f \in \mathcal{K}$  is the same for all  $\mathcal{K}$ -orthonormal bases and bounded above by  $\|f\|_{\mathcal{K}}$  independent of the placement and number of data points.*

**Proof:** Equation (3.14) yields

$$\sum_{j=1}^N u_j^2(x) = \|U(x)\|_2^2 = K(x, x) - P^2(x) \leq K(x, x) \text{ for all } x \in \Omega,$$

and this proves the first assertion, because the power function is basis-independent. Writing an interpolant  $s_f$  in the form (3.8) is an orthonormal representation. Thus, being basis-independent as well,

$$\|s_f\|_{\mathcal{K}}^2 = \sum_{j=1}^N \lambda_j^2(f) \leq \|f\|_{\mathcal{K}}^2 \quad (3.17)$$

proves the second.  $\square$

In conclusion, the divided differences corresponding to  $\mathcal{K}$ -orthonormal data-driven bases are orthonormal as functionals, and also  $\mathcal{K}$ -orthonormal bases will lead to stable results in function space even for nearly-coalescing data points, provided that the data come from a function in the native space. The functionals act like divided differences and have unity norm in the dual of the native space, irrespective of the placement of data points.

From Theorem (3.3.1) we get the bound

$$|s_f(x)|^2 \leq K(x, x) \|f\|_{\mathcal{K}}^2 \text{ for all } f \in \mathcal{K}, x \in \Omega \quad (3.18)$$

for all  $\mathcal{K}$ -orthonormal bases  $U$ , implying that the evaluation of the interpolant is stable provided that the  $U_j(x)$  and  $\lambda_j(f)$  can be evaluated stably. The basic equations for these are

$$U(x) = T(x)B^{-1} \text{ for all } x \in \Omega, \Lambda_U(f) = (B^{-1})^T E(f) \text{ for all } f \in \mathcal{K},$$

and we see that in both cases the matrix  $B^{-1}$  is involved, or a system with coefficient matrix  $B$  has to be solved. For the condition of  $B$  we have

**Theorem 3.4.3.** *The general solution  $B$  of (3.16) is always of the form  $B = Q_1 \Sigma Q$  with an orthogonal matrix  $Q_1$ , when  $A = Q^T \Sigma^2 Q$  is a SVD of  $A$ . Thus the spectrum of  $A$  is factored by (3.16), and the spectral condition of  $B$  is the square root of the spectral condition of  $A$ .*

**Proof:** From  $A = Q^T \Sigma^2 Q = B^T B$  we get  $I = \Sigma^{-1} Q B^T B Q^T \Sigma^{-1}$ . Thus  $Q_1 := B Q^T \Sigma^{-1}$  is orthogonal and  $B = Q_1 \Sigma Q$ . The matrix  $B$  has a SVD with singular values being the square roots of those of  $A$ .  $\square$

Thus all  $\mathcal{K}$ -orthonormal bases divide the ill-conditioning of  $A$  fairly between the function and the functional part. Later, we shall consider special adaptive algorithms for the Newton case.

### 3.5 Discretely Orthonormal Bases

In line with all  $\mathcal{K}$ -orthonormal bases, the framework for discretely orthonormal bases comes out from the discrete Gramian matrix  $\Gamma_U$ , since

$$\Gamma_U = \left( \sum_m U_j(x_m) U_k(x_m) \right)_{j,k} = V_U^T V_U = C_U^T A^T A C_U.$$

In order to construct discretely orthonormal bases, the value matrix  $V_U$  should be the orthogonal matrix, so that  $V_U^T V_U = I$  becomes

$$\Gamma_U = (u_j, u_k)_{X, l_2} = \delta_{j,k}.$$

In other words

**Theorem 3.5.1.** *Each data-dependent discretely orthonormal basis arises from a decomposition*

$$A = Q \cdot B$$

with  $Q = V_U$  orthogonal and  $B = C_U^{-1} = Q^T \cdot A$ .

**Proof:** By the formula  $\Gamma_U = C_U^T \cdot A^T \cdot A \cdot C_U$  for the discrete Gramian  $\Gamma_U$ , and setting  $Q := A \cdot C_U$ , we see that  $\Gamma_U = I$  is equivalent to orthogonality of  $Q$ .  $\square$

This is a matrix factorization, the so-called QR decomposition,  $A = QR$ . Here  $R$  is upper triangular, while  $Q$  is orthogonal. A sequence of Householder reflections is applied to the columns of  $A$  to produce the matrix  $R$ , the  $j$ th column of  $R$  is a linear combination of the first  $j$  columns of  $A$ . Consequently, the elements of  $R$  below the main diagonal are zero. If we have the full factorization, so the full  $Q$  is an orthogonal matrix. If  $A$  is nonsingular, then this factorization is unique if we require that the diagonal elements of  $R$  are positive. A standard  $QR$  decomposition  $A = QR$  will lead to a basis we shall denote by  $O$  with  $C_O = R^{-1}$ ,  $V_O = Q$ . This is nothing else than Gram-Schmidt orthonormalization of the values of the translate basis  $T$  on  $X$ . Thus the QR basis is

$$U^O(x) = (R^{-1})^T T(x).$$

To continue it will be useful to find the definition of the basis respect to the Lagrange basis. Since we have  $AL(x) = T(x)$ , which implies  $L(x) = QU^O(x)$ , the QR basis are correspondent to the Lagrange basis by a unitary matrix  $Q$ , in addition  $U^O(x) = Q^T L(x)$ . That means we applied a sequence of Householder reflections on the Lagrange basis to compute the QR basis. The bases arising from QR decompositions are not orthogonal in native spaces. Since  $\|U^O(x)\|_2 = \|Q^T L(x)\|_2 = \|L(x)\|_2$ , for all  $x \in \Omega$ , implies that  $QR$  bases are bounded. The power function formula is

$$P^2(x) = K(x, x) - (U^O(x))^T R(x),$$

and the interpolant can be written as

$$s_f(x) = \sum_j (\Delta_j^O f)^T u_j^O(x),$$

where

$$\Delta_j^O f = e_j^T Q^T E(f).$$

Moreover, we have  $e_j^T \Delta(u_k) = \delta_{jk}$ ,  $0 \leq j, k \leq n$ . The orthogonality of the matrix  $Q$  implies

$$\|\Delta^O f\|_\infty \leq \|\Delta^O f\|_2 = \|Q^T E(f)\|_2 = \|E(f)\|_2,$$

thus

$$\begin{aligned} |s_f(x)| &= |\sum_j (\Delta_j^O f)^T u_j^O(x)| \\ &\leq \|\Delta^O f\|_2 \|U^O(x)\|_2 = \|E(f)\|_2 \|U^O(x)\|_2, \\ &= \|E(f)\|_2 \|L(x)\|_2. \end{aligned}$$

Thus the evaluation of  $s_{X,f}(x)$  via the QR basis is stable, provided that the divided differences and the QR bases can be stably calculated.

The second case comes from rescaling an SVD basis. In fact, any SVD basis  $A = Q^T \Sigma^2 Q$  can be split into  $A = Q \cdot B$  with  $B = \Sigma^2 Q$ . This makes the value matrix orthonormal, while the ill-conditioning is completely shifted into the construction matrix. Thus, if scaling is ignored, all SVD bases are both discretely and  $\mathcal{K}$ -orthogonal. The converse is also true:

**Theorem 3.5.2.** *All data-dependent bases which are discretely and  $\mathcal{K}$ -orthogonal are scaled SVD bases.*

**Proof:** Any such basis  $U$  can be rescaled to be  $\mathcal{K}$ -orthonormal. We then have  $A = B^T \cdot B$  with  $B = C_U^{-1}$  and  $\Gamma_U = C_U^T \cdot A^2 C_U = D^2$  with a nonsingular diagonal matrix  $D$ . This implies  $I = D^{-1} C_U^T A^T A C_U D^{-1}$  and that

$$Q := A \cdot C_U \cdot D^{-1} = B^T \cdot B \cdot C_U \cdot D^{-1} = (C_U^{-1})^T \cdot D^{-1}$$

is orthogonal. But then

$$\begin{aligned} A \cdot Q &= A \cdot A \cdot C_U \cdot D^{-1} \\ &= (C_U^{-1})^T D^{-1} \\ &= Q \cdot D^2 \end{aligned}$$

leads to the SVD of  $A = Q \cdot D^2 \cdot Q^T$  with  $B = C_U^{-1} = D \cdot Q^T$ .  $\square$

For any discretely orthonormal basis  $U$ , the  $\mathcal{K}$ -Gramian is

$$G_U = C_U^T \cdot A \cdot C_U = Q^T \cdot A^{-1} \cdot A \cdot A^{-1} \cdot Q = Q^T \cdot A^{-1} \cdot Q$$

and thus spectrally equivalent to  $A^{-1}$ . In view of Theorem (3.3.1), this is comparable to the Lagrange and the translates basis.

**Theorem 3.5.3.** *The duals of the discretely orthonormal bases arise from decompositions*

$$A = B \cdot Q$$

with  $Q$  orthogonal.

**Proof:** Following our rule-of-thumb for a discretely orthonormal basis  $U$ , we get

$$A = Q \cdot B = A^T = B^T \cdot Q^T = V_{U^*} \cdot C_{U^*}^{-1}, \quad V_{U^*} = B^T, \quad C_{U^*} = Q. \quad \square$$

These bases have orthogonal construction matrices instead of orthogonal value matrices. Again, a scaled SVD basis is a special case, and also the transpose  $A = R^T \cdot Q^T$  of a  $QR$  decomposition  $A = Q \cdot R$ . The  $\mathcal{K}$ -Gramians of these bases are of the form  $Q \cdot A \cdot Q^T$ , and thus again spectrally equivalent to the translate and Lagrange bases, as far as the spectral condition is concerned.

### 3.6 SVD Bases

Singular values are well conditioned, which means they are not extremely sensitive to perturbations in  $A$ . The sensitivity of a singular vector depends on how small the gap is between its singular value and any other singular value, the smaller the gap, the more sensitive the singular vector.

The singular value decomposition is useful for numerical determination of the rank of the matrix, and for solving linear least squares problems, especially when they are rank deficient, or nearly so.

Now we look at the basis coming out of a singular Value Decomposition of the kernel matrix. The eigenvalue problem for a symmetric matrix is considerably simpler and more stable with respect to round-off. Therefore, the eigenvectors have the convenient mathematical property of orthogonality and span the entire space of  $A$ . That is, the form a basis or minimum spanning set with

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

where the columns of  $Q$  form the set of orthonormal eigenvectors of  $A$ . Though their computation is rather involved, the SVD bases have some nice properties, as we have seen in the previous sections, in particular in Theorem (3.5.2). Going back to Section 2.2 and using an integration formula with well-distributed points and equal weights  $w_k = w > 0$ , we see that the discretized solution  $v_j^{(n)}$ ,  $1 \leq j, n \leq N$  of the eigenvalue problem (2.8) is related to the eigenvalue problem of  $A$  itself, and thus is a scaled SVD basis. Thus we can expect that SVD bases for large and well-placed data point sets are approximations of the data-independent eigenfunction basis. For kernels with rapidly decaying eigenvalues, one has to expect numerical rank loss in the kernel matrix  $A$ , but the SVD is the best known way to control this.

In theory, the SVD basis  $S$  is given by  $S(x) = T(x) \cdot Q^T \cdot \Sigma^{-1}$ , while the value matrix is  $V_S = Q^T \cdot \Sigma$ . Moreover  $V_S$  is the correspondence between the Lagrange basis and the SVD bases where  $L(x) = S(x)\Sigma^{-1} \cdot Q$ . Following the power function

$$P_X^2(x) = K(x, x) - \sum_{j=1}^n s_j^2(x)$$

as for all  $\mathcal{K}$ -orthonormal bases, this is a nice property but we should be sure that the  $s_j(x)$  are not too big when  $\sigma_j$  are too small. They are discretely bounded but not in general. The representation of the interpolant in terms of the SVD basis is

$$s_f(x) = S(x)V_S^{-1}E(f) = S(x)\Sigma^{-1} \cdot QE(f).$$

Then we get another version of generalized divided differences by  $\Delta^s f = V_S^{-1}E(f)$ . We see that the Riesz representers of the generalized divided difference  $\Delta^s f$  are given by the SVD basis, and the divided differences form a dual orthogonal basis. Consequently, we get

$$\|\Sigma\Delta_s f\|_2 = \|E(f)\|_2,$$

which provides a balance between singular values and generalized divided differences. If singular values  $\sigma_j^2$  are sorted to decrease with increasing  $j$ , the columns of the value matrix have decreasing norms  $\|V_U e_j\|_2 = \sigma_j$  for increasing  $j$ . The usual Tychonov regularization will replace small  $\sigma_j < \epsilon$  by zero, thus making the basis shorter. In that case, the numerical result of the reconstruction of  $f$  from given data  $E(f)$  will be a non-interpolatory projection of  $f$  into the span of the selected SVD basis functions, and if these are corresponding to the eigenfunctions with large eigenvalues, the result will be an accurate and stable reproduction of the projection of  $f$  into the span of these eigenfunctions. If, however,  $f$  has a large projection onto higher eigenfunctions, this will lead to unavoidable errors, but these often look like noise, while the numerical solution looks smooth. This makes the SVD useful for a lot of applications where deterministic worst-case error bounds make no sense, and in particular where the data are noisy anyway and exact interpolation is not desirable.

### 3.7 Newton Basis

A square symmetric and positive matrix  $A$  has a special, more efficient, triangular factorization, so-called Cholesky decomposition. Cholesky decomposition is about a factor of two faster than alternative methods for solving linear equations. Instead of seeking arbitrary lower and upper triangular factors  $L$  and  $U$ , Cholesky decomposition constructs a lower triangular matrix  $L$  whose

transpose  $L^T$  can itself serve as the upper triangular part. The Cholesky decomposition is roughly twice as efficient as the  $LU$  decomposition for solving systems of linear equations. A Cholesky decomposition  $A = L \cdot L^T$  with a nonsingular lower triangular matrix  $L$  leads to the Newton basis  $N$  with  $N(x) = T(x) \cdot C_N = T(x) \cdot (L^T)^{-1}$  and value matrix  $V_N = L$ . Given a symmetric positive definite matrix  $A$ , the Cholesky decomposition yields a lower triangular matrix  $L$  with strictly positive diagonal entries. This basis was already treated in [18], but with a different normalization that concealed its  $\mathcal{K}$ -orthonormality. Theorem (3.4.2) and the stability bound (3.18) are not in [18], but proved here for general  $\mathcal{K}$ -orthonormal bases. The identity  $V_N = L$  implies

$$\begin{aligned} N_k(x_j) &= 0, \quad 1 \leq j < k \leq N \\ N_k(x_k) &> 0, \quad 1 \leq k \leq N. \end{aligned}$$

The connection to the Lagrange basis is  $N(x) = L(x)V_N$ , hence, the power function is

$$P^2(x) = K(x, x) - V_N V_N^T = K(x, x) - \sum_{j=1}^n N_j^2(x)$$

like all  $\mathcal{K}$ -orthonormal bases. The interpolant can be written as

$$s_{X,f}(x) = T(x)\alpha_X = N(x)L^{-1}E(f)$$

where the generalized divided differences shall be  $\Delta_N(f) = {}^{-1}E(f)$ , then

$$s_{X,f}(x) = N(x)\Delta_N(f).$$

The Newton basis is orthonormal in the native space of the kernel. The basis functions are Riesz representers of the divided differences, and thus the divided differences with respect to the Newton basis are a dual orthonormal basis.

In practice, the Cholesky decomposition would be pivoted, but we shall describe another adaptive algorithm below.

The construction of the Newton basis is recursive like the Cholesky algorithm, and this means that the first  $n$  basis functions  $N_1, \dots, N_n$  need not be recalculated when going over from  $n$  to  $n+1$ . This has some remarkable consequences which were not noted in [18]. In particular, we can interpret (3.14) in the form

$$\sum_{j=1}^n N_j^2(x) = K(x, x) - P_n^2(x), \quad (3.19)$$

recursively, if we denote the power function on the  $n$  points of  $X_n := \{x_1, \dots, x_n\}$  by  $P_n$ . Thus

**Theorem 3.7.1.**

$$N_n^2(x) = P_{n-1}^2(x) - P_n^2(x) \leq P_{n-1}^2(x). \quad (3.20)$$

If  $x_n$  is chosen recursively as

$$x_n := \arg \max P_{n-1}^2(\cdot),$$

then

$$N_n^2(x) \leq N_n^2(x_n) \text{ for all } x \in \Omega,$$

i.e. the basis has no parasitic maxima.

**Proof:** The first statement follows from (3.19), and the second from  $N_n^2(x) \leq P_{n-1}^2(x) \leq P_{n-1}^2(x_n) = N_n^2(x_n)$  because of  $P_n(x_n) = 0$ .  $\square$

This argument was already used in [8] for the Lagrange basis. If  $x_1, \dots, x_n$  are fixed, the functions  $L_n$  and  $N_n$  differ only by a normalization factor, but  $L_n$  will change when we go over to  $n + 1$ .

In (3.19), one can take the limit  $n \rightarrow \infty$  without problems, and it was proven in [8] that

$$\sum_{j=1}^{\infty} N_j^2(x) = K(x, x)$$

if the points  $x_1, x_2, \dots$  get dense in a bounded domain  $\Omega \subset \mathbb{R}^d$ . On such a domain, and for a continuous kernel  $K$ , we also get

$$\sum_{j=1}^n \|N_j\|_{L_2(\Omega)}^2 \leq \int_{\Omega} K(x, x) dx$$

by integration of (3.19), for  $n \rightarrow \infty$ , and for the craziest possible point distributions. Together with (3.20), this shows that the Newton basis does not seriously degenerate when points get close. This is in line with the nondegeneracy of the data provided by (3.17), if the data come from a function  $f$  in the native space. It is also in line with the behavior of the Newton basis for polynomial interpolation of smooth univariate functions. The basis functions do not degenerate and the divided differences turn into derivatives.

By construction, the functions  $N_1, \dots, N_n$  are an orthonormal basis for the span of the translates  $K(\cdot, x_1), \dots, K(\cdot, x_n)$ . Thus the action of the reproducing kernel  $K$  on that space is given by

$$K_n(x, y) = \sum_{j=1}^n N_j(x) N_j(y)$$

and the action of the kernel  $K_n$  on the orthogonal complement

$$\{f \in \mathcal{K} : f(x_j) = 0, 1 \leq j \leq n\}$$

is given by the kernel  $K(x, y) - K_n(x, y)$ . In [17], the latter was called the *power kernel* and represented differently, without using the Newton basis see Definition (2.1.2). If the points  $x_1, x_2, \dots$  are dense in a bounded domain  $\Omega$ , this argument proves the series representation

$$K(x, y) = \sum_{j=1}^{\infty} N_j(x)N_j(y) \quad (3.21)$$

of the kernel. From Section 2.2 we know that there may be a good low-rank approximation to the kernel, and thus we have to anticipate that, for a special ordering of the points, convergence of the series may be rapid and connected to the decay of the eigenvalues of the kernel. This means that one should consider adaptive point selections that make the series converge fast. This will be the topic of the next section.

### 3.8 Generalized Interpolation

We consider data  $\{x_i, \lambda_i f\}$ ,  $i = 1, \dots, N$ , where  $\Lambda = \{\lambda_1, \dots, \lambda_N\} \subset N_{\mathcal{K}}^*$ , is a linearly independent set of continuous linear functionals and  $f$  is some (smooth) data function. For example, it could denote evaluation of some derivative at the point  $x_i$ . Therefore, we assume the generalized Hermite interpolation to be of the form:

$$s_{\Lambda}(x) = \sum_{j=1}^N \alpha_j \lambda_j^y K(x, x_j), \quad x \in \mathbb{R}^s,$$

from the space

$$N_{\Lambda} = \text{span}\{\lambda_j^y K(x, y) : j = 1, \dots, N\},$$

and require it to satisfy

$$\lambda_i^x s_{\Lambda} = \lambda_i^x f, \quad i = 1, \dots, N$$

The linear system  $A_{\Lambda} \alpha = f_{\Lambda}$ , which arises in this case has matrix entries

$$A_{\Lambda} = (\lambda_i^x \lambda_j^y K(x_i, x_j))_{1 \leq i, j \leq N},$$

and right-hand side  $f_{\Lambda} = [\lambda_1 f, \dots, \lambda_N f]^T$ . The generalized kernel matrix  $A_{\Lambda}$  is symmetric and positive definite whenever  $\lambda_j$  are linearly independent over  $N_{\mathcal{K}}(\Omega)$ . Then a generalized recovery problem would seek to find a function  $s \in \mathcal{K}$  such that

$$\lambda_j(s) = \lambda_j(f), \quad 1 \leq j \leq N,$$

for some  $f$  in the native space. We will call  $s$  a generalized interpolant. The optimal recovery problem searches for the norm-minimal generalized interpolant,

$$\|s^*\| = \min\{\|s\| : s \in \mathcal{K}, \lambda_j(s) = \lambda_j(f), 1 \leq j \leq N\}.$$

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

### 3.9 Notes and comments

In this chapter, we sought to establish a methodology for all possibilities to construct bases of data-dependent subspaces of reproducing kernel Hilbert spaces. We described bases in general and then proceeded to special bases with different properties. In particular, we characterized all  $\mathcal{K}$ -orthonormal bases and proved that they lead to stable results in function space even for nearly-coalescing data points, provided that the data come from a function in the native space. We also considered dual bases that depend on the functionals induced by any basis change. The functionals act like divided differences, which are bounded in the dual of the native space for all data sets. The  $\mathcal{K}$ -orthonormality of bases carries over to  $\mathcal{K}$ -orthonormality of functionals. The Newton basis seems to be a particularly good choice, since it is  $\mathcal{K}$ -orthonormal, allows stable evaluation in the sense of (3.10) and can be calculated recursively and adaptively. Depending on point selections, convergence rates of series like (3.19) and (3.21) should be investigated further. Efficient adaptive algorithms for calculating the Newton basis along the lines of orthogonal matching pursuit will provide some numerical results in the next chapter.

# Adaptive Calculation of Newton Bases

---

The Newton basis seems to be a particularly good choice, since it is  $\mathcal{K}$ -orthonormal, allows stable evaluation in the sense of (3.10) and also can be calculated recursively and adaptively, which we will show it in this chapter. We take this material from [14]

## 4.1 $P$ -dependent Point Selection Strategy

We first consider the case where we want to find a good basis for *all* possible interpolation problems, i.e. we do not care for single data and focus on point selection instead. Let  $\Omega \subset \mathbb{R}^d$  be a bounded domain with a continuous positive definite kernel  $K$  on it. For applications, users will take a large finite subset  $X = \{x_1, \dots, x_N\} \subset \Omega$  but avoid to form the huge  $N \times N$  kernel matrix  $A$  for all points of  $X$ . Instead, one can use a column-based version of a pivoted Cholesky decomposition which performs  $m$  steps of computational complexity  $\mathcal{O}(Nm)$  to stop at rank  $m$ . Its overall complexity thus is  $\mathcal{O}(Nm^2)$ , and it requires a total storage of  $\mathcal{O}(mN)$ . It builds the first  $m$  columns of the value matrix  $V_N$  and thus the first  $m$  Newton basis functions on the full set  $X$ .

Though it is well-known how to perform a pivoted column-based Cholesky decomposition, we describe the algorithm here because we want to make use of the kernel background and to end up with functions, not just with matrices. Having (3.19) in mind, we start the algorithm by choosing  $x_1$  by permutation of points as

$$x_1 = \arg \max \{K(x, x) : x \in X\}.$$

Since we always assume the kernel to be easily evaluated, this will use  $\mathcal{O}(N)$  operations, and we store the vector

$$z := (K(x_1, x_1), \dots, K(x_N, x_N))^T$$

for later use. The first Newton basis function then is

$$N_1(x) := \frac{K(x, x_1)}{\sqrt{K(x_1, x_1)}} \tag{4.1}$$

due to (2.5), and we evaluate it on  $X$ , giving us the first column  $v_1$  of the value matrix  $V_N$ . Another  $N$ -vector  $w$  should store the values  $v_1^2(x_j)$ ,  $1 \leq j \leq N$ .

Assume now that we have fixed the first  $m$  points and the first  $m$  columns of  $V_N$ , forming an  $N \times m$  matrix  $V_m$ . The vector  $w$  should contain the values

$$w_j = \sum_{k=1}^m v_k^2(x_j), \quad 1 \leq j \leq N.$$

For points  $x_j \in X$ , the power function  $P_m$  is

$$P_m^2(x_j) = K(x_j, x_j) - \sum_{k=1}^m v_k^2(x_j) = z_j - w_j, \quad 1 \leq j \leq N,$$

and we find its maximum and assume that it is attained at  $x_{m+1}$ . Note that the first  $m$  components of  $z - w$  should be zero, giving us some check for roundoff.

The algorithm stops if this maximum is smaller than a tolerance  $\epsilon^2$ . In that case,  $P_m(x) \leq \epsilon$  holds on all of  $X$ , and all functions  $f$  in  $\mathcal{K}_X$  can be replaced by their interpolants  $s_f$  in  $x_1, \dots, x_m$  with errors

$$|f(x) - s_f(x)| \leq \epsilon \|f\|_{\mathcal{K}} \text{ for all } x \in X.$$

Thus, using more than our  $m$  well-selected points of  $X$  is useless if we can tolerate the above error.

If we decide to continue, we now generate the column  $Ae_{m+1}$  of  $A$  consisting of values  $K(x_{m+1}, x_j)$ ,  $1 \leq j \leq N$  and form the vector  $u := Ae_{m+1} - V_m \cdot (V_m^T e_{m+1})$  at cost  $\mathcal{O}(Nm)$ . This contains the values on  $X$  of the function

$$u(x) := K(x, x_{m+1}) - \sum_{j=1}^m N_j(x)N_j(x_{m+1}), \quad (4.2)$$

and this function satisfies

$$(u, N_k)_{\mathcal{K}} = N_k(x_{m+1}) - \sum_{j=1}^m N_j(x_{m+1})(N_j, N_k)_{\mathcal{K}} = 0, \quad 1 \leq k \leq N.$$

Since the span of  $N_1, \dots, N_m$  coincides with the span of  $K(\cdot, x_1), \dots, K(\cdot, x_m)$ , we also have  $u_j = u(x_j) = 0$ ,  $1 \leq j \leq m$ , giving us another check on roundoff.

We then define

$$N_{m+1}(x) := \frac{u(x)}{\|u\|_{\mathcal{K}}}$$

and use  $\|u\|_{\mathcal{K}}^2 = z_{m+1} - w_{m+1} = P_m^2(x_{m+1})$  for this. To prove this identity, we employ orthonormality of the  $N_j$  in (4.2) to get

$$\begin{aligned} \sum_{j=1}^m N_j^2(x_{m+1}) &= \|u - K(\cdot, x_{m+1})\|_{\mathcal{K}}^2 \\ &= \|u\|_{\mathcal{K}}^2 - 2(u, K(\cdot, x_{m+1}))_{\mathcal{K}} + K(x_{m+1}, x_{m+1}) \\ &= \|u\|_{\mathcal{K}}^2 - 2u(x_{m+1}) + K(x_{m+1}, x_{m+1}) \\ = w_{m+1} &= \|u\|_{\mathcal{K}}^2 - 2u_{m+1} + z_{m+1} \end{aligned}$$

and insert  $u_{m+1} = z_{m+1} - w_{m+1}$ . We update  $w$  and add the vector  $u/\|u\|_{\mathcal{K}}$  as a new column to  $N_m$  to finish step  $m + 1$ .

This algorithm provides the first  $m$  Newton basis functions on  $N$  points at cost  $\mathcal{O}(Nm^2)$ . It is particularly useful if the kernel has a good low-rank approximation and if the user wants results on a large but fixed point set  $X$ .

If data  $E(f)$  of a function  $f$  are given on  $X$ , one might simply set up the overdetermined system

$$V_m c = E(f)$$

and solve it in the least-squares sense for a coefficient vector  $c \in \mathbb{R}^m$ . Another possibility is to take only the first  $m$  rows of this system, thus getting away with an interpolant on  $x_1, \dots, x_m$ . This system is triangular, can be solved at cost  $\mathcal{O}(m^2)$  and usually is quite sufficient, because the main algorithm usually is stopped when the power function is very small on all of  $X$ .

If we denote the top  $m \times m$  part of  $V_m$  by  $L_m$ , we get the values of the Lagrange basis for nodes  $x_1, \dots, x_m$  on all of  $X$  as the matrix  $V_m \cdot L_m^{-1}$ . The divided differences for the Newton basis  $N_1, \dots, N_m$  are obtainable as  $L_m^{-1} E_m(f)$ , if we take the first  $m$  components of  $E(f)$  as  $E_m(f)$ .

For use with meshless methods, whose bases should be expressed “entirely in terms of nodes” [2], we suggest not to use the Lagrange basis based on function values in  $x_1, \dots, x_m$ , but rather the Newton basis, the divided differences being the parametrization instead of the function values at those nodes. If a result is expressed as a coefficient vector  $c \in \mathbb{R}^m$  in this parametrization, the resulting values at all nodes of  $X$  are given by  $V_m \cdot c$ .

If the values of the Newton basis are needed at other points, or if derivatives are to be calculated, we can use the standard equation (3.2) and insert  $V_U = C_U^{-1}$  for a  $\mathcal{K}$ -orthonormal basis  $U$ . Then we get the linear system

$$V_N \cdot N^T(x) = T(x)^T$$

for the Newton basis. If the basis is shortened to  $m$  functions, this system is shortened to be  $m \times m$ , and then it has the lower triangular coefficient matrix  $L_m$  we had before. If linear maps  $\mathcal{L}$  like derivatives have to be evaluated, we use the system

$$V_N \cdot \mathcal{L}(N^T(\cdot)) = \mathcal{L}(T(\cdot)^T)$$

in shortened form. These evaluations are of complexity  $\mathcal{O}(m^2)$  each, which is compatible with the complexity  $\mathcal{O}(Nm^2)$  we already had for getting values on  $N$  points.

## 4.2 $f$ -dependent Point Selection Strategy

If we are given a specific vector  $E(f)$  of data on a point set  $X$ , there are fully data-dependent adaptive techniques for approximation of  $f$ . A particularly simple one is based on a pivoted  $QR$  decomposition and can be viewed as an instance of Orthogonal Matching Pursuit [16] in the column space of the kernel matrix. Starting theoretically from a huge linear system (3.7) based on  $N$  points, the basic idea is to project the right-hand side into the column space and to select those columns that allow to reproduce the right-hand side with good accuracy. In principle, this could be done by applying a column-pivoted  $QR$  algorithm to the system

$$\begin{pmatrix} -E(f) & A \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

making sure that the first column is used first. In addition, the QR algorithm should not start from full matrices, but rather work on columns only. Again, this is well-known from Numerical Linear Algebra, but we want to describe a different algorithm that performs data-dependent Orthogonal Matching Pursuit in the Hilbert space  $\mathcal{K}$  and uses the Newton basis.

Let  $X$  be a large set of  $N$  points  $x_1, \dots, x_N$ , and assume the data vector  $E(f) = (f(x_1), \dots, f(x_N))^T$  to be given from a function  $f \in \mathcal{K}$  we do not know explicitly, but which we want to approximate. By (2.4), we know that the data have the semantics

$$f(x_j) = (f, K(\cdot, x_j))_{\mathcal{K}}, \quad 1 \leq j \leq N$$

though we do not know  $f$ . Selecting  $x_1$  to be the point where the data vector  $E(f)$  attains its maximum absolute value means that we have selected the kernel translate that will approximate  $f$  best in  $\mathcal{K}$ . We now proceed like in the previous algorithm, forming the first Newton basis function  $N_1 := K(\cdot, x_1)/\sqrt{K(x_1, x_1)}$ . But then we replace the data of  $f$  by the data of the error  $f_1 := f - (f, N_1)_{\mathcal{K}}N_1$  of its best approximation by multiples of  $N_1$  or  $K(\cdot, x_1)$  in  $\mathcal{K}$ . For this, we only need

$$(f, N_1)_{\mathcal{K}} = f(x_1)/\sqrt{K(x_1, x_1)}.$$

Then we proceed by choosing  $x_2$  as the point where  $f_1$  attains its maximum absolute value.

This algorithm constructs a Newton basis, but with a different, now  $f$ -dependent selection of points. In the notation of the above algorithm, let us

assume that we already have  $V_m$  and need

$$f_m := f - \sum_{j=1}^m (f, N_j)_{\mathcal{K}} N_j = f_{m-1} - (f, N_m)_{\mathcal{K}} N_m \quad (4.3)$$

on the full set  $X$ . To perform this recursively, we look at step  $m+1$  and use (4.2) to get

$$(f, u)_{\mathcal{K}} = f(x_{m+1}) - \sum_{j=1}^m (f, N_j)_{\mathcal{K}} N_j(x_{m+1})$$

and, with proper normalization,

$$(f, N_{m+1})_{\mathcal{K}} = (f, u)_{\mathcal{K}} / \|u\|_{\mathcal{K}}$$

with  $\|u\|_{\mathcal{K}}^2 = z_{m+1} - w_{m+1}$  as shown before. Thus we have a recursion for these inner products, and inserting them into (4.3) allows to find  $x_{m+1}$  as the point where  $f_m$  attains its maximal absolute value.

This algorithm is Orthogonal Matching Pursuit in  $\mathcal{K}$ , and it should be terminated when  $|(f, N_{m+1})_{\mathcal{K}}|$  is sufficiently small. By orthogonality to the span of  $K(\cdot, x_j)$  for  $1 \leq j \leq m$ , the result is the interpolant to  $f$  on  $x_1, \dots, x_m$ . It should finally be noted that the method is a reformulation of the greedy method of [24] in terms of the Newton basis.

### 4.3 Numerical Tests

We consider the domain  $\Omega$  defined by the unit disk with the third quadrant cut away. We select a large set  $X$  of points on a fine grid on  $[-1, 1]^2$  that fall into  $\Omega$ . Then we run the adaptive algorithm of Section (4.1) to generate a selection of well-distributed points. For the Gaussian at scale 2, Figure (4.1) shows the first 30 selected points and the decay of the maximum of Power function for that case. For 100 points and inverse multiquadrics of the radial form

$$\phi(r) = (1 + r^2/8)^{-2},$$

similar plots are in Figure (4.2). The Newton basis function  $v_{25}$  for the Gaussian at scale 2 is in Figure (4.6). We now turn to the  $f$ -dependent point selection strategy. Figure (4.5) shows the results for the function  $f(x, y) := \exp(|x - y|) - 1$  on the same domain. One can see the accumulation of selected points close to the derivative discontinuity at  $x = y$ . And also, the Newton basis function  $v_{25}$  for the Wendland functions at scale 4 is in Figure (4.6). We ran the algorithm of all cases of data-dependent bases. The case of interpolation by the Gaussian at scale 2 are illustrated in following figures up to (4.8)

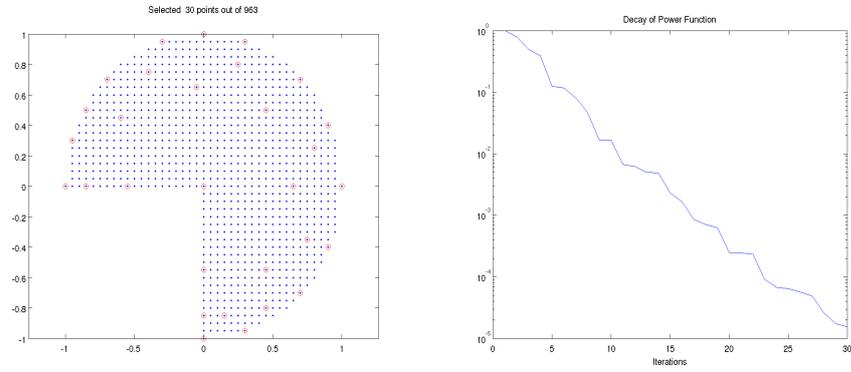


Figure 4.1: Selected points and power function decay for the Gaussian

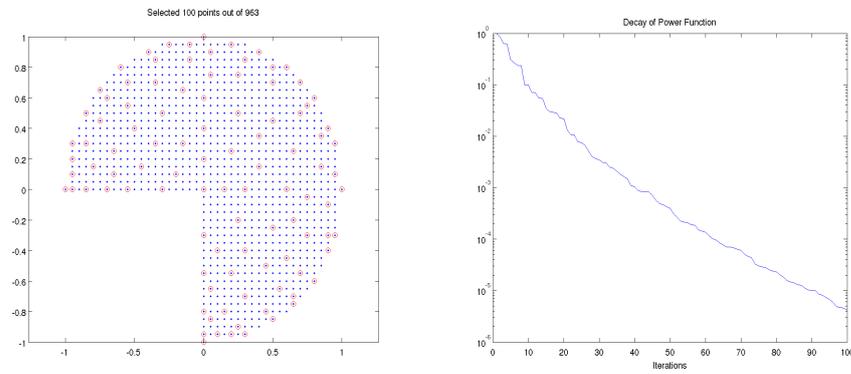


Figure 4.2: Selected points and power function decay for an inverse multi-quadratic

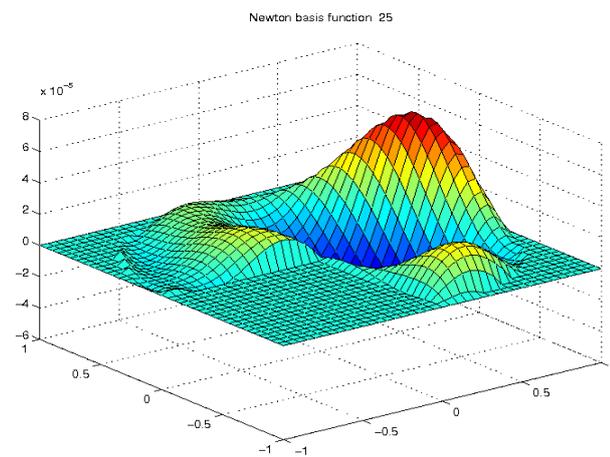


Figure 4.3: Newton basis function  $v_{25}$  for the Gaussian

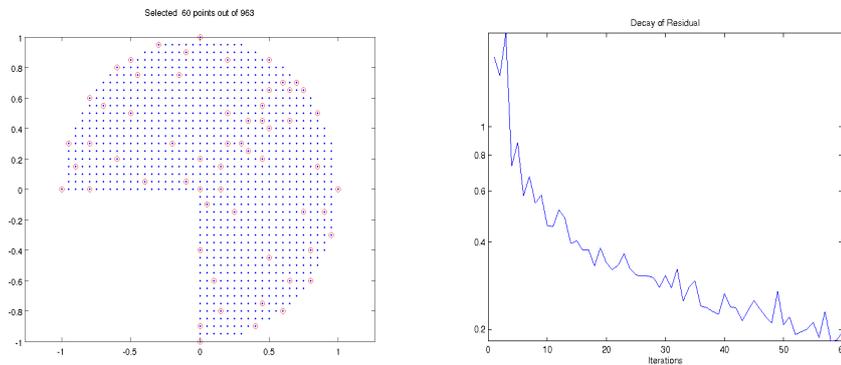


Figure 4.4: Data-dependent greedy point selection and power function decay for the Wendland functions

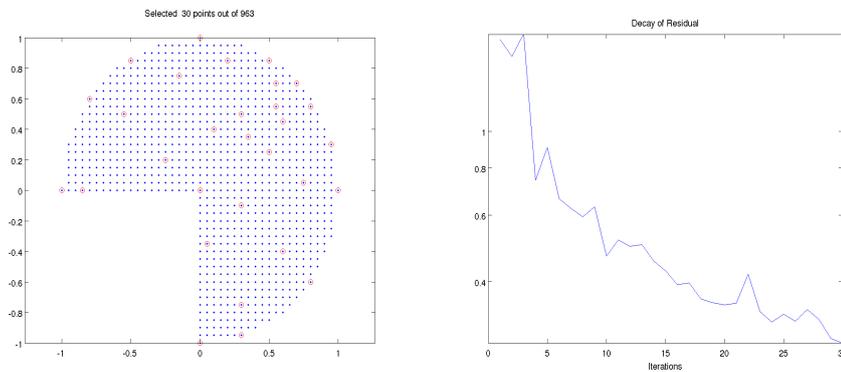


Figure 4.5: Selected points and power function decay for the Wendland functions scale 4

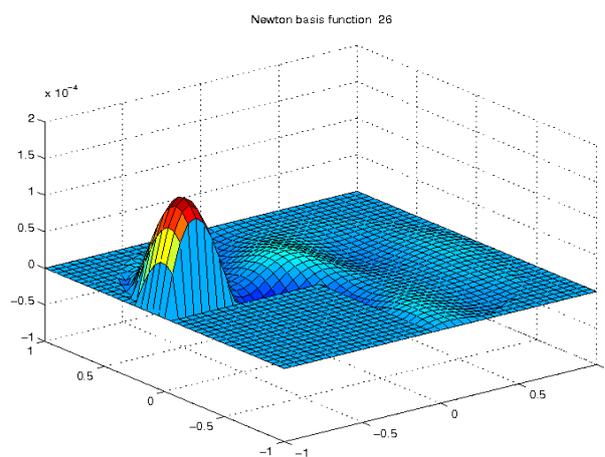


Figure 4.6: Newton basis function  $v_{25}$  for the Wendland functions

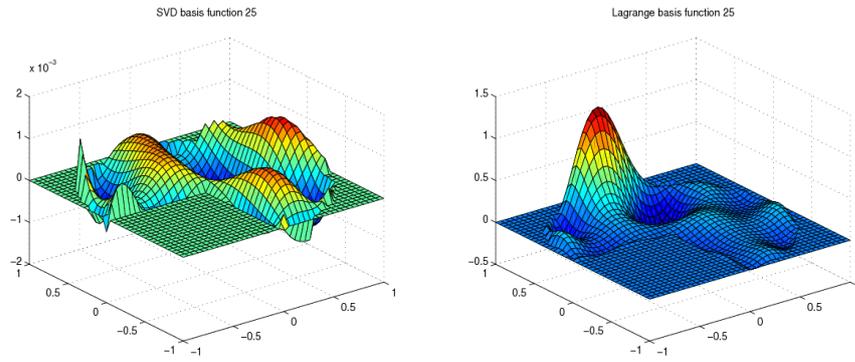


Figure 4.7: SVD basis and Lagrange basis for the Gaussian

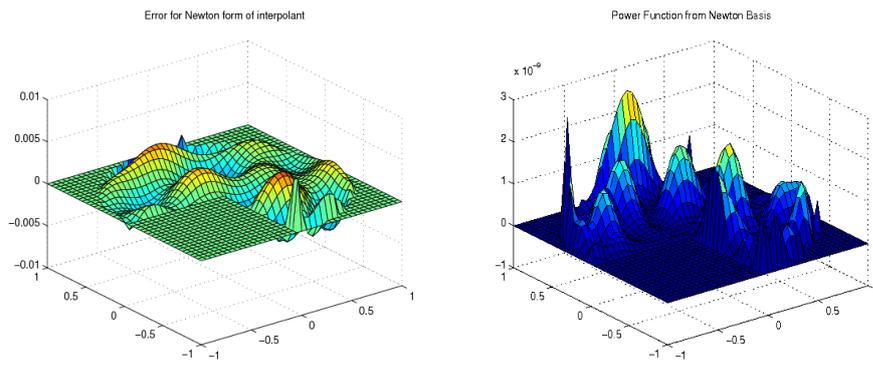


Figure 4.8: Interpolation error and power function for Gaussian

# The Conditionally Positive Definite Bases

---

This chapter extends the previous chapter to the important case of conditionally positive kernels such as thin-plate splines or polyharmonic kernels. We go back to the notation in Section 2.3, and use  $m$  as the order of a  $\mathcal{P}_m^d$ -conditionally positive definite kernel  $K$ , we use here. The interpolation sets  $X = \{x_1, \dots, x_N\}$  assumed to be  $\mathcal{P}_m^d$ -unisolvent. The case  $m = 0$  of the chapter concerning bases for unconditionally positive definite kernels, will serve as a guideline. Again, particularly interesting cases are bases of Lagrange or Newton type, and bases which are orthogonal or orthonormal, either discretely (i.e. via their function values on the centers) or as elements of the underlying “native” space for the given kernel. This chapter first looks at bases from a general point of view, i.e. not assigning a special rôle to polynomials, and not reordering points. In particular, we shall consider bases with orthogonality properties, Lagrange and Newton bases, and bases resulting from certain modifications of the given kernel. We close with a few numerical examples, including an adaptively constructed and partially orthonormal Newton-type basis.

## 5.1 Notation and Basic Facts

The goal is to provide useful bases for  $\mathcal{K}_X$  of (2.14) and to exhibit relations between these bases. Due to their dependence on the given unisolvent set  $X$ , we call these bases *data-dependent*. It will turn out that Section 2.1 does not generalize in a straightforward way, since there are many possibilities to proceed. For instance, one can preselect a polynomial basis and complete it by  $N - Q$  other functions, or let all basis functions contain some polynomial part. Another option is to preselect a minimal unisolvent subset of  $Q$  points first and then use a “reduced” kernel. Finally, one can try to reduce the conditional positive definite case to the unconditionally positive definite case by a suitable change of the kernel, but even this change can be done in different ways. These different approaches sometimes lead to the same result. With the space  $\mathcal{P}_m^d$  of  $d$ -variate real-valued polynomials of order at most  $m$ , let  $Q$  be the dimension of  $\mathcal{P}_m^d$  and define a basis  $p_1, \dots, p_Q$  of  $\mathcal{P}_m^d$  which does not necessarily consist of monomials. We restrict the admissible point sets  $X = \{x_1, \dots, x_N\} \subset \Omega \subset \mathbb{R}^d$  to be  $\mathcal{P}_m^d$ -unisolvent and fix a conditionally

positive (semi-) definite kernel  $K : \Omega \times \Omega \rightarrow \mathbb{R}^d$  of order  $m$  on a set  $\Omega \subseteq \mathbb{R}^d$ . In view of the Definition (2.3.2) and (2.14), we define the spaces  $M_{X,m}$  and  $\mathcal{K}_X$ . Our goal is to derive bases of  $\mathcal{K}_X$  with certain properties. Since we fix  $X$ , we call these bases *data-dependent*, and as long as we shall deal with a fixed  $X$ , we shall drop  $X$  in the notation later.

Following (2.17), we can split each  $s_{X,c_X,b_X} \in \mathcal{K}_X$  into

$$\begin{aligned} s_{X,c_X,b_X} &= s_{X,c_X,b_X}^K + s_{X,c_X,b_X}^P &= s_{X,c_X} + p_{X,b_X}, \\ s_{X,c_X,b_X}^K &:= s_{X,c_X} &:= K_X(x)c_X, \quad c_X \in \mathbb{R}^{|X|}, \quad P_X^T c_X = 0, \\ s_{X,c_X,b_X}^P &:= p_{X,b_X} &:= p_X(x)b_X, \quad b_X \in \mathbb{R}^Q. \end{aligned} \tag{5.1}$$

**Theorem 5.1.1.** *The split in (5.1) is unique.*

**Proof:** It suffices to show that any function  $s_{X,c_X}$  of the above form with a vector  $c_X \in \mathbb{R}^{|X|}$  satisfying moment conditions  $P_X^T c_X = 0$  vanishes, if it coincides with a polynomial  $p = p_{X,b_X} \in \mathcal{P}_m^d$ . But since then  $s_{X,c_X,-b_X} = s_{X,c_X} - p_{X,b_X}$  is in  $\mathcal{K}_X$  and vanishes on  $X$ , the nonsingularity of  $\mathcal{A}$  implies  $c_X = 0$  and  $b_X = 0$ .  $\square$

The bilinear form (2.13) now can be used to define a bilinear form

$$(s_{X,c_X,b_X}, s_{Y,c_Y,b_Y}) := c_X^T A_{XY} c_Y = (s_{X,c_X}, s_{Y,c_Y}) \tag{5.2}$$

on interpolants on unisolvent sets  $X$  and  $Y$ . Using this bilinear form, we shall look at bases with orthogonality properties. We do not need the full Hilbert space background here, but we refer the reader to Section 2.3.

For later use, we define the notation

$$s_c := \sum_{x_j \in X} c_j K(x_j, \cdot) \tag{5.3}$$

for functions in  $\mathcal{K}_X$  depending on a vector  $c \in M = M_{X,m}$ , and similarly

$$\lambda_c := \sum_{x_j \in X} c_j \delta_{x_j} : f \mapsto \sum_{x_j \in X} c_j f(x_j) \tag{5.4}$$

for an associated linear functional, which vanishes on  $\mathcal{P}_m^d$  due to the moment conditions implied by  $a \in M$ .

**Lemma 5.1.1.** *In (2.14), the sum is direct, the space  $\mathcal{K}_X$  has dimension  $N = |X|$ , and interpolation on  $X$  by functions of  $\mathcal{K}_X$  is uniquely possible.*

**Proof** If we have  $s_c = p$  for some  $c \in M$  and some  $p \in \mathcal{P}_m^d$  using (5.3), then  $\lambda_c(s_c) = Q(c) = 0$  and thus  $c = 0$  due to conditional positive

definiteness. This proves the first assertion, and the second will follow from the third.

The argument above also works if we have  $s_c = p$  only on the points of  $X$ . This implies that there is no nonzero function in  $\mathcal{K}_X$  that vanishes on  $X$ , and this proves uniqueness of interpolants, if they exist. If we want to find a function  $p+s_c \in \mathcal{K}_X$  with prescribed function values  $f_k = p(x_k)+s_c(x_k)$ ,  $1 \leq k \leq N$ , we have to solve the linear system

$$\begin{pmatrix} A & P \\ P^T & 0_{Q \times Q} \end{pmatrix} \begin{pmatrix} c \\ b \end{pmatrix} = \begin{pmatrix} f \\ 0_{Q \times 1} \end{pmatrix}$$

where  $f = (f_1, \dots, f_N)^T \in \mathbb{R}^N$ ,  $c \in \mathbb{R}^N$ , and  $b \in \mathbb{R}^Q$ , while the coefficient matrix is the  $(N+Q) \times (N+Q)$  matrix

$$\mathcal{A} := \begin{pmatrix} A & P \\ P^T & 0_{Q \times Q} \end{pmatrix}. \quad (5.5)$$

Note that  $P^T c = 0$  is equivalent to  $c \in M$ , making up the final  $Q$  equations. Solvability of the homogeneous system means that some  $s_c$  with  $c \in M$  and some

$$p_b := \sum_{i=1}^Q b_i p_i \in \mathcal{P}_m^d \quad (5.6)$$

exist such that  $p_b + s_c$  vanishes on  $X$ , and this is impossible unless both  $p_b$  and  $s_c$  are identically zero. Thus the matrix (5.5) is nonsingular and interpolation of arbitrary data is uniquely possible.  $\square$

## 5.2 General Data-Dependent Bases

Each data-dependent basis  $w_1, \dots, w_N$  must have a nonsingular  $N \times N$  value matrix  $V_w = (w_i(x_j))$  that leads to the system

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} C_w \\ B_w \end{pmatrix} = \begin{pmatrix} V_w \\ 0 \end{pmatrix}. \quad (5.7)$$

The *moment conditions* are employed via  $P^T C_w = 0$ , while the *coefficient matrices*  $C_w$  and  $B_w$  satisfy

$$AC_w + PB_w = V_w, \quad (5.8)$$

that expresses the  $N \times N$  matrix  $C_w$  and the  $Q \times N$  matrix  $B_w$  uniquely in terms of  $V_w$ . Clearly, this matrix is necessarily nonsingular because the basis must allow unique interpolation on  $X$ . Here will be no special rôle of polynomials, and no reordering of the points of  $X$ .

**Definition 5.2.1.** The value matrix of a basis  $w = (w_1, \dots, w_N)$  is the  $N \times N$  matrix

$$V_w = (w_i(x_j))_{1 \leq i, j \leq N} \in \mathbb{R}^{N \times N}$$

where  $j$  is the row and  $i$  is the column index.

The value matrix  $V_w$  defines a basis  $w$  uniquely. Clearly the associated basis  $w$  is then determined by the columns of  $C_w$  and  $B_w$  via

$$w_k = \sum_{j=1}^N c_{jk} K(x_k, \cdot) + \sum_{i=1}^Q b_{ik} p_i, \quad 1 \leq k \leq N \quad (5.9)$$

where we omitted  $w$  in the notation of the matrix elements.

**Theorem 5.2.1.** A basis  $w$  is uniquely defined by either a nonsingular  $N \times N$  value matrix  $V_w$  or by a  $N \times N$  matrix  $C_w$  and a  $Q \times N$  matrix  $B_w$  such that  $P^T C_w = 0$  holds and  $\begin{pmatrix} C_w \\ B_w \end{pmatrix}$  has rank  $N$ .

**Proof** The first part is known already. Reading (5.7) from left to right, one can start with matrices  $C_w$  and  $B_w$  satisfying the hypotheses, define a basis  $w$  by (5.9), and get the value matrix via (5.7).  $\square$

**Theorem 5.2.2.** Using the notation (5.5) and (5.8), the  $N \times N$  matrix  $C_w$  has rank  $N - Q$ , and the  $(N + Q) \times N$  matrix

$$\begin{pmatrix} C_w \\ B_w \end{pmatrix} = \mathcal{A}^{-1} \begin{pmatrix} V_w \\ 0 \end{pmatrix}$$

has rank  $N$ .

**Proof** The second statement is trivial, and the moment conditions imply that there are  $Q$  linear independent linear relations between rows of  $C_w$ . Thus the row rank of  $C_w$  is at most  $N - Q$ , but since the second statement holds and  $B_w$  is an  $Q \times N$  matrix, the row rank of  $C_w$  must be  $N - Q$ .  $\square$

The basis functions themselves have the representation

$$(w_1(x), \dots, w_N(x)) = (K(x, x_1), \dots, K(x, x_N)) \cdot C_w + (p_1(x), \dots, p_Q(x)) \cdot B_w \quad (5.10)$$

which splits them additively via (5.1) into

$$w_j = w_j^K + w_j^P, \quad 1 \leq j \leq N \quad (5.11)$$

with

$$\begin{aligned} w_j^K(x) &= (K(x, x_1), \dots, K(x, x_N)) C_w e_j, \\ w_j^P(x) &= (p_1(x), \dots, p_Q(x)) B_w e_j. \end{aligned} \quad (5.12)$$

into polynomials and linear combinations of kernel translates.

At this point we can already define the unique Lagrange basis  $L$  by requiring that the value matrix  $V_L$  is the identity matrix. Its construction matrices are then uniquely defined by

$$\begin{pmatrix} C_L \\ B_L \end{pmatrix} = \begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} I_{N \times N} \\ 0 \end{pmatrix}.$$

We could require the basis  $p_1, \dots, p_Q$  of  $\mathcal{P} = \mathcal{P}_m^d$  to be of Lagrange form on the first  $Q$  points of  $X$ . However, this might have an influence on the general form of data-dependent bases. But:

**Theorem 5.2.3.** *Given a nonsingular value matrix  $V_w$ , the corresponding basis  $w$  is uniquely defined and independent of basis changes in  $\mathcal{P}_m^d$ . In particular, the coefficient matrix  $C_w$  and the split (5.11) are unique, while the  $B_w$  matrices change.*

**Proof:** Each basis change in  $\mathcal{P}_m^d$  is given by the transition  $Q := P \cdot C$  with a nonsingular  $Q \times Q$  matrix  $C$ . Then, by easy calculations, the matrix  $B_w$  goes over to  $C^{-1}B_w$ , but by (5.10), the basis functions  $w_j$  are not changed, as is the matrix  $C_w$  by (5.8). In the second equation of (5.12), the matrix  $C$  cancels out.  $\square$

Using a general basis  $w$ , we can write the interpolant  $s_f$  to a data vector  $f \in \mathbb{R}^N$  as

$$s_f = \sum_{x_j \in X} w_j \lambda_{j,w}(f) \quad (5.13)$$

where the coefficients  $\lambda_{j,w}(f)$  must take linear combinations of the elements of  $f$ . If the data come from a function  $f$  in the sense that the above vector  $f$  is  $(f(x_1), \dots, f(x_N))^T$ , the coefficients  $\lambda_{j,w}(f)$  are linear combinations of point evaluation functionals  $\delta_{x_1}, \dots, \delta_{x_N}$ .

**Definition 5.2.2.** *The functionals  $\lambda_{j,w}(f)$  in (5.13) are the associated functionals to the basis  $w$ .*

If we assemble the associated functionals into a column vector  $\Lambda_w(f) \in \mathbb{R}^N$ , and if we insert points  $x_k$  into (5.13), we get  $f = V_w \Lambda_w(f)$  and thus

$$\Lambda_w(f) = V_w^{-1} f. \quad (5.14)$$

Thus, like in the unconditionally positive definite case see (3.9), the inverse of the value matrix governs how the associated functionals are composed out of the point evaluation functionals. Note the similarity to divided differences here. If, for interpolation by polynomials, we consider the Newton formula, we see the ‘‘Newton’’ basis and the divided differences making up an interpolant of the form (5.13). In such a case,  $V_w$  is a nonsingular triangular matrix, as is  $V_w^{-1}$ . In general, (5.14) implies

**Theorem 5.2.4.** *The calculation of the associated functionals to a general basis follows a triangular scheme if and only if the basis has a triangular pattern of zeroes. Similarly, the associated functionals are orthogonally composed out of the point evaluation functionals if and only if the basis has an orthogonal value matrix.*  $\square$

The identity (5.14) has some influence on the stability of evaluation of the interpolant (5.13). If we ignore the stability of calculating the basis  $w$  and the functionals  $\lambda_{j,w}(f)$ , the numerical stability of evaluating (5.13) at some point  $x$  is governed by the expression

$$\sum_{x_j \in X} |w_j(x)| |\lambda_{j,w}(f)| \leq \|w(x)\|_p \|\Lambda_w(f)\|_q$$

with  $1/p + 1/q = 1$ . We shall evaluate the right-hand side of this for various bases.

Using (5.8) and  $P^T C_w = 0$  we see that the matrix

$$(C_w^T, B_w^T) \mathcal{A} \begin{pmatrix} C_w \\ B_w \end{pmatrix} = C_w^T A C_w = G_w = C_w^T V_w$$

is symmetric. If we use the semi-inner product (5.2), the Gramian matrix

$$G_w = ((w_i, w_j))_{1 \leq i, j \leq N}$$

of semi-inner products consists of the bilinear form acting on columns of  $C_w$ , i.e.

$$G_w = C_w^T A C_w = C_w^T V_w = ((w_i^K, w_j^K))_{1 \leq i, j \leq N}, \quad (5.15)$$

but this matrix will usually be only positive semidefinite since we have a semi-inner product. In particular, Theorem (5.2.2) yields

**Corollary 5.2.1.** *The Gramian in the semi-inner product (5.2) of all data-dependent bases has rank  $N - Q$ . In particular, it is impossible to have a full orthonormal basis of  $N$  functions of  $\mathcal{K}_X$  if  $Q > 0$ .*  $\square$

Note that there is another form of the Gramian due to discrete  $\ell_2$  inner products on  $X$  is

$$H = V_w^T V_w,$$

since the rows of  $V_w$  correspond to points, while the columns correspond to functions. Clearly, this matrix is positive definite, and there are plenty of  $\ell_2$ -orthonormal bases of  $\mathcal{K}_X$ , in particular the Lagrange basis we shall consider in Section 5.7.

**Theorem 5.2.5.** *All  $\ell_2$ -orthonormal bases  $w$  on  $X$  arise from orthogonal value matrices  $V_w$  along the above lines. In particular, Theorem (5.2.4) is applicable here.*  $\square$

### 5.3 Polynomial Projectors

We now want to get rid of the polynomial part in (5.8). To this end, we need a  $N \times N$  matrix  $\Pi$  with  $\Pi P = 0$  and full rank  $N - Q$ . There are many ways to get such a matrix. Since the  $Q \times Q$  matrix  $P^T P$  is nonsingular due to unisolvency, we can define a symmetric projector via

$$\Pi := I - P(P^T P)^{-1} P^T. \quad (5.16)$$

For any vector  $z \in \mathbb{R}^N$  the vector  $\Pi z$  satisfies the moment conditions, and that for all vectors satisfying the moment conditions there is such a representation. This follows from the fact that  $\Pi$  is the orthogonal projector on  $\mathbb{R}^N$  to the kernel of  $P^T$ . Note that here  $\Pi$  is not the general projector defined in (2.18).

Another way to get an  $N \times N$  matrix  $\Pi$  with  $\Pi P = 0$  and full rank  $N - Q$  is to reorder the points of  $X$  to let the set  $X_Q := \{x_1, \dots, x_Q\}$  be unisolvent and to let the polynomial basis be a Lagrange basis on  $X_Q$ . Then we have  $P = \begin{pmatrix} I_{Q \times Q} \\ P_2 \end{pmatrix}$  and can define

$$\Pi = \begin{pmatrix} 0_{Q \times Q} & 0_{Q \times N-Q} \\ -P_2 & I_{N-Q \times N-Q} \end{pmatrix}. \quad (5.17)$$

This arises naturally here and will be useful for what follows. If seen as actions on data vectors, the special projector  $\Pi$  of (5.16) replaces the data by the error of the  $\ell_2$ -optimal polynomial recovery on  $X$ , while (5.17) replaces the data by the error of interpolation in  $X_Q$ . In both cases, the data of polynomials are mapped to zero, i.e.  $\Pi P = 0$  holds. However, the matrix  $\Pi$  of (5.16) has the advantage to be symmetric, idempotent, and independent of the choice of basis in  $\mathcal{P}_m^d$ .

We collect a few facts that directly follow from discrete least-squares approximation and linear algebra, In all cases, the maximal rank of  $\Pi$  and the property  $\Pi P = 0$  imply

**Theorem 5.3.1.** *1. For all vectors  $z$  in  $\mathbb{R}^N$  seen as values on  $X$ , the vector  $\Pi z$  is the error of the best  $\ell_2$  approximation to  $z$  by the range of  $P$ , i.e. by the values of polynomials of  $\mathcal{P}_m^d$  on  $X$ .*

*2.  $I - \Pi$  is the orthogonal projector to the range of  $P$ .*

*3.  $\Pi$  is the orthogonal projector on  $\mathbb{R}^N$  to the kernel of  $P^T$ .*

*4. For any vector  $z \in \mathbb{R}^N$ , the vector  $\Pi z$  satisfies the moment conditions.*

*5. For all vectors  $c$  satisfying the moment conditions there is a representation as  $c = \Pi z$  for some  $z \in \mathbb{R}^N$ .  $\square$*

**Theorem 5.3.2.** *The projector  $\Pi$  is independent of the choice of basis in  $\mathcal{P} = \mathcal{P}_m^d$ .*

**Proof:** If we go from  $P$  to  $P \cdot C$  with a nonsingular  $Q \times Q$  matrix  $C$ , the projector does not change. But the assertion also follows from the previous theorem, since  $\Pi$  depends only on the range of  $P$ .  $\square$

**Theorem 5.3.3.** *The matrix  $\Pi^T A \Pi$  is symmetric and positive semidefinite with rank  $N - Q$ .*

**Proof** The quadratic form  $z \mapsto z^T \Pi^T A \Pi z$  is positive semidefinite, because the vectors  $\Pi z$  satisfy the moment conditions. Its kernel is the kernel of  $\Pi$  and thus equal to the range of  $P$ , i.e. of dimension  $Q$ .  $\square$

If we multiply (5.8) from the left by  $\Pi$ , we get

$$\Pi A C_w = \Pi V_w \quad (5.18)$$

also for the general situation. Note that equation (5.18) generalizes the identity  $A C_w = V_w$  obtained in the unconditionally positive definite case see (3.4). For later use, we note that Theorem (5.3.1) implies that we can factorize

$$C_w = \Pi F_w \quad (5.19)$$

with a nonunique but nonsingular  $N \times N$  matrix  $F_w$ . In particular, a very natural candidate for a basis  $w^*$  is definable by

$$C_{w^*} := \Pi, \quad B_{w^*} = P^T,$$

due to the properties of  $\Pi$ . In the two special cases for  $\Pi$  provided above, the matrix  $\begin{pmatrix} \Pi^T \\ P^T \end{pmatrix}$  has rank  $N$ , and thus Theorem (5.2.2) is applicable, yielding a basis in both cases. Since Theorem (5.3.1) implies that  $\begin{pmatrix} \Pi \\ P \end{pmatrix}$  has rank  $N$ , and thus Theorem (5.2.1) is applicable. The value matrix of this basis is

$$V_{w^*} = A \Pi + P P^T.$$

### 5.3.1 Partially Orthonormal Bases

We now can look for all bases which have Gramians  $G_o$  which are  $N \times N$  diagonal matrices with  $Q$  zeros and  $N - Q$  ones on the diagonal. With the projector

$$\Pi := I - P(P^T P)^{-1} P^T$$

that is independent of the choice of polynomial basis, we can use (5.19) as  $C_w = \Pi F_w$  and write down the necessary equation

$$C_w^T A C_w = F_w^T \Pi^T A \Pi F_w = G_o = G_o^2.$$

Since  $F_w$  is nonsingular (but also nonunique), we get the decomposition

$$\Pi^T A \Pi = (F_w^T)^{-1} G_o^2 F_w^{-1} = E_w^T E_w$$

with the matrix

$$E_w := G_o F_w^{-1}.$$

**Theorem 5.3.4.** *All partially orthonormal bases of  $\mathcal{K}_X$  arise from factorizations*

$$\Pi^T A \Pi = F^T F \quad (5.20)$$

with  $N \times N$  matrices  $F$  of rank  $N - Q$ .

**Proof:** We saw above that the factorization is necessary. To prove that it is sufficient, we start from (5.21) and perform a Singular Value Decomposition  $F^T F = U^T \Sigma U$  of the positive semidefinite and symmetric matrix  $F^T F$ , with an orthogonal  $N \times N$  matrix  $U$ . The matrix  $\Sigma$  of singular values will have exactly  $Q$  zeros on the diagonal, and we introduce  $G_o := \text{sgn } \Sigma = G_o^2$  to reflect this pattern of zeros. If the invertible diagonal matrix  $S$  is obtained from  $\Sigma$  by replacing the zeros on the diagonal by ones, we get

$$\Sigma = \sqrt{S} G_o \sqrt{S}$$

and

$$\Pi^T A \Pi = F^T F = U^T \sqrt{S} G_o^2 \sqrt{S} U.$$

This allows us to define the  $N \times N$  matrix

$$C := \Pi U^T S^{-1/2}$$

that satisfies moment conditions and is of rank  $N - Q$ . We have to complement it by a  $Q \times N$  matrix  $B$  such that  $(C^T, B^T)$  has rank  $N$ , and then we get a basis  $w$  by taking  $C_w := C$  and  $B_w := B$  in (5.10). By construction, the Gramian  $G_w$  is

$$\begin{aligned} G_w &= C^T A C \\ &= S^{-1/2} U \Pi A \Pi U^T S^{-1/2} \\ &= S^{-1/2} U F^T F U^T S^{-1/2} \\ &= S^{-1/2} U U^T \sqrt{S} G_o^2 \sqrt{S} U U^T S^{-1/2} \\ &= G_o^2 = G_o. \quad \square \end{aligned}$$

Note that the value matrix  $V_w$  then is obtained via (5.7).

There is a simpler form of this.

**Theorem 5.3.5.** *All partially orthonormal bases of  $\mathcal{K}_X$  arise from factorizations*

$$\Pi^T A \Pi = E^T E \quad (5.21)$$

with  $N \times N$  matrices  $E$  of rank  $N - Q$  that have the property that  $E = G_o F^{-1}$  with a nonsingular  $N \times N$  matrix  $F$ , and where  $G_o$  is a  $N \times N$  diagonal matrix with  $Q$  zeros and  $N - Q$  ones on the diagonal.

**Proof:** We define a basis  $w$  via  $C_w := \Pi F$ . Then

$$\begin{aligned} C_w^T A C_w &= F^T \Pi A \Pi F \\ &= F^T E^T E F \\ &= F^T F^{T-1} G_o^T G_o F^{-1} F \\ &= G_o. \quad \square \end{aligned}$$

**Corollary 5.3.1.** *These partially orthonormal bases are not unique. Even if  $C_w$  is fixed along the above lines, the matrices  $B_w$  for application of Theorem (5.2.1) are still free.*  $\square$

We now consider bases that arise from pivoted Cholesky decompositions of the symmetric positive semidefinite  $N \times N$  matrix  $\Pi^T A \Pi$  of rank  $N - Q$ . We assume that after reordering of points and  $N$  steps of the Cholesky algorithm, we get

$$\Pi^T A \Pi = L G_o L^T$$

with a nonsingular lower triangular matrix  $L$  and a diagonal matrix  $G_o$  with the first  $N - Q$  diagonal elements being one and the final  $Q$  diagonal elements being zero.

Then, in the above context,  $F = (L^T)^{-1}$  and  $C = \Pi(L^T)^{-1}$  and  $G_o = C^T V = L^{-1} \Pi V$  leading to

$$\Pi V = L G_o,$$

generalizing what we had in the Newton case for unconditionally positive definite kernels. If we decompose in SVD style

$$\Pi^T A \Pi = U G_o U^T$$

with an  $N \times N$  orthogonal matrix  $U$  and a nonnegative matrix  $G_o$  of singular values with exactly  $N - Q$  positive ones, then

$$C = \Pi U, \quad \Pi V = U G_o,$$

again generalizing what we have in the standard case see Section 3.6.

## 5.4 Eliminating Moment Conditions

Unfortunately, the moment conditions in  $M$  are implicit additional conditions, blowing the interpolation problem up to dimension  $N + Q$  under  $Q$  conditions. To eliminate the conditions, we should construct an orthonormal basis of the orthogonal complement of  $P(\mathbb{R}^Q)$ , which is the kernel  $M$  of  $P^T$  in (2.11). This means that there is a  $N \times (N - Q)$  matrix  $Z$  with

$$M = Z(\mathbb{R}^{N-Q}), \quad M \ni a = Zc, \quad c \in \mathbb{R}^{N-Q}$$

whose columns are orthonormal. The matrix  $Z$  can be obtained as the left part of the  $N \times N$  orthogonal matrix  $Q$  arising from a  $QR$  decomposition of  $P$ . This means that the vectors  $a$  with

$$a_j = \sum_{i=1}^{N-Q} z_{ji} c_i$$

for vectors  $c \in \mathbb{R}^{N-Q}$  make up all of  $M$ , and thus we can write

$$s_a = \sum_{i=1}^{N-Q} c_i \underbrace{\sum_{x_j \in X} z_{ji} K(x_j, \cdot)}_{=: z_{i+Q}} = \sum_{i=1}^{N-Q} c_i z_i$$

for functions

$$z_{i+Q} := \sum_{x_j \in X} z_{ji} K(x_j, \cdot), \quad 1 \leq i \leq N - Q.$$

If we also define  $z_i = p_i$ ,  $1 \leq i \leq Q$ , we get a ‘‘EMC’’ basis of  $S$  obtained by eliminating **moment conditions**. We shall analyze it below, but note that the interpolation problem now needs solving the  $N \times N$  system

$$(AZ, P) \begin{pmatrix} c \\ b \end{pmatrix} = f$$

with the notation introduced above. Because of  $Z^T P = 0$ , this can be split into solving

$$\begin{aligned} Z^T AZ c &= Z^T f, \\ Pb &= f - AZc, \end{aligned}$$

where the second set of equations just is a solvable polynomial interpolation problem on the full unisolvent set  $X$ . As a byproduct, we get that the  $(N - Q) \times (N - Q)$  matrix  $Z^T AZ$  is symmetric and positive definite. Unfortunately, forming  $AZ$  needs  $N^2(N - Q)$  operations that can possibly be avoided by using other bases.

## 5.5 Points and Polynomials First

One way to solve systems with the coefficient matrix  $\mathcal{A}$  is to sort out  $Q$  points forming an unisolvent subset, and to reorder these points to become  $x_1, \dots, x_Q$ . Splitting the matrices  $A$  and  $\mathcal{A}$  similarly, we get

$$\mathcal{A} = \begin{pmatrix} A_{11} & A_{12} & P_1 \\ A_{12}^T & A_{22} & P_2 \\ P_1^T & P_2^T & 0 \end{pmatrix}.$$

Note that since we chose a fixed polynomial basis first, the matrices  $P_1$  and  $P_2$  are determined by this basis choice. Thus the system

$$\begin{pmatrix} A_{11} & A_{12} & P_1 \\ A_{12}^T & A_{22} & P_2 \\ P_1^T & P_2^T & 0 \end{pmatrix} \cdot \begin{pmatrix} C_1 \\ C_2 \\ b \end{pmatrix} = \begin{pmatrix} V_1 \\ V_2 \\ 0 \end{pmatrix}$$

has a nonsingular  $Q \times Q$  matrix  $P_1$ , thus by elimination, we get the following equations

$$\begin{aligned} C_1 &= -(P_1^T)^{-1} P_2^T C_2 \\ b &= P_1^{-1} (V_1 + A_{11} (P_1^T)^{-1} P_2^T C_2 - A_{12} C_2) \\ \tilde{A} C_2 &= V_2 - P_2 P_1^{-1} V_1 \end{aligned} \quad (5.22)$$

with the symmetric  $(N - Q) \times (N - Q)$  matrix

$$\tilde{A} := A_{22} - A_{12}^T (P_1^T)^{-1} P_2^T - P_2 P_1^{-1} A_{12} + P_2 P_1^{-1} A_{11} (P_1^T)^{-1} P_2^T. \quad (5.23)$$

These equations can be solved in reverse order due to

**Theorem 5.5.1.** *The matrix  $\tilde{A}$  is positive definite. The resulting matrices  $C_1$  and  $C_2$  are independent of the choice of basis in  $\mathcal{P}_m^d$ , as is  $\tilde{A}$ .*

**Proof:** For an arbitrary vector  $z \in \mathbb{R}^{N-Q}$  we get moment conditions

$$(P_1^T, P_2^T) \begin{pmatrix} -(P_1^T)^{-1} P_2^T z \\ z \end{pmatrix} = 0$$

and thus

$$\begin{pmatrix} -(P_1^T)^{-1} P_2^T z \\ z \end{pmatrix}^T A \begin{pmatrix} -(P_1^T)^{-1} P_2^T z \\ z \end{pmatrix} = z^T \tilde{A} z \geq 0$$

by conditional positive definiteness of the kernel. If the quadratic form is zero, then  $z = 0$ . A basis change in  $\mathcal{P}_m^d$  will not change  $P_2 P_1^{-1}$ , because  $P_1$  and  $P_2$  go over to  $P_1 C$  and  $P_2 C$  with a nonsingular  $Q \times Q$  matrix  $C$ .  $\square$

By a simple recalculation of (5.15), we also get

**Theorem 5.5.2.** *In the above form, the Gramian of a basis  $w$  is*

$$G_w = \tilde{C}_w^T \tilde{A} \tilde{C}_w$$

if we split

$$C_w = \begin{pmatrix} \hat{C}_w \\ \tilde{C}_w \end{pmatrix}. \quad \square$$

For a split

$$V_w = \begin{pmatrix} V_{1,w} \\ V_{2,w} \end{pmatrix}$$

and

$$\tilde{V}_w := V_{2,w} - P_2 P_1^{-1} V_{1,w}$$

we have

$$\tilde{A}\tilde{C}_w = \tilde{V}_w,$$

which is a generalization of the corresponding equation in the unconditional case. But note that both  $\tilde{C}_w$  and  $\tilde{V}_w$  cannot be inverted in general, since they are  $(N - Q) \times N$  matrices of rank  $N - Q$ .

Now let the points of a  $\mathcal{P}_m^d$ -unisolvent set  $X = \{x_1, \dots, x_N\}$  be ordered in such a way that the first  $Q$  points form a  $\mathcal{P}_m^d$ -unisolvent subset  $X_Q = \{x_1, \dots, x_Q\} \subset X$ , and assume that  $p_1, \dots, p_Q$  are a basis of  $\mathcal{P}_m^d$ , then

**Definition 5.5.1.** *Any data-dependent basis  $w$  consisting of  $w_1 = p_1, \dots, w_Q = p_Q$  and  $N - Q$  other basis functions  $w_{Q+1}, \dots, w_N$  is called a points and polynomials first (PPF) basis.*

We now proceed to characterize all PPF bases. We define  $P^T = (P_1^T, P_2^T)$  with matrices  $P_1 \in \mathbb{R}^{Q \times Q}$ ,  $P_2 \in \mathbb{R}^{(N-Q) \times Q}$  using our ordering, and let the value matrix of the complete basis take the form

$$V_w := \begin{pmatrix} P_1 & V_P \\ P_2 & V_R \end{pmatrix}$$

with matrices  $V_P \in \mathbb{R}^{Q \times (N-Q)}$ ,  $V_R \in \mathbb{R}^{(N-Q) \times (N-Q)}$ . By construction, the matrix  $P_1$  is nonsingular. When taking a polynomial Lagrange basis to start with, we get  $P_1 = I$ .

**Theorem 5.5.3.** *The identity (5.7) is necessarily of the form*

$$\begin{pmatrix} A_{11} & A_{12} & P_1 \\ A_{12}^T & A_{22} & P_2 \\ P_1^T & P_2^T & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & C_P \\ 0 & C_R \\ I & B \end{pmatrix} = \begin{pmatrix} P_1 & V_P \\ P_2 & V_R \\ 0 & 0 \end{pmatrix} \quad (5.24)$$

with a  $Q \times (N - Q)$  matrix  $B$  and a nonsingular  $(N - Q) \times (N - Q)$  matrix  $C_R$ . If  $C_R$  is fixed, all choices of either  $B$  or  $V_P$  are possible, and each such choice defines a PPF basis.

**Proof:** For a PPF basis, the system (5.7) has the general form

$$\begin{pmatrix} A_{11} & A_{12} & P_1 \\ A_{12}^T & A_{22} & P_2 \\ P_1^T & P_2^T & 0 \end{pmatrix} \cdot \begin{pmatrix} R & C_P \\ S & C_R \\ T & B \end{pmatrix} = \begin{pmatrix} P_1 & V_P \\ P_2 & V_R \\ 0 & 0 \end{pmatrix}$$

and consequently

$$\begin{aligned} A_{11}R + A_{12}S + P_1T &= P_1, \\ A_{12}^T R + A_{22}S + P_2T &= P_2, \\ P_1^T R + P_2^T S &= 0. \end{aligned}$$

A simple elimination then yields  $\tilde{A}S = 0$  with the matrix  $\tilde{A}$  from (5.23) which is nonsingular due to Theorem (5.5.1). Thus  $S = 0$ , and  $P_1$  is nonsingular, this implies that  $R = 0$  and  $T = I$ . It is necessary to prove (5.24).

By elimination in (5.24), we get necessary the equations

$$\begin{aligned} C_P &= -(P_1^T)^{-1}P_2^T C_R, \\ B &= P_1^{-1}(V_P + (A_{11}(P_1^T)^{-1}P_2^T - A_{12})C_R), \\ V_R &= \tilde{A}C_R + P_2P_1^{-1}V_P \end{aligned} \quad (5.25)$$

with the symmetric positive definite  $(N - Q) \times (N - Q)$  matrix  $\tilde{A}$  from (5.23). The nonsingularity of the value matrix implies that  $V_R - P_2P_1^{-1}V_P$  must be nonsingular, and since this is  $\tilde{A}C_R$ , we get that  $C_R$  necessarily is nonsingular. The above arguments can be pursued backwards to show that each nonsingular  $C_R$  and arbitrary choices of either  $B$  or  $V_P$  yield a PPF basis.  $\square$

For a split

$$V_w = \begin{pmatrix} P_1 & V_P \\ P_2 & V_R \end{pmatrix}$$

of the corresponding value matrix and

$$\tilde{V}_w := V_R - P_2P_1^{-1}V_P$$

we have

$$\tilde{A}C_R = \tilde{V}_w. \quad (5.26)$$

There is a particular special case:

**Definition 5.5.2.** A PPF basis is called canonical, if it satisfies  $V_P = 0$ .

For canonical PPF bases, we have the simple identity

$$\tilde{A}C_R = V_R \quad (5.27)$$

between nonsingular  $(N - Q) \times (N - Q)$  matrices. Either  $V_R$  or  $C_R$  can be prescribed for a canonical PPF basis to define it uniquely.

The identities (5.26) and (5.27) generalize what we had in the unconditionally positive definite case. It will turn out later that canonical PPF bases arise from using the *reduced kernel*.

In principle, one may put  $B := 0$ , and  $C_R := I$ . This special case of a PPF basis will be treated in the next section after a short detour.

**Theorem 5.5.4.** *All PPF bases  $w$  with orthonormal functions  $w_{Q+1}, \dots, w_N$  arise from a factorization*

$$\tilde{A} = (C_R^T)^{-1}(C_R)^{-1} = E^T E,$$

where  $E := C_R^{-1}$ . Then the value matrix is

$$\tilde{V}_w = (C_R^T)^{-1}$$

and for a canonical PPF basis we have  $V_R = (C_R^T)^{-1}$ .

**Proof:** By a simple recalculation of (5.15) using

$$C_w = \begin{pmatrix} 0 & C_P \\ 0 & C_R \end{pmatrix},$$

we get the Gramian of a PPF basis  $w$  to be

$$G_w = C_R^T \tilde{A} C_R.$$

If this is the identity, the first assertion follows. The rest is a consequence of (5.26) and (5.27).  $\square$

This is similar to the results in (3.6) for the unconditionally positive definite case. There different decompositions like Cholesky and SVD were used to construct orthonormal bases. Here, partially orthonormal PPF bases arise from SVD or Cholesky decompositions of the matrix  $\tilde{A}$ .

We now want to determine the discretely orthonormal PPF bases. If we split the value matrix  $V_w$  by reordering the  $Q$  unisolvent points, then

$$V_w = \begin{pmatrix} P_1 & V_P \\ P_2 & V_R \end{pmatrix}.$$

To construct all discretely orthonormal PPF bases, we start from  $H = V_w^T V_w = I$  and get

$$\begin{aligned} P_1^T P_1 + P_2^T P_2 &= I \\ V_P^T V_P + V_R^T V_R &= I \\ P_1^T V_P + P_2^T V_R &= 0, \end{aligned} \tag{5.28}$$

where  $P_1$  is nonsingular. Then from (5.26) we get

$$\tilde{A} C_R = (I + P_2(P_1)^{-1}(P_1^T)^{-1}P_2^T)V_R,$$

which is another generalization of what we know in the unconditionally positive definite case.

**Theorem 5.5.5.** *Canonical PPF bases are not discretely orthonormal.*

**Proof:** Canonical PPF bases have  $V_P = 0$ . Then  $V_R$  is nonsingular, and the system (5.28) shows that  $P_2$  must be zero. Thus all polynomials must be zero on all points, contradicting unisolvency.  $\square$

## 5.6 Standard Basis

Here in this section, we would like to use the  $K(x, x_j)$  as a basis, but we have to obey the moment conditions. We order the points in such a way that the first  $Q$  are unisolvent, and we let  $p_1, \dots, p_Q$  be a Lagrange basis of polynomial interpolation on  $x_1, \dots, x_Q$ . Then we can reproduce

$$p(y) = \sum_{m=1}^Q p(x_m) p_m(y) \text{ for all } p \in \mathcal{P}, y \in \Omega,$$

and for the set  $\{y\} \cup \{x_1, \dots, x_Q\}$  we have a coefficient vector

$$1, -p_1(y), \dots, -p_Q(y)$$

that satisfies the moment conditions on that set, i.e.

$$1 \cdot p(y) - \sum_{m=1}^Q p(x_m) p_m(y) = 0 \text{ for all } p \in \mathcal{P}.$$

This allows to define the functions

$$s_y(x) := K(x, y) - \sum_{m=1}^Q p_m(y) K(x, x_m), \quad x \in \Omega$$

for all  $y \in \Omega \setminus \{x_1, \dots, x_Q\}$  which are in  $\mathcal{K}_X$ . We use this to write down a generalization of the usual reproduction formula. To this end, we take an arbitrary  $s_{c,b} \in \mathcal{K}_X$  of the form (2.17) and evaluate

$$\begin{aligned} (s_{c,b}, s_y) &= \sum_{i=1}^N c_i \left( K(x_i, y) - \sum_{m=1}^Q p_m(y) K(x_i, x_m) \right) \\ &= \sum_{i=1}^N c_i K(x_i, y) - \sum_{m=1}^Q p_m(y) \sum_{i=1}^N c_i K(x_i, x_m) \\ &= s_{c,b}(y) - p_b(y) - \sum_{m=1}^Q p_m(y) (s_{c,b}(x_m) - p_b(x_m)) \\ &= s_{c,b}(y) - \sum_{m=1}^Q p_m(y) s_{c,b}(x_m), \end{aligned}$$

Then a standard basis generalizing the translates  $K(x, x_j)$  consists of  $p_1, \dots, p_Q$  and of simple linear combinations of translates of  $K$ , i.e.

$$s_j(x) := K(x, x_j) - \sum_{m=1}^Q p_m(x_j) K(x, x_m), \quad Q+1 \leq j \leq N.$$

If we add  $s_j := p_j$  for  $1 \leq j \leq Q$ , we have  $N$  functions and need to prove linear independence.

**Theorem 5.6.1.** *This set of functions is a basis.*

**Proof:** Define  $P^T = (I_{Q \times Q}, P_2^T)$  due to the Lagrange property on the first  $Q$  points. We split  $A$  and  $\mathcal{A}$  to get

$$\mathcal{A} = \begin{pmatrix} A_{11} & A_{12} & I \\ A_{12}^T & A_{22} & P_2 \\ I & P_2^T & 0 \end{pmatrix}.$$

The value matrix has  $P$  as its left  $N \times Q$  submatrix. The right  $N \times (N - Q)$  submatrix is

$$\begin{pmatrix} A_{12} \\ A_{22} \end{pmatrix} - \begin{pmatrix} A_{11} \\ A_{12}^T \end{pmatrix} \cdot P_2^T$$

due to the definition of the basis. The identity (5.7) then is

$$\begin{pmatrix} A_{11} & A_{12} & I \\ A_{12}^T & A_{22} & P_2 \\ I & P_2^T & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & -P_2^T \\ 0 & I \\ I & 0 \end{pmatrix} = \begin{pmatrix} I & A_{12} - A_{11} P_2^T \\ P_2 & A_{22} - A_{12}^T P_2^T \\ 0 & 0 \end{pmatrix}$$

which could also have been obtained via the coefficient matrices and (5.10). The first matrix is nonsingular, the second has rank  $N$ , and thus the value matrix has rank  $N$ .  $\square$

We see that we have got a PPF basis with  $B = 0$  and  $C_R = I$ . In terms of the split (5.12), the standard basis satisfies

$$\begin{aligned} s_j &= 0 & + & s_j^P, & 1 \leq j \leq Q, \\ s_j &= s_j^K & + & 0, & Q + 1 \leq j \leq N. \end{aligned}$$

The Gramian follows from (5.15) and takes the form

$$\begin{pmatrix} 0 & 0 \\ 0 & \tilde{A} \end{pmatrix}$$

with the symmetric matrix

$$\tilde{A} := P_2 A_{11} P_2^T - P_2 A_{12} - A_{12}^T P_2^T + A_{22}$$

that can later be interpreted as a kernel matrix of the reduced kernel. This generalizes the fact that the kernel matrix  $A$  itself is the Gramian of the standard basis in the unconditionally positive definite case see (3.6).

## 5.7 Lagrange Basis

The Lagrange basis  $u_1(x), \dots, u_N(x)$  has the value matrix  $V_u = I$  and is defined via

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} u(x) \\ v(x) \end{pmatrix} = \begin{pmatrix} K_X(x) \\ p(x) \end{pmatrix}$$

with

$$\begin{aligned} K_X(x)^T &:= (K(x_1, x), \dots, K(x_Q, x)) \in \mathbb{R}^N, \\ p(x)^T &:= (p_1(x), \dots, p_Q(x)) \in \mathbb{R}^Q, \\ u(x)^T &:= (u_1(x), \dots, u_N(x)) \in \mathbb{R}^N, \\ v(x)^T &:= (v_1(x), \dots, v_Q(x)) \in \mathbb{R}^Q. \end{aligned}$$

The construction matrices of the Lagrange basis follow from Theorem (5.2.2) by setting  $V_w = I$ . For practical calculations, one can also use the equations (5.22) backwards. The same equations yield the Gramian of the Lagrange basis as the symmetric positive semidefinite  $N \times N$  matrix

$$(-P_2 P_1^{-1}, I)^T \tilde{A}^{-1} (-P_2 P_1^{-1}, I)$$

of rank  $N - Q$ . Note that  $A^{-1}$  is the Gramian of the Lagrange basis in the unconditionally positive definite case.

We now look at arbitrary other bases  $w = (w_1, \dots, w_N)$  with nonsingular value matrices

$$V_w := (w_j(x_k))_{1 \leq j, k \leq N}.$$

Here and elsewhere we stick to the notation that points vary with the row index and functions vary with the column index. Then we see that

$$(w_1(x), \dots, w_N(x)) = (u_1(x), \dots, u_N(x)) V_w$$

is the way the basis is composed out of the Lagrange basis.

It is tempting to ask for PPF Lagrange bases. But:

**Theorem 5.7.1.** *For  $N > Q$  there is no PPF Lagrange basis.*

**Proof:** If there were a PPF Lagrange basis, we must have  $P_2 = 0$  in (5.24). But then all polynomials including 1 must vanish on  $x_{Q+1}, \dots, x_N$ .  $\square$

But there clearly is a canonical PPF basis that is partially Lagrange in the sense that  $V_R = I$  and  $V_P = 0$ . It has  $C_R = \tilde{A}^{-1}$  by (5.27), and the equations (5.25) yield  $B$  and  $C_P$  for this case. Since not necessarily  $P_2 = 0$ , the basis is not Lagrange in the true sense.

In terms of the Lagrange basis, we can form the *power kernel* [17]

$$\begin{aligned} K_X(x, y) &:= K(x, y) - \sum_{i=1}^N u_i(x) K(x_i, y) - \sum_{i=1}^N K(x, x_i) u_i(y) \\ &\quad + \sum_{i,j=1}^N K(x_i, x_j) u_i(x) u_j(y) \\ &= K(x, y) - u^T(x) K_X(y) - u^T(y) K_X(x) + u^T(x) A u^T(y) \end{aligned}$$

which is symmetric and positive definite on  $\Omega \setminus X$  and vanishes when one of its arguments is in  $X$ . The quadratic term can be eliminated using

$$u^T(x)Au(y) = u^T(x)(K_X(y) - Pv(y)) = u^T(x)K_X(y) - p^T(x)v(y)$$

at the expense of bringing polynomials into the representation.

The *power function* on  $X$  then is defined via

$$P_X^2(x) := K_X(x, x) \text{ for all } x \in \Omega.$$

## 5.8 Newton Basis, Iterative Construction

We start with  $X_Q := \{x_1, \dots, x_Q\}$  to be unisolvent, and there we take the standard Lagrange basis  $v_1, \dots, v_Q$ . We need all values  $v_i(x_j)$ ,  $1 \leq j \leq N$ ,  $1 \leq i \leq Q$  for further calculation. We already have  $v_i(x_j) = 0$  for  $1 \leq i < j \leq Q$ . We shall need the definition of the reduced kernel

$$\begin{aligned} K_Q(x, y) &:= K(x, y) - \sum_{j=1}^Q v_j(x)K(y, x_j) - \sum_{k=1}^Q v_k(y)K(x, x_k) \\ &\quad - \sum_{j=1}^Q \sum_{k=1}^Q v_j(x)v_k(y)K(x_j, x_k) \end{aligned}$$

which is symmetric and unconditionally positive semi-definite kernel on  $\Omega \setminus X_Q$  and vanishes if one of the arguments is in  $X_Q$ .

We store the actual values of the power function

$$P_Q^2(x) := K_Q(x, x)$$

on the points of  $X$ . They are zero on  $X_Q$ . For what follows, we start with  $m := Q$ . Now, for induction, we assume that we already have a construction of a basis  $v_1, \dots, v_m$  for some  $Q \leq m < N$  with the orthonormality properties

$$(v_i, v_j) = \delta_{ij}, \quad Q + 1 \leq i, j \leq m.$$

Formally, we also assume that we have the values of the power function  $P_m(x)$  on  $X$ . Note that the associated power kernel satisfies the recursion

$$K_m(x, y) = K_{m-1}(x, y) - \frac{K_{m-1}(x, x_m)K_{m-1}(y, x_m)}{K_{m-1}(x_m, x_m)}$$

for  $m > Q$  (see [21, 17]) and has the properties

$$K_m(x_j, y) = K_m(x, x_j) = 0, \quad 1 \leq j \leq m, \quad P_m^2(x) = K_m(x, x) \text{ for all } x, y \in \Omega.$$

Then we define

$$x_{m+1} := \arg \max \{P_m(x) : x \in X\}$$

and stop if  $P_m(x_{m+1})$  is zero or very small, because then we are done. Now we formally define

$$v_{m+1}(x) := \frac{K_m(x, x_{m+1})}{\sqrt{K_m(x_{m+1}, x_{m+1})}} \text{ for all } x \in \Omega$$

to get that

$$v_{m+1}(x_j) = 0, \quad 1 \leq j \leq m,$$

as required. But we have to show how  $v_{m+1}$  and  $P_{m+1}$  can be calculated efficiently on  $X$ .

We see immediately that

$$P_{m+1}^2(x) = P_m^2(x) - v_{m+1}^2(x) = P_Q^2(x) - \sum_{j=Q+1}^{m+1} v_j^2(x)$$

holds by construction and induction. Thus we only need  $v_{m+1}$  on  $X$ . We consider the recursion

$$\begin{aligned} K_j(x, y) &= K_{j-1}(x, y) - \frac{K_{j-1}(x, x_j)K_{j-1}(y, x_j)}{K_{j-1}(x_j, x_j)} \\ &= K_{j-1}(x, y) - v_j(x)v_j(y) \end{aligned}$$

that comes to

$$v_{m+1}(x) = K_Q(x, x_{m+1}) - \sum_{j=Q+1}^m v_j(x_{m+1})v_j(x)$$

which is computable from the values  $K_Q(x, x_{m+1})$  for all  $x \in X$ . But these are obtainable from

$$\begin{aligned} K_Q(x, x_{m+1}) &= K(x, x_{m+1}) - \sum_{j=1}^Q v_j(x)K(x_{m+1}, x_j) - \sum_{k=1}^Q v_k(x_{m+1})K(x, x_k) \\ &\quad - \sum_{j=1}^Q \sum_{k=1}^Q v_j(x)v_k(x_{m+1})K(x_j, x_k) \end{aligned}$$

at reasonable cost. It remains to show that

$$(v_{m+1}, v_j) = 0, \quad Q+1 \leq j \leq m,$$

and this follows from

$$(K_Q(x, x_{m+1}), v_j) = v_j(x_{m+1}), \quad Q+1 \leq j \leq m$$

because these  $v_j$  vanish on  $x_1, \dots, x_Q$ . Furthermore,

$$\begin{aligned} P_m^2(x_{m+1})(v_{m+1}, v_{m+1}) &= (K_m(x, x_{m+1}), K_m(x, x_{m+1})) \\ &= K_m(x_{m+1}, x_{m+1}) \\ &= P_m^2(x_{m+1}) \end{aligned}$$

by what is known for power kernels from [17]. This proves orthonormality. The value matrix of this canonical PPF basis is

$$\begin{pmatrix} I & 0 \\ P_2 & L \end{pmatrix}$$

with a lower triangular value matrix  $L = V_R$ , and we have orthonormality of  $v_{Q+1}, \dots, v_N$ . Now Theorem (5.5.4) implies that we have  $C_R = (L^{-1})^T$  and a Cholesky decomposition  $\tilde{A} = LL^T$ . This proves

**Theorem 5.8.1.** *The above construction generates the canonical PPF basis with  $P_1 = I$ , which coincides with the extension of the Lagrange basis of  $\mathcal{P}_m^d$  by the Newton basis for  $X \setminus X_Q$  and the reduced kernel.  $\square$*

If we use the reduced kernel for constructing a Lagrange basis on  $X \setminus X_Q$  and use a basis of  $\mathcal{P}_m^d$  for the first  $Q$  basis functions, we get the canonical PPF and partially Lagrange basis that we mentioned at the end of Section 5.7.

## 5.9 Dual Bases

In the unconditionally positive definite case, it was easy to define a *dual basis* to a given basis  $w$  by taking the Riesz representers of the functionals defined in (5.13). This is not possible in the conditionally positive definite case, since the Riesz representation theorem needs a Hilbert space and can be only applied to functionals  $\lambda_c$  as in (5.4) that satisfy moment conditions, i.e. with  $c \in M$ . Thus, after starting from a basis  $w$  and forming the  $N$  linearly independent functionals in (5.13), one has to proceed towards a subset of  $N - Q$  linearly independent functionals that satisfy the usual  $Q$  moment conditions.

This can be done in different ways. We start with the PPF situation and assume that the space  $\mathcal{P}_m^d$  is spanned by a basis  $p_1, \dots, p_Q$  on a minimal  $\mathcal{P}_m^d$ -unisolvent set  $X_Q = \{x_1, \dots, x_Q\} \in \Omega$ . Then there is a linear projector  $\Pi_Q$  onto  $\mathcal{P}_m^d$  defined by

$$(\Pi_Q(f))(x) := \sum_{m=1}^Q f(x_m)p_m(x), \quad \text{for all } x \in \Omega$$

and each functional  $\lambda$  can be turned into another that vanishes on  $\mathcal{P}_m^d$  by defining

$$(\Pi_Q^*(\lambda))(f) := \lambda(f - \Pi_Q(f)).$$

If we do this for the  $N$  linear functionals of (5.13), we get  $N$  functionals

$$\mu_k(f) := \lambda_k(f - \Pi_Q(f)), \quad 1 \leq k \leq N$$

which will have Riesz representers  $v_1, \dots, v_N$  of the form (5.3).

**Theorem 5.9.1.** *These functions can be expressed via the reduced kernel  $\tilde{K}$  as*

$$v_k(x) = \mu_k^y \tilde{K}(x, y) = \lambda_k^y \tilde{K}(x, y), \quad 1 \leq k \leq N, \quad x \in \Omega.$$

**Proof:** On each function  $s_c$  of the form (5.3) we have

$$(s_c, v_k) = \mu_k(s_c) = \lambda_k(s_c - \Pi_Q(s_c))$$

and consequently

$$\begin{aligned} v_k(x) &= (v_k, \tilde{K}(x, \cdot)) \\ &= \mu_k^y \tilde{K}(x, y) \\ &= \lambda_k^y (\tilde{K}(x, y) - \Pi_Q(\tilde{K}(x, \cdot))(y)) \\ &= \lambda_k^y \tilde{K}(x, y). \end{aligned}$$

□

Clearly, the functions  $v_1, \dots, v_N$  will only span a space of dimension  $N - Q$ , since they are Riesz representers of  $N$  functionals satisfying moment conditions. They vanish on the first  $Q$  points, and need to be augmented by certain polynomials.

## 5.10 Numerical Tests

We consider the domain  $\Omega$  similar to the one in Section 4.3, defined by the unit disk with the third quadrant cut away, and select a large set  $X$  of points on a fine grid on  $[-1, 1]^2$  that fall into  $\Omega$ . We ran the algorithm for all cases of bases that are mentioned in this chapter. For illustration, we take the thin-plate spline kernel which is conditionally positive definite of order 2. As a fixed set  $X$  to start with, we construct a unisolvent set for  $\mathcal{P}_4^2$  of 10 points by a pivoted QR decomposition of the value matrix of polynomials on a fine point set, see Figure (5.1). This set, when used for order  $m = 2$ , has a unisolvent subset of 3 points that we determine similarly. We reorder the data sites to have the unisolvent set in front. Then we calculate the 3 polynomial Lagrange basis functions for the unisolvent subset for further internal use (see Figure (5.2) on the left for one of these Lagrangians). One

of the full Lagrange basis functions based on Section 5.7 is in Figure (5.2) on the right. The standard basis of section 5.6 vanishes on the unisolvent set, see Figure (5.3) on the left, while on the right there is a case of the “projector-based” strategy of (5.17). For the partially orthonormal bases see Figure (5.4), in the Cholesky and the SVD case. Finally, Figure (5.6) shows one of the adaptively constructed Newton basis functions for 25 points and the associated power function based and selection points in (5.5) on Section 5.8. Note that this starts from the first  $Q = 3$  fixed points and then adaptively selects points from the fine set in Figure (5.1).

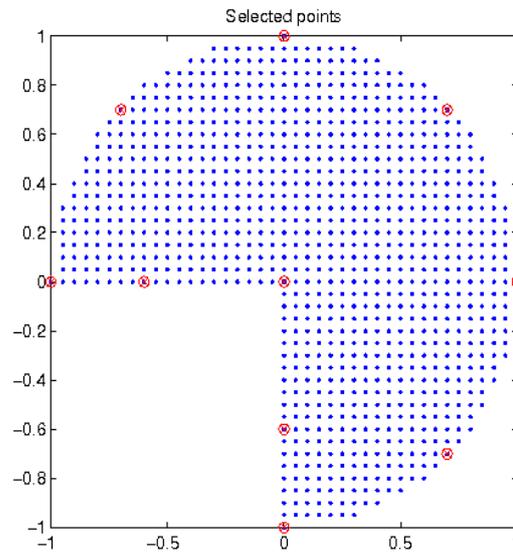


Figure 5.1: Selected 10 interpolation points

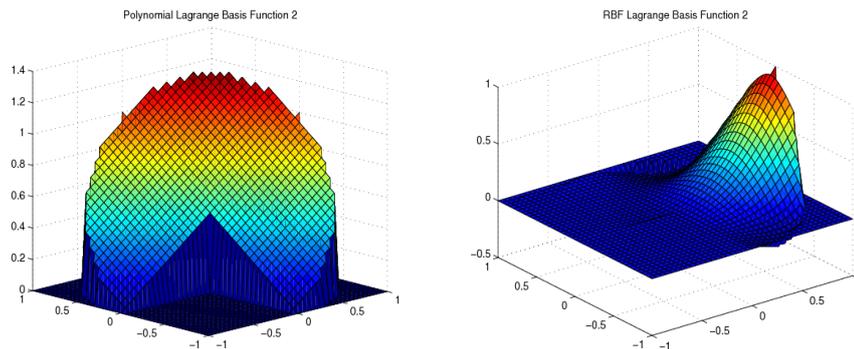


Figure 5.2: Polynomial Lagrange basis function and RBF Lagrange basis function

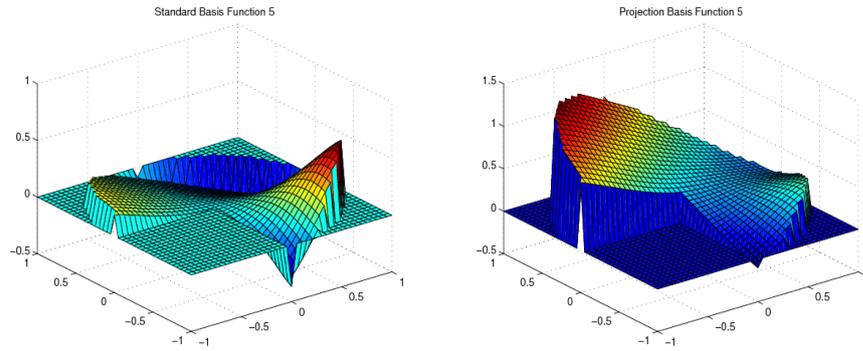


Figure 5.3: Standard basis function and projector basis function

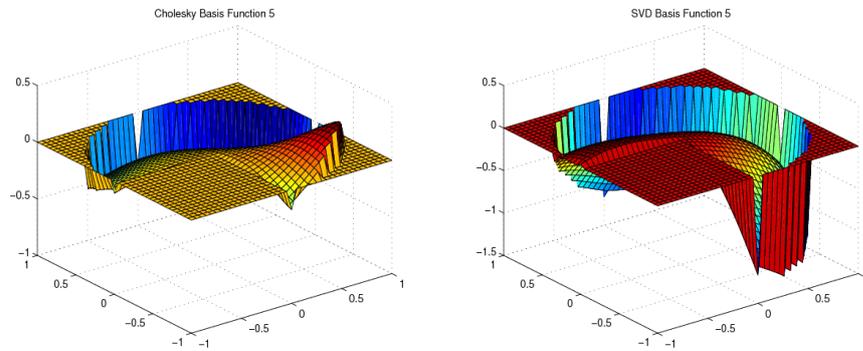


Figure 5.4: Cholesky basis and SVD basis

## 5.11 Notes and Comments

Altogether, we provided data-dependent bases for  $\mathcal{K}_X$  and prescribed their relations between these bases. The results cover all kernels in practical use. There are unexpectedly many choices how to proceed to different bases from starting with the general form of bases. For instance, preselecting points leads to basis functions containing polynomial parts, or using a reduced kernel allows reduction to the unconditional case on a subset of points. Moreover, we investigated bases of Lagrange or Newton type, and orthogonal or orthonormal bases. Finally, one can reduce the conditional positive definite case completely to the unconditionally positive definite case by use of the extended kernel, but we have not commented on this in detail.

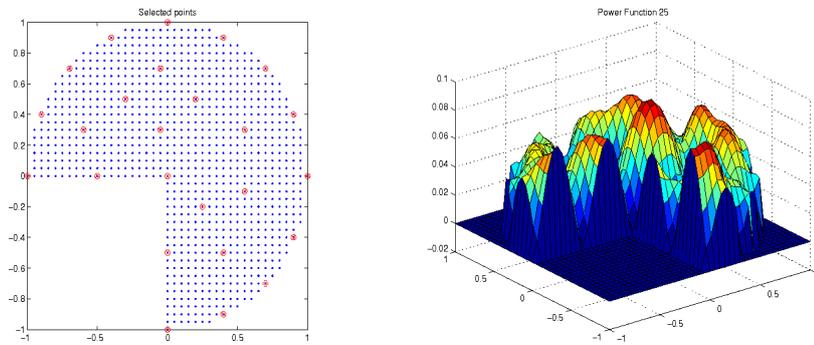
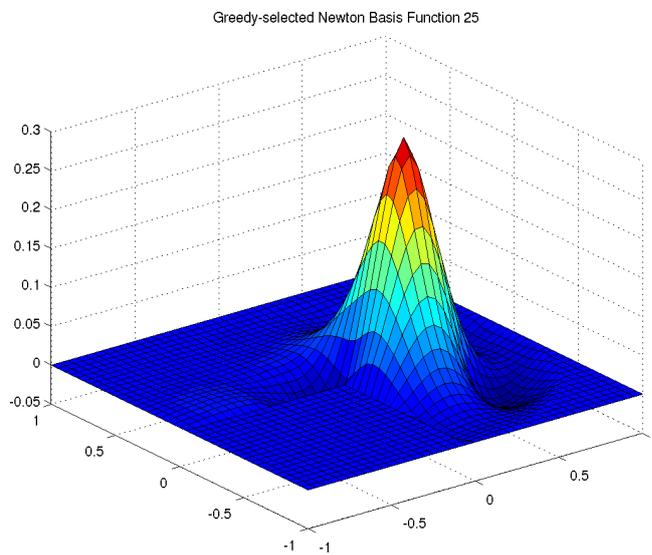


Figure 5.5: Point selection and power function

Figure 5.6: Greedy-selected Newton basis  $v_{25}$



# Conclusion

---

We considered general bases, while we specialized to  $\mathcal{K}$ -orthonormal and discretely orthonormal bases in Sections 3.4, 3.5 and 5.3.1. Our goal is to sort all possible bases into certain classes and to prove as much as we can about their properties and their relations. This leads us to consider different data-dependent bases and to compare them with respect to stability, recursive computability, duality, and orthogonality properties. We now have a thorough account of the possibilities to construct bases of data-dependent subspaces of reproducing kernel Hilbert spaces. The Newton basis seems to be a particularly good choice, since it is  $\mathcal{K}$ -orthonormal, allows stable evaluation in the sense of (3.10) and also they can be calculated recursively and adaptively.

But if no data set is specified, we should also consider “data-independent” bases. The most natural is the *eigenfunction basis* coming from the Mercer theorem. We described it in Section 2.2 and compared it with data-dependent bases. Of course, data-independent bases are much more “natural” than data-dependent bases, and they provide intrinsic information about the kernel.

We can investigate the connection of the Newton basis to Expansion kernels. Assume that a positive definite kernel has a rapidly converging series expansion

$$K(x, y) = \sum_{k=0}^{\infty} \lambda_k \phi_k(x) \phi_k(y),$$

like in section 2.2. Then the kernel on finite point set  $X$  with  $N$  points will be represented by

$$\begin{aligned} K_N(x_i, x_j) &= \sum_{k=0}^N \lambda_k \phi_k(x_i) \phi_k(x_j) \\ &= e_j^T \Phi_{X,N} \Lambda_N \Phi_{X,N}^T e_k, \end{aligned}$$

with matrix  $\Phi_{X,N} = (\phi_j(x_i))_{0 \leq i, j \leq N}$ , and diagonal matrix  $\Lambda_N$  with element  $\lambda_0, \dots, \lambda_N$  on diagonal. We obtain the kernel matrix

$$A_X = \Phi_{X,N} \sqrt{\Lambda_N} \sqrt{\Lambda_N} \Phi_{X,N}^T,$$

showing the connection to the general bases were already defined by decomposition of the kernel matrix. Particularly, we use the Cholesky decomposition  $A_X = R^T R$ , and take the QR decomposition of  $\sqrt{\Lambda_N} \Phi_{X,N}^T = R$ ,

then

$$R = \Phi_{X,N} \sqrt{\Lambda_N} Q,$$

holds. Consequently, the Newton basis is related to the eigenfunction basis, the only difference is by a scaling and orthogonal transformation in function space. This can be also investigated for the other choices of data-dependent bases. In case of data-independent bases, problems may arise when working on a set  $X$  of centers later, because the Mairhuber–Curtis theorem [15] shows that matrices of values of the basis on points of  $X$  have a data-dependent rank, if one is in a truly multivariate and nontrivial situation. This is why we used the data-independent eigenfunction basis only for theoretical comparisons.

An important question for future studies is to construct bases iteratively via Krylov subspace methods and using neighboring points. Such algorithms can be very effective, in particular in cases where local Lagrangian basis functions decay exponentially, e.g. for thin-plate splines. We refer to [12] for details. Also, we neglect bases that come from special preconditioning techniques. Instead, we mainly focused here on bases that can be derived by standard linear algebra operations on the usual kernel matrix. Consequently, a future approach can be to construct bases by using preconditioners for the kernel matrix. In many algorithms it is necessary to solve a linear system  $AX = Y$  with the kernel matrix  $A$ . In numerical analysis, a preconditioner  $T$  of a matrix  $A$  is a matrix such that  $T^{-1}A$  has a smaller condition number than  $A$ . Which means one can solve the preconditioned system

$$T^{-1}(AX - Y) = 0,$$

instead of solving the original linear system above.

This implies the same solution as the original system so long as the preconditioner matrix  $T$  is nonsingular. Usually the preconditioned matrix is computed in iterative methods. Since the operator  $T^{-1}$  must be applied at each step of the iterative linear solver, it should have a small cost (computing time) of applying the  $T^{-1}$  operation. Preconditioned iterative methods for  $AX - Y = 0$  are, in most cases, mathematically equivalent to standard iterative methods applied to the preconditioned system  $T^{-1}(AX - Y) = 0$ . Examples of popular preconditioned iterative methods for linear systems include the *preconditioned conjugate gradient method*, the *biconjugate gradient method*, and the *generalized minimal residual method*. Iterative methods, which use scalar products to compute the iterative parameters, require corresponding changes in the scalar product together with substituting  $T^{-1}(AX - Y) = 0$  for  $AX - Y = 0$ .

Popular preconditioners like Jacobi, SSOR and ILU can be used to improve the convergence, however, these preconditioners for  $N$  points in the domain have an  $\mathcal{O}(N^2)$  space and computation requirement for dense matrices, which would ruin any advantage gained by the fast matrix-vector products. A good preconditioner will improve the convergence of the iterative

approach at the expense of an increased cost per iteration. A preconditioner is useful in case that the total time taken by the preconditioned approach should be less than the unpreconditioned approach. This is achieved by using fast matrix-vector product algorithms, resulting in a computationally efficient solver with faster convergence.

Consequently, preconditioners are useful in iterative methods to solve a linear system, since the rate of convergence for most iterative linear solvers increases as the condition number of a matrix decreases as a result of preconditioning. Iterative Krylov approaches are often used with fast matrix vector products for efficient solution. When the underlying system is ill-conditioned, there is a degradation of the performance of iterative approaches, necessitating the use of a preconditioner for the Krylov iterations. alternatively, preconditioned Krylov methods can be considered as accelerations. This also can be investigated for the conditional case, especially it can be crucial for the iteratively constructed conditional Newton basis.



# Bibliography

- [1] *Computer experiments. In Design and analysis of experiments.* Handbook of Statist., North-Holland, Amsterdam.
- [2] T. Belytschko, Y. Krongauz, D.J. Organ, M. Fleming, and P. Krysl. Meshless methods: an overview and recent developments. *Computer Methods in Applied Mechanics and Engineering, special issue*, 139:3–47, 1996.
- [3] M. D. Buhmann. *Radial Basis Functions.* Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2004.
- [4] M.D. Buhmann. New developments in the theory of radial basis function interpolation. In K. Jetter and F.I. Utreras, editors, *Multivariate Approximations: From CAGD to Wavelets*, pages 35–75. World Scientific, Singapore, 1993.
- [5] M.D. Buhmann. Radial basis functions. *Acta Numerica*, 10:1–38, 2000.
- [6] N. Cressie, editor. *Multivariate Geostatistics: An Introduction with Applications.* Wiley and Stons, New–York, 1993.
- [7] C. de Boor. What is the inverse of a basis? *BIT*, 41:880–890, 2001. [ftp.cs.wisc.edu/Approx/inverse\\_basis.pdf](ftp.cs.wisc.edu/Approx/inverse_basis.pdf).
- [8] St. De Marchi and R. Schaback. Stability of kernel-based interpolation. *Adv. in Comp. Math.*, 32:155–161, 2010.
- [9] J. Duchon. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *Rev. Française Automat. Informat. Rech. Opér. Anal. Numer.*, 10:5–12, 1976.
- [10] N. Dyn. Interpolation of scattered data by radial functions. In C.K. Chui, L.L. Schumaker, and F.I. Utreras, editors, *Topics in Multivariate Approximation*, pages 47–61. Academic Press, Boston, 1987.
- [11] G. F. Fasshauer. *Meshfree Approximation Methods with MATLAB*, volume 6 of *Interdisciplinary Mathematical Sciences.* World Scientific Publishers, Singapore, 2007.
- [12] A. C. Faul and M. J. D. Powell. Proof of convergence of an iterative technique for thin plate spline interpolation in two dimensions. *Appl. Math. Comp.*, 11(2-3):183–192, 1998.

- [13] R.L. Hardy. Theory and applications of the multiquadric-biharmonic method: 20 years of discovery. *Comput. Math. Appl.*, 19:163–208, 1990.
- [14] R. Schaback M. Pazouki. Bases for kernel-based spaces. *Computational and Applied Mathematics*, 236(4):575–588, 2011.
- [15] J. C. Mairhuber. On Haar’s theorem concerning Chebychev approximation problems having unique solutions. *Proc. Amer. Math. Soc.*, 7:609–615, 1956.
- [16] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, pages 3397–3415, 1993.
- [17] M. Mouattamid and R. Schaback. Recursive kernels. *Analysis in Theory and Applications*, 25:301–316, 2009.
- [18] St. Müller and R. Schaback. A Newton basis for kernel spaces. *Journal of Approximation Theory*, 161:645–655, 2009. doi:10.1016/j.jat.2008.10.014.
- [19] M.J.D. Powell. The theory of radial basis function approximation in 1990. In W.A. Light, editor, *Advances in Numerical Analysis II: Wavelets, Subdivision Algorithms, and Radial Basis Functions*, pages 105–210. Oxford Univ. Press, Oxford, 1992.
- [20] R. Schaback. Creating surfaces from scattered data using radial basis functions. In T. Lyche M. Daehlen and L.L. Schumaker, editors, *Mathematical Methods for Curves and Surfaces*, pages 477–496. Vanderbilt University Press, Nashville, TN, 1995.
- [21] R. Schaback. Reconstruction of multivariate functions from scattered data. Manuscript, available via <http://www.num.math.uni-goettingen.de/schaback/research/group.html>, 1997.
- [22] R. Schaback. Improved error bounds for scattered data interpolation by radial basis functions. *Mathematics of Computation*, 68:201–216, 1999.
- [23] R. Schaback. Kernel-based meshless methods. Technical report, Numerical And Applied Mathematics Department, University of Göttingen, 2011.
- [24] R. Schaback and H. Wendland. Numerical techniques based on radial basis functions. In A. Cohen, C. Rabut, and L.L. Schumaker, editors, *Curve and Surface Fitting*, pages 359–374. Vanderbilt University Press, 2000.

- [25] R. Schaback and H. Wendland. Characterization and construction of radial basis functions. In N. Dyn, D. Leviatan, D. and Levin, and A. Pinkus, editors, *Multivariate Approximation and Applications*, pages 1–24. Cambridge University Press, 2001.
- [26] H. Wackernage, editor. *Multivariate Geostatistics: An Introduction with Applications*. Springer, 2003.
- [27] H. Wendland. *Scattered Data Approximation*. Cambridge University Press, 2005.
- [28] H. Wendland, editor. *Spatial coupling in aeroelasticity by meshless kernel-based methods*. ECCOMAS CFD, 2006.