Reconstruction of Multivariate Functions from Scattered Data

R. Schaback

January 21, 1997

Contents

1 Introduction **6**
1.1 Recovery, Interpolation, and Approximation 7
1.2 Input and Output Data 7
1.3 Restrictions on the Choice of Spaces 8
1.4 Data-dependent Spaces 9
1.5 Evaluation, Interpolation and Approximation 11
1.6 Conditionally Positive Definite Functions 13
1.7 Basic Equations for Conditionally Positive Definite Functions . 14

2 Working with Basis Functions **15**
2.1 General Practical Considerations 16
2.2 Current Basis Functions 19
2.3 Computational Complexity of Solving the System 20
2.4 Multilevel Algorithms 21
2.5 Numerical Examples 22

3 Hilbert Space Theory **26**
3.1 Optimal Recovery in Hilbert Spaces 28
3.2 Spaces of Functions 42
3.3 Native Spaces 54
3.4 Standardized Notation 65
3.5 Restrictions, Extensions, and Infinite Problems 67
3.6 Example: Cubic Splines in One Variable 75

4 Power Functions and Applications **80**
4.1 Power functions 80
4.2 Optimal Recovery Redefined 82
4.3 Example: Optimal Interpolation in $W_\infty^1[a,b]$ 83
4.4 Recursive Constructions 93
4.5 Condition 97
4.6 Uncertainty Relation 99

5 Error Bounds **101**
5.1 Upper Bounds for the Optimal Power Function 101
5.2 Approximation Error in the Radial Case 104
5.3 Jackson-Bernstein Theorems and Applications 107
5.4 Lebesgue Constants 111
5.5 Error Bounds in Terms of Fill Distance 120
5.6 Doubling the Approximation Order 123

5.7 Improvement by Localization 124

6 Advanced Results on $\mathbb{R}^d$ 124

6.1 Transforms of Translation-Invariant Basis Functions 124
6.2 Connection to $L_2(\mathbb{R}^d)$ 128
6.3 Characterization of Native Spaces 130
6.4 Condition Numbers 133
6.5 Error Bounds 139
6.6 Error Bounds and Scaling 142

7 Special Theory 142

7.1 Results for General Transforms 142
7.2 Theory on the Torus using Fourier Series 145
7.3 Theory on Spheres using Expansions 145
7.4 Lower Bounds for Eigenvalues 145
7.5 Generalizations of Results Using Transforms 146

8 Theory on Grids 146

8.1 Sampling Theory 146
8.2 Strang-Fix Theory 146
8.3 Application to Radial Basis Functions 146
8.4 Shift Invariant Spaces 146

9 Construction of Positive Definite Functions 146

9.1 General Construction Techniques 146
9.2 Construction of Positive Definite Radial Functions on $\mathbb{R}^d$ . . . 148
9.3 Positive Definite Functions on Topological Groups 171
9.4 Positive Definite Zonal Functions on Spheres 171

10 Special Algorithms 171

10.1 Reduction of Enlarged System, Method 1 171
10.2 Reduction of Enlarged System, Method 2 173

11 Computational Geometry Techniques 176

11.1 Voronoi Diagrams 176

12 Analytic Background 176

12.1 Calculus Facts 176
12.2 Hilbert Space Basics 177
12.3 Special Functions and Transforms 186
12.4 Lebesgue Integration 194
12.5 Fourier Transforms on $\mathbb{R}^d$ 197

12.6 Sobolev Spaces 206

13 Appendix 207

13.1 Basis Functions 207

13.2 MATLAB routines 208

List of Figures

1 Franke's function 24

2 Reconstruction of Franke's function from thin-plate splines on $M = 81$ points 24

3 Residuals for thin-plate splines on $M = 81$ points 24

4 Recovery using Wendland's C^2 function with support radius 2 on $M = 81$ points 24

5 Residuals using Wendland's C^2 function with support radius 2 on $M = 81$ points 24

6 Recovery using Wendland's C^2 function with bandwidth 21 on $M = 289$ points 24

7 Residuals using Wendland's C^2 function with bandwidth 21 on $M = 289$ points 24

8 Residuals using Wendland's C^2 function with bandwidth 21 on $M = 4225$ points 24

9 Recovery using Wendland's C^2 function with bandwidth 45 on $M = 289$ points 24

10 Residuals using Wendland's C^2 function with bandwidth 45 on $M = 4225$ points 24

11 Recovery using Wendland's C^2 function with bandwidth 21 on $M = 289$ points, multilevel method 27

12 Residuals using Wendland's C^2 function with bandwidth 21 on $M = 289$ points, multilevel method 27

13 Recovery using Wendland's C^2 function with bandwidth 5 on $M = 9$ points, multilevel method 27

14 Recovery using Wendland's C^2 function with bandwidth 5 on $M = 25$ points, multilevel method 27

15 Recovery using Wendland's C^2 function with bandwidth 5 on $M = 81$ points, multilevel method 27

16 Recovery using Wendland's C^2 function with bandwidth 5 on $M = 289$ points, multilevel method 27

17 Recovery using Wendland's C^2 function with bandwidth 5 on $M = 4225$ points, multilevel method 27

18 Comparison of Power Functions(*FigPFEx1*) 88
19 Optimal Knot Placement (*FigPFEx2*) 89

List of Tables

1 Conditionally Positive Definite Functions (*TCPDFct*) 19
2 Unconditionally Positive Definite Functions (*TPDFct*) 19
3 Errors for interpolation of Franke's function, Nonstationary Case 23
4 Errors for interpolation of Franke's function, Stationary Case . 25
5 Errors for interpolation of Franke's function, Stationary Case, Interpolation of residuals 26
6 L_∞ Bounds of Power Function Based on Lagrange Data (*TCPDEB*) 121
7 Lower Bounds of Smallest Eigenvalue Based on Lagrange Data with Separation Distance q (*TCPDC*) 139
8 Conditionally Positive Definite Functions (*TCPDFct2*) 207
9 Unconditionally Positive Definite Functions (*TPDFct2*) . . . 207
10 Transforms (*TFT*) 207

Foreword

This text currently serves two purposes:

- it backs up the lecture on reconstruction of multivariate functions as given in Göttingen in summer 1996, and
- it serves as a gradually growing reference manual for research of the group in Göttingen and related places.

It may finally develop into a monograph, but as to now it is rather preliminary and not intended for general distribution. Suggestions, corrections, addenda, and any form of criticism are welcome.

R. Schaback Göttingen, January 21, 1997

e-mail: schaback@namu01.gwdg.de

Layout for this test version:

- Logical LaTeX labels printed out in slanted font.
- Three-level cumulative numbering of environments and equations.
- The index is just a preliminary and (possibly) unsorted list of keywords.

1 Introduction

(*SectIntro*) The following is intended to give the basic motivation for what follows in later chapters. It shows that the reconstruction of multivariate functions f from certain function spaces $\mathcal{F}$ requires dependence of $\mathcal{F}$ on the data. Such data-dependent spaces are provided by conditionally positive definite functions, and these are in the focus of this text. Their optimality properties, as proven in later sections, justify this point of view. After definition of spaces generated by conditionally positive definite functions, this section introduces the standard algorithms for recovery of functions from such spaces. Examples, generalizations, proofs, theoretical details, and implementation problems will be added in later sections.

1.1 Recovery, Interpolation, and Approximation

In almost all practical applications, a function f is given not as a formula, but as a bunch of data. These data often take the form of approximate values $f(x_1), \ldots, f(x_M)$ of f at some scattered locations $x_1, \ldots, x_M$ of the domain Ω of definition of f. The **recovery problem** then consists in the reconstruction of f as a formula from the given data. This reconstruction can be done in two ways:

- **interpolation** tries to match the data exactly, taking f from a large class $\mathcal{F}$ of functions that is actually able to meet all of the data, while

- **approximation** allows f to miss the data somewhat, but selects the reconstruction function from a smaller class $\mathcal{F}$ of functions that will not in general be able to reconstruct the data exactly.

The selection between interpolation and approximation will depend on the application, and especially on the choice of function classes $\mathcal{F}$ and the necessity of exact reproduction of data.

We shall address both problems here, and there will be some hidden links discovered between the two approaches. Furthermore, we shall allow a much wider class of recovery problems in later sections, but the basic motivation is better shown by the above simplified "Lagrange" setting.

1.2 Input and Output Data

We shall consider reconstruction of d-variate functions f defined on a domain Ω. In most cases, Ω will be a subset of $I\!R^d$, but many results will hold on general sets. Right from the start we keep in mind that d might be large and that the domain Ω may be all of $I\!R^d$ or something special like a subdomain of $I\!R^d$ or the $d-1$ dimensional **sphere**, i.e. the surface $\{ x \in I\!R^d \ : \ \|x\|_2 = 1 \}$ of the **unit ball** $\{ x \in I\!R^d \ : \ \|x\|_2 \leq 1 \}$, where $\|.\|_2$ denotes the usual **Euclidean norm** on $I\!R^d$. In addition, we also may encounter very large sets of data, and these usually come up in two parts:

- a finite set $X = \{x_1, \ldots, x_M\}$ of M possibly wildly scattered points in some domain $\Omega \subseteq I\!R^d$, and

- real numbers $f_1, \ldots, f_M$ that represent approximate values of f at the given points.

The reconstruction should supply some function f defined on a domain Ω that contains all the data locations, and the data are to be reproduced approximately in the sense

$$f_j \approx f(x_j), \quad 1 \leq j \leq M.$$

But there are two other important input data for the recovery process:

- the domain Ω should be prescribed by the user, and
- the reconstruction should be confined to some prescribed class $\mathcal{F}$ of functions in order to avoid unpredictable results.

These will finally fix the set of formulas that are allowed as the output of the recovery process. Their choice will very much depend on the application and on additional knowledge of the user. For instance, somebody might want the resulting function f to be defined on all of $I\!R^d$, while somebody else is interested in a much more local reconstruction, e.g. in the convex hull of the data locations.

Furthermore, there may be different requirements on the smoothness of the recovered function or on its decay further away from the data. These have to be incorporated into the choice of $\mathcal{F}$, in addition to further information the user can provide.

1.3 Restrictions on the Choice of Spaces

There are two good reasons to assume that the class $\mathcal{F}$ of functions should be a linear space:

- If the values f_j are multiplied by a fixed scalar factor α, then the new data should be recovered by the function αf instead of f.
- If data f_j and g_j at the same locations $x_j \in I\!R^d$ are recovered by functions f and g, respectively, then the data $f_j + g_j$ should be recovered by the function $f + g$.

Note that this does not only require the class $\mathcal{F}$ to be a linear space: it also enforces the whole recovery process to consist of linear maps that associate a function to each data set. Furthermore, the recovery process will have a nonunique solution and thus be numerically unstable, if there is a function g in $\mathcal{F}$ that vanishes at all data locations in $X = \{x_1, \ldots, x_M\}$, because then all functions of the form αg can be added to a solution f without altering the data reproduction.

Definition 1.3.1 *(DefNond) If $\mathcal{F}$ is a space of functions on a domain Ω, then a subset X of Ω is called $\mathcal{F}$-***nondegenerate***, if zero is the only function from $\mathcal{F}$ that vanishes on X.*

We see that only the $\mathcal{F}$-nondegenerate subsets X of Ω can be used for stable reconstruction. It would be nice if any finite set X or at least (if $dim\mathcal{F} = M$) any set $X = \{x_1, \ldots, x_M\}$ would be nondegenerate for a given space $\mathcal{F}$.

But in truly multivariate situations this turns out to be *impossible*. In fact, if a linear subspace $\mathcal{F}$ of dimension $M \geq 2$ of a space of multivariate functions is fixed independent of the set $X = \{x_1, \ldots, x_M\}$, there always is a degenerate set X. This surprising and disappointing observation dates back to Mairhuber and Curtis (cf. [8](*Braess:86-1*)):

Theorem 1.3.2 *(MCTheorem) Let $\mathcal{F}$ be an M-dimensional space of continuous real-valued functions on some domain $\Omega \subseteq I\!R^d$ which is truly d-dimensional in the sense that it contains at least an open subset Ω_1 of $I\!R^d$. Assume further that any set $X = \{x_1, \ldots, x_M\} \subseteq \Omega_1$ is $\mathcal{F}$-nondegenerate. Then either $M = 1$ or $d = 1$ hold, i.e. either the function space or the underlying domain are just one-dimensional.*

Proof. We can assume $\Omega = \Omega_1$ without loss of generality. If the continuous functions $v_1, \ldots, v_M$ are a basis of $\mathcal{F}$, then the function $D(x_1, \ldots, x_M) = \det(v_j(x_k))$ is a continuous function of its M arguments. Due to our assumption this function can vanish only if two or more of the arguments coincide. Let us assume $M \geq 2$, and let Ω be at least truly 2-dimensional. Then one can swap the points x_1 and x_2 by a continuous motion that avoids coincidence with any of the arguments. Thus there is a continuous transition between $D(x_1, x_2, x_3, \ldots, x_M)$ and $D(x_2, x_1, x_3, \ldots, x_M) = -D(x_1, x_2, x_3, \ldots, x_M)$ that keeps D away from zero. This is impossible. $\square$

1.4 Data-dependent Spaces

(*SubSectDDSpaces*) The Mairhuber-Curtis theorem 1.3.2 (*MCTheorem*) forces us to let the space $\mathcal{F}$ depend on the data. But since for given $X = \{x_1, \ldots, x_M\}$ there should be a linear recovery map

$$R_X \; : \; I\!R^M \to \mathcal{F}. \quad (f_1, \ldots, f_M) \mapsto f,$$

it is reasonable to let $\mathcal{F}$ depend on the data **locations** $X = \{x_1, \ldots, x_M\}$ only, not on the data *values* $f_1, \ldots, f_M$. The formulas for the construction of functions $f(x)$ in $\mathcal{F}$ thus must depend on $X = \{x_1, \ldots, x_M\}$ and generate a linear space. The most straightforward way to achieve this is to combine the arguments x and x_j into a *single* function

$$\Phi : \Omega \times \Omega \to I\!R$$

and to view each $\Phi(x, x_j)$ as a data-dependent function of the variable x. Superposition of such functions results in a space

(*calfdef*)

$$\mathcal{F}_{X,\Phi} := \left\{ \sum_{j=1}^{M} \alpha_j \Phi(x, x_j) \quad : \quad \alpha_j \in I\!R \right\} \tag{1.4.1}$$

that may serve our purposes. It will turn out later that there are strong arguments to support this definition of a data-dependent space of functions. Under quite weak and general assumptions it can be proven that there is no better way to do it. Details of this will be given in 3.1.5 (*Necessity)*, but we cite the basic features here to support some useful simplifications. If for some Φ the union of all function spaces $\mathcal{F}_{X,\Phi}$ for varying sets X is required to have **translation invariance**, then the function Φ should be of the special form

$$\Phi(x, y) = \phi(x - y), \quad \phi \ : I\!R^d \to I\!R^d.$$

If we add **rotational invariance**, we end up with **radial basis functions**

$$\Phi(x, y) = \phi(\|x - y\|_2), \quad \phi \ : \ I\!R_{\geq 0} \to I\!R.$$

Note that in the latter case there is only a single *univariate* function required to generate a large class of spaces of *multivariate* functions. If we are working on the unit sphere in $I\!R^d$ and assume rotational invariance, we get **zonal** functions

$$\Phi(x, y) = \phi(x^T y), \quad \phi \ : I\!R_{\geq 0} \to I\!R^d,$$

where x^T stands for transposition of the vector x such that $\Phi(x, y)$ just is a univariate function $\phi(x^T y)$ of the scalar product $x^T y$. Details are provided in section 3.2.4 (*SecIP)*.

Of course there are other methods to generate data-dependent linear spaces of functions. The most prominent one is used widely in the theory of **finite elements**. There, the data set $X = \{x_1, \ldots, x_M\}$ is first used to generate a triangulation of its convex hull, and then one constructs functions on each subset of the triangulation, which are finally patched together to form smooth global functions. This approach is very effective if the space dimension d is small and if the functions to be recovered need not be very smooth. We refer the reader to the vast literature on this subject, and we proceed without considering triangulations of domains and patching of functions.

1.5 Evaluation, Interpolation and Approximation

(*subsecEIA*) The representation of functions in (1.4.1, *calfdef*) now serves as the reconstruction formula, and all one has to do when solving the reconstruction problem is to determine the vector $\alpha = (\alpha_1, \ldots, \alpha_M)$ of the coefficients of the resulting function with the representation

(*falphadef*)

$$f_\alpha(x) := \sum_{j=1}^{M} \alpha_j \Phi(x, x_j), \quad x \in \Omega \subseteq I\!R^d. \tag{1.5.1}$$

Before we turn to this problem, we note that evaluation of such a function at large numbers of different locations $x \in \Omega$ can be quite costly if M is large. However, the strong dependence on M can be relaxed if the values $\Phi(x, x_j)$ vanish whenever x and x_j are not near to each other. Examples of such functions will be given later.

Reconstruction by *interpolation* on $X = \{x_1, \ldots, x_M\}$ will now require to solve the linear system

(*EQsys1*)

$$\sum_{j=1}^{M} \alpha_j \Phi(x_k, x_j) = f_k, \quad k = 1, \ldots, M \tag{1.5.2}$$

for $\alpha_1, \ldots, \alpha_M$. We shall write this in shorthand matrix form as

$$A\alpha \ = \ f,$$

but in cases where the dependence on X and Φ is crucial, we add capital subscripts:

$$A_{X,\Phi}\alpha_{X,\Phi} \ = \ f_X, \quad A_{X,\Phi} = (\Phi(x_k, x_j))_{1 \leq j,k \leq M}\,.$$

To make the system uniquely solvable, the matrix A must be nonsingular. Looking at approximation, we shall soon have additional reasons to assume that $A_{X,\Phi}$ should even be positive definite. Thus it is more or less unavoidable to assume $A_{X,\Phi}$ to be positive definite for all X, when the function Φ is fixed. For these reasons we require the function Φ to satisfy

Definition 1.5.3 *(DPD) A real-valued function*

$$\Phi : \Omega \times \Omega \to I\!R$$

is a **positive definite function** *on* Ω*, iff for any choice of finite subsets* $X = \{x_1, \ldots, x_M\} \subseteq \Omega$ *of* M *different points the matrix*

$$A_{X,\Phi} = (\Phi(x_k, x_j))_{1 \le j,k \le M}$$

is positive definite.

At first sight it seems to be a miracle that a fixed function Φ should be sufficient to make all matrices of the above form positive definite, no matter which points are chosen and no matter how many. It is even more astonishing that one can often pick radial functions like $\Phi(x, y) = \exp(\|x - y\|_2^2)$ to do the job, and to work for **any** space dimension.

Turning to approximation, the space $\mathcal{F}_{X,\Phi}$ of (1.4.1, *calfdef*) should depend on less data than those given to determine the approximation. We simply assume some other data on some (large) Lebesgue-measurable subset $\Omega_1 \subseteq \Omega$ to be specified, and approximation should take place in the space $L_2(\Omega_1)$, for instance, which we assume to contain $\mathcal{F}_{X,\Phi}$. This covers discrete and continuous least-squares fits on the set Ω_1 by functions of the form f_α from (1.5.1, *falphadef*). The normal equations for the approximation are

$$\sum_{j=1}^{M} \alpha_j (\Phi(\cdot, x_k), \Phi(\cdot, x_j))_{L_2(\Omega_1)} = (\Phi(\cdot, x_k), f(\cdot))_{L_2(\Omega_1)}, \quad k = 1, \ldots, M.$$

Introducing new functions

(*Psidef*)

$$\Psi(x, y) := (\Phi(\cdot, x), \Phi(\cdot, y))_{L_2(\Omega_1)} \tag{1.5.4}$$

$$g(y) := (\Phi(\cdot, y), f(\cdot))_{L_2(\Omega_1)}$$

we see that this is exactly an *interpolation* system of the form

$$A_{X,\Psi} \alpha_{X,\Psi} = g_X.$$

Thus approximation reduces to interpolation by functions from a similar, but somewhat different function space.

At this point we see how positive definiteness comes in: the above matrix $A_{X,\Psi}$ is a Gramian with respect to the functions $\Phi(\cdot, x_k)$ in the inner-product space $L_2(\Omega_1)$. Thus it is positive definite whenever these functions are linearly independent in $L_2(\Omega_1)$. But the latter requirement is unavoidable for stable approximation in $L_2(\Omega_1)$.

From these preliminary considerations we conclude that positive definite functions should be investigated further, and we note in passing that (1.5.4, *Psidef*) yields a first method to construct such functions Ψ from linear independent functions $\Phi(\cdot, x_k)$, $1 \leq k \leq M, x_k \in \Omega$. We shall consider such constructions in detail in section 9.1 (*SecGCT*), but we remark in passing that the **Gaussian**

$$\Phi(x, y) := \exp(-\alpha \|x - y\|_2^2)$$

is positive definite on $\mathbb{R}^d$ for all d and all $\alpha > 0$. Since the proof requires tools like Fourier transforms, we defer it to Theorem 12.5.6 (*GaussPD*) on page 199.

1.6 Conditionally Positive Definite Functions

Positive definite functions (formerly defined in a slightly different way) have a long history that is nicely surveyed by Stewart [44](*Stewart:76-1*). However, the first cases of radial basis functions used widely and successfully in applications were

- the **thin-plate spline** $\Phi(x, y) = \phi(\|x - y\|_2) = -\|x - y\|_2 \log \|x - y\|_2$ introduced by Duchon [10](*duchon:76-1*), [11](*duchon:78-1*), [12](*duchon:79-1*),
- the **multiquadric** $\Phi(x, y) = \phi(\|x - y\|_2) = \sqrt{c^2 + \|x - y\|_2^2}$ and
- the **inverse multiquadric** $\Phi(x, y) = \phi(\|x - y\|_2) = \dfrac{1}{\sqrt{c^2 + \|x - y\|_2^2}}$

  used by the geophysicist Hardy [17](*hardy:71-1*)

but the first two of these are *not* positive definite. The corresponding matrices $A_{X,\Phi}$ naturally define quadratic forms

(*QFdef*)

$$Q_{X,\Phi} \;:\; (\alpha_1, \ldots, \alpha_M) \mapsto \alpha^T A_{X,\Phi} \alpha := \sum_{j,k=1}^{M} \alpha_j \alpha_k \Phi(x_j, x_k) \qquad (1.6.1)$$

on $\mathbb{R}^d$, where ^{T} stands for vector transposition, but these forms are positive definite only on a proper subspace of $\mathbb{R}^M$. More precisely, for certain positive values of m the above functions Φ satisfy the following

Definition 1.6.2 *(DCPD) A real-valued function*

$$\Phi : \Omega \times \Omega \to \mathbb{R}$$

is a **conditionally positive definite function** *of order m on $\Omega \subseteq I\!R^d$, iff for any choice of finite subsets $X = \{x_1, \ldots, x_M\} \subseteq \Omega$ of M different points the value*

$$\alpha^T A_{X,\Phi}\alpha := \sum_{j,k=1}^{M} \alpha_j \alpha_k \Phi(x_j, x_k)$$

of the quadratic form (1.6.1, QFdef) is positive, provided that the vector $\alpha = (\alpha_1, \ldots, \alpha_M) \in I\!R^M \setminus \{0\}$ has the additional property

(CPDef)

$$\sum_{j=1}^{M} \alpha_j p(x_j) = 0 \tag{1.6.3}$$

for all d-variate polynomials p of order (=degree-1) up to m. The linear space of such polynomials will be denoted by $I\!P^d_m$, and its dimension is

$$q := \binom{m-1+d}{d}.$$

It is a major problem to prove that multiquadrics are conditionally positive definite of a fixed order m for all space dimensions d. This was done (among other things) in Micchelli's fundamental paper [27](*micchelli:86-1*) that boosted the research on radial basis functions.

1.7 Basic Equations for Conditionally Positive Definite Functions

If Φ is conditionally positive definite of order m on $\Omega \subseteq I\!R^d$, then the additional condition (1.6.3, *CPDef*) reduces the M degrees of freedom of $\alpha \in I\!R^M$ by at most q, the dimension of the space $I\!P^d_m$ of polynomials. Thus it is reasonable to add q new degrees of freedom to the recovery process by adding $I\!P^d_m$ to the space of admissible functions. Then (1.4.1, *calfdef*) has to be replaced by

(*calfdef2*)

$$\mathcal{G}_{X,\Phi} := I\!P^d_m + \mathcal{F}_{X,\Phi} = I\!P^d_m + \left\{ \sum_{j=1}^{M} \alpha_j \Phi(x, x_j) \;:\; \alpha_j \in I\!R \text{ with (1.6.3, } CPDef) \right\}. \tag{1.7.1}$$

Now the $M \times M$ system (1.5.2, *EQsys1*) goes over into the $(M+q) \times (M+q)$ system

(*EQsys2*)

$$
\begin{array}{lcccl}
\sum_{j=1}^{M}\alpha_j\Phi(x_k,x_j) & + & \sum_{i=1}^{q}\beta_i p_i(x_k) & = & f_k, \quad 1\le k\le M \\
\sum_{j=1}^{M}\alpha_j p_i(x_j) & + & 0 & = & 0, \qquad 1\le i\le q
\end{array}
\tag{1.7.2}
$$

for vectors $\alpha = (\alpha_1,\ldots,\alpha_M) \in I\!R^M$ and $\beta = (\beta_1,\ldots,\beta_q) \in I\!R^q$, where the polynomials $p_1,\ldots,p_q$ are a basis of $I\!P^d_m$. Introducing a matrix

$$
P := P_X := (p_i(x_j))_{1\le i\le q, 1\le j\le M},
$$

of values of polynomials, this system reads in matrix form as

(*BDef*)

$$
\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}. \tag{1.7.3}
$$

The coefficient matrix of this enlarged linear system will be abbreviated by B or $B_{X,\Phi}$. The solvability of (1.7.2, *EQsys2*) is described by

Theorem 1.7.4 *(Nonsing1) Let Φ be conditionally positive definite of order m on $\Omega \subseteq I\!R^d$, and let the data set $X = \{x_1,\ldots,x_M\} \subseteq \Omega$ be $I\!P^d_m$-nondegenerate. Then the system (1.7.2, EQsys2) is uniquely solvable. Furthermore, there are linear algebra techniques using at most $\mathcal{O}(Mq^2 + M^2q)$ operations to reduce it to a positive definite $(M-q)\times(M-q)$ system.*

Proof. Let a pair of vectors $\alpha \in I\!R^M$ and $\beta \in I\!R^q$ solve the homogeneous system with matrix (1.7.3, *BDef*). Then we have $A\alpha + P\beta = 0$ and $P^T\alpha = 0$. Multiplying the first equation with α^T and inserting the second in transposed form, we get $\alpha^T A\alpha + 0 = 0$. Now $\alpha = 0$ follows from conditional positive definiteness, and we are left with $P\beta = 0$. This in turn implies $\beta = 0$, because X is $I\!P^d_m$-nondegenerate. The second assertion will be proven by two explicit algorithms in 10.1 (*Red1*) and 10.2 (*Red2*). □

2 Working with Basis Functions

This section is intended for readers working on applications. It contains tables of the currently known conditionally positive definite functions and provides guidelines for picking the right function Φ from the tables. These guidelines are based on both numerical experience and theoretical insight.

However, this chapter will not attempt to prove any of the statements inherent in the guidelines, but rather refer the reader to subsequent sections.

Right after giving the general guidelines, we turn to efficiency considerations. Special strategies for system reduction, iterative solution, sparse matrices, and preprocessing techniques for large sets of data points are delayed to section 10 (*SecSA)*.

A series of examples serves for illustration. Since these examples are quite convincing in general, they justify the considerable amount of theoretical background to be developed in later sections.

2.1 General Practical Considerations

Before picking a suitable function Φ for recovering a function f in an application, the user first has to consider the following issues:

- How smooth should f be?
- What is the required behaviour near the boundary of the convex hull or outside of the data set $X = \{x_1, \ldots, x_M\}$?
- Are the data locations evenly or very unevenly distributed?
- Is exact reproduction of the data required?
- Are M and/or the space dimension d so large that efficiency considerations are predominant over reproduction quality questions?

2.1.1 Uncertainty Relation

(*GPCUP*) When considering the above questions, the user has to keep in mind that every good thing has its price. This basic fact of real life occurs here in the form of an *Uncertainty Relation*:

> If you go for good reproduction quality, you have to sacrifice numerical stability. If you go for good stability, you have to sacrifice reproduction quality.

This wishy-washy statement will be made precise in 4.6 (*URT)*, and there it turns out that both reproduction quality and numerical instability are linked to both data density and smoothness of Φ (and, in cases with compact support, to the size of the support radius of Φ). Furthermore, if large linear systems with positive definite coefficient matrices are solved by the conjugate gradient method, numerical stability is directly linked to efficiency via the condition of the matrices. This is why for large problems one can replace "stability" by "efficiency" in the Uncertainty Relation.

If the data density is considered fixed, the Uncertainty Relation suggests that the user should be very careful about the smoothness of the function Φ. It should be as low as the application tolerates, and any excessive smoothness will have negative effects on stability.

But if reproduction quality or stability is fixed, there is a trade-off between data density and smoothness of Φ. For sparse data one can work with smooth functions, and for large and dense data sets one has to work with low smoothness of Φ in order to avoid numerical problems. If working with compactly supported functions Φ, this is a standard way to escape the inherent numerical problems with very large and dense data sets. One can split the data set into subsets of increasing density and use compactly supported functions with decreasing support radii on these data sets. If things work out nice, one can expect to work at a fixed stability level, but with incereasing local resolution. We treat such *multilevel techniques* in detail in 2.4 (*MLA)* but the next paragraph will add some other arguments in favor of it.

Compactly supported functions offer computational advantages due to sparsity of the corresponding matrices. If supports are small, the effect of such functions will be strictly local, and this has both advantages and disadvantages. The disadvantage is that global effects cannot be nicely recovered, and thus small supports should be used only in cases where the global behavior is already recovered by any other method. The usual trick is to

- first apply a global method (possibly using a small but global data set),
- take the residuals (data minus values of the recovery function) and then
- handle the local effects by reconstruction the residuals using compactly supported functions on the full data set.

This three-stage process is quite common in applications and amounts to solve for the global *trend* first and then to model the local effects on a finer scale. The last two steps can be iterated using smaller and smaller supports, and this is the multilevel method that we look at in 2.4 (*MLA)*

2.1.2 Unevenly Distributed Data

(*GPCUDD*) The above statements assume a more or less evenly scattered data set. If there are local clusters of data points or areas without data, some other aspects come into the game. In fact, for a fixed function Φ the numerical stability and the reproduction quality are connected to two similar, but different quantities which roughly coincide for evenly distributed data sets. The stability is connected to the **separation distance**

(*SDDef*)

$$s := s_X := \frac{1}{2} \min_{1 \leq j \neq k \leq M} \|x_j - x_k\|_2 \tag{2.1.1}$$

while the reproduction quality on the domain Ω is ruled by a somewhat more complicated quantity (see (5.5, *hrhodef*)) that can roughly be described for practical purposes by the **fill distance**

(*DDDef*)

$$h := h_{X,\Omega} := \sup_{x \in \Omega} \min_{1 \leq j \leq M} \|x - x_j\|_2. \tag{2.1.2}$$

Separation distance measures the minimal distance that separates any two data locations, i.e. it is the minimal distance from any point of the data set to its nearest data point, while fill distance measures the way how the data fill the domain, i.e. it is the maximal distance from any point of the domain to its nearest data point. Thus fill distance is never smaller than separation distance, but hazardous cases have a very small separation distance relative to the fill distance. We call a data set *unevenly distributed* if this happens, and the quotient

$$\delta_{X,\Omega} := \frac{h_{X,\Omega}}{s_X} \geq 1$$

is a good measure for the unevenness of a data distribution X with respect to a domain Ω.

Now the naive treatment of unevenly distributed data sets will induce "additional" numerical instabilities caused by the irregularity of the data distribution. If these instabilities are severe, some action must be taken. If caused by a few points that are extremely near to other data locations with comparable data values, the user can simply throw these "duplicates" out of the data set and proceed, expecting that the nearby data points are sufficient for the required reconstruction.

But there are cases where the data show local clusters which themselves consist of nicely distributed data locations. Then the problem lives on more than one density scale, and the obvious technique to treat such cases is by working in several steps with increasing local resolution. This is another good reason for the multilevel approach in 2.4 (*MLA)*.

2.2 Current Basis Functions

(*SecCBF*) Table 1 (*TCPDFct)* lists some of the currently known radial functions that are conditionally positive definite of *positive* order m on $I\!R^d$. A more or less complete list will be in the Appendix under 13.1 (*SecBF)*. Note that these have some polynomial growth towards infinity, and they always generate non-sparse matrices. They work for any space dimension d, and they are especially useful for cases where decay towards infinity is a disadvantage. Thus they should not be applied to residuals but rather to the original data, and their power lies in good reproduction of the global overall shape of the function to be reconstructed, especially in areas away from the data locations.

We now turn to *unconditionally* positive definite functions defined on $I\!R^d$.

$\phi(r)$	Parameters	m
r^β	$\beta > 0,\ \beta \notin 2I\!N$	$m \geq \lceil \beta/2 \rceil$
$r^\beta \log r$	$\beta > 0,\ \beta \in 2I\!N$	$m > \beta/2$
$(r^2+c^2)^{\beta/2}$	$\beta > 0,\ \beta \notin 2I\!N$	$m \geq \lceil \beta/2 \rceil$

Table 1: Conditionally Positive Definite Functions (*TCPDFct*)

$\phi(r)$	Parameters	Smoothness	Dimension	Name/Reference
$e^{-\beta r^2}$	$\beta > 0$	$C^\infty(I\!R^d)$	$d < \infty$	Gaussian
$(r^2+c^2)^{\beta/2}$	$\beta < 0$	$C^\infty(I\!R^d)$	$d < \infty$	inv. Multiquadric
$r^\nu K_\nu(r)$	$\nu > 0$	$C^{\lfloor \nu \rfloor}$	$d < \infty$	Sobolev spline
$(1-r)_+^2(2+r)$		C^0	$d \leq 3$	Wu [47](*wu:95-2*)
$(1-r)_+^4(1+4r)$		C^2	$d \leq 3$	Wendland [46](*wendland:95-1*)

Table 2: Unconditionally Positive Definite Functions (*TPDFct*)

These have decay towards infinity and come in two variations: compactly supported or not. Due to results given in 9.2.12 (*NECSAlld)* there are no compactly supported positive definite functions that work for all space dimensions. Thus one has to check the space dimension d when working with compactly supported functions. Table 2 (*TPDFct)* lists some of the currently known cases and provides information about smoothness and admissible space dimensions for positive definiteness. See 13.1 (*SecBF)* for further cases and details.

The decay towards infinity may be an unwanted feature when applied to raw data, but it is very convenient when applied to residuals. Compact supports provide sparse matrices, but the adjustment of the support radius can be hazardous. If chosen too small, the resulting matrices $A_{X,\Phi}$ tend to be nicely diagonal, making the numerical solution very stable and efficient, but the reproduction quality is disastrous, because one reproduces the data by extremely narrow and isolated "delta" peaks. On the contrary, a large support radius very much improves reproduction quality, but at the expense of matrix fill-in and increasing condition. This is another consequence of the Uncertainty Relation.

2.3 Computational Complexity of Solving the System

(*CompEffort*) We now investigate the numerical effort required to solve the system (1.7.3, *BDef*). Assuming that q usually is zero or small compared to M, we roughly have a positive definite and symmetric $M \times M$ system to solve. If the condition is reasonable and M is not too large, Cholesky factorization will do the job at about $M^3/6 + \mathcal{O}(M^2)$ computational cost. However, this is not acceptable for large M. In particular, the value of M can be even too large to form the full matrix in storage. Therefore one has to look for iterative methods and sparse matrix techniques. Some special tricks due to Beatson [5](*beatson-newsam:92-1*) and Powell [38](*powell:92-1*)[37](*powell:92-2*)[4](*PowellEffTPSSystem*) are possible for specific basis functions, but we concentrate here on the solution via compactly supported functions.

In this case the matrix is sparse and its bandwidth depends on the relative size δ/h of the support radius δ and the fill distance h. For a fixed compactly supported positive definite function Φ the effect of an increase of δ yields

- an increase of the bandwidth of the matrix in (1.7.3, *BDef*) via an increase of δ/h,
- an increase of the reproduction quality via an increase of δ/h (see 6.6 (*SecError)*), and
- an increase of its condition via an increase of δ/q (see 4.5 (*SecCondition)*).

This is another version of the Uncertainty Relation, and the user has to fix the support radius δ to be sufficiently large to get good reproduction quality while keeping it small enough to let the solution of the system be computationally effective. A general rule of thumb is to work at the limits of the computational resources, and to switch to multilevel techniques (see 2.4 (*MLA*)) in cases where the reproduction quality still is inadequate.

If the ratio δ/q is kept bounded, the norm of the inverse (and thus a major part of the condition) of the matrix in (1.7.3, *BDef*) is bounded. Solving the system by conjugate gradients uses only a fixed number of iterations for fixed precision requirements, if the condition is bounded. Furthermore, each iteration takes only $\mathcal{O}(M \cdot B)$ operations for bandwidth B. Thus the numerical cost of solving the system (1.7.3, *BDef*) can be kept roughly at $\mathcal{O}(M)$, if the user keeps the ratios of h, q, and δ within reasonable bounds.

We finally check the complexity of evaluating (1.5.1, *falphadef*) at a single argument x. In general one has to expect $\mathcal{O}(M)$, but since one has to evaluate the function in at least $\mathcal{O}(M)$ or many more points, the cost for evaluation will even be underestimated by $\mathcal{O}(M^2)$. For large values of M this cannot be tolerated. Using stencils [38](*powell:92-1*) and Laurent expansions [37](*powell:92-2*) Powell has overcome these difficulities in case of thin-plate splines. For compactly supported basis functions with maximally B points in their support (this coincides with the bandwidth of the system (1.7.3, *BDef*)) one has $\mathcal{O}(B)$ operations for each evaluation, which is a significant advantage if many evaluations have to be made. However, each evaluation then requires to solve the B-nearest-neighbor problem of computational geometry, because for each x one has to pick the B data points x_j with nonzero $\Phi(x, x_j)$ in an effective way. If the data are not too wildly scattered, one can employ preprocessing techniques of complexity at most $\mathcal{O}(M)$ to solve this problem at $\mathcal{O}(1)$ for each x. In general, preprocessing of cost $\mathcal{O}(M \log M)$ is necessary to provide a $\mathcal{O}(\log M)$ complexity of solving the B-nearest-neighbor problem for each x. Details will be provided in section 11 (*SecCGT*)

2.4 Multilevel Algorithms

(*MLA*) The basic idea here is to work at levels indexed by j, where one uses a basis function Φ_j that usually will be compactly supported with a support radius δ_j. On level j the data is confined to a subset X_j of the full data set X, and the corresponding fill distance and separation distance will be denoted by h_j and q_j, respectively. The function f_j to be recovered by some other function s_j at level j consists of the *residuals* of the preceding step, i.e.

$$f_j := f_{j-1} - s_{j-1},\ j \geq 1,\ f_0 := f.$$

The *ratios* of the three quantities h_j, q_j, and δ_j are kept at reasonable values that make the computations possible, while the quantities themselves decrease with increasing j.

The rationale behind this multilevel techniques is to recover the function f at different levels of resolution, starting from global reconstruction of slowly varying features from coarse global data and ending up with highly local reconstruction of fine details from densely distributed data. The numerical performance of this technique is superior to single-level techniques in applications with very large data sets (see 2.5 (*SecExamples*) and [13](*floater-iske:95-1*) [14](*floater-iske:96-1*)), but its theoretical treatment, starting in [5](*NRSW*), still is incomplete. The numerical cost can be kept to $\mathcal{O}(M)$ by proper choice of supports and fill distances.

2.5 Numerical Examples

(*SecExamples*) Here are some first examples of reconstructions of functions from multivariate scattered data. For easy presentation, we restrict ourselves to two-dimensional cases and use MATLAB for the computations. The corresponding MATLAB M-files and MEX-files are in the appendix.

We start with the reconstruction of Franke's function [15](*franke:82-1*) rescaled to $\Omega = [0,1] \times [0,1] \subset I\!R^2$ from data on a grid $(i/2n, j/2n)$, $0 \leq i,j \leq 2n$ such that $M = (2n+1)^2$. The matrix in (1.7.3, *BDef*) then has approximately $4n^4$ entries, and the computational cost of Cholesky factorization is about $4n^6/3$. If the matrix is non-sparse, only very moderate values of n can be treated.

The function itself (Figure 1 (*FigFranke33*)) is nicely reconstructed up to graphical precision by thin-plate splines $\phi(r) = r^2 \log r$ from information on $M = 81$ data points (Figure 2 (*FigTPS81Fct33*)). The effects of higher values of M are visualized by plotting *residuals* (see Figure 3 (*FigTPS81Res*) for $M = 81$, and note the scale on the z-axis for plots of residuals).

Working on more than $M = 225$ points becomes very ineffective for non-sparse cases. Thus we now consider examples with Wendland's compactly supported radial basis function $\phi(r) = (1-r)_+^4(4r+1)$ with support scaled to radius δ. On $M = 81$ data points one can still compare with the previous case while using a large support radius $\delta = 2$ (Figures 4 (*FigWF81Supp2Fct33)*, 5 (*FigWF81Res2)*). To handle larger values of M, the support radius has to be decreased to introduce sparsity. We start with examples having bandwidth 21 on $M = 289$ and $M = 4225$ points (Figures 6 (*FigWF289Ban21Fct65)*, 7 (*FigWF289Ban21)*, 8 (*FigWF4225Ban21)*). Note that the reproduced function is overlaid by some high-frequency wiggles that arise from the small support of the radial basis functio used: the approximation is somewhat too spiky. A look at the residuals supports this, but also implies that the larger errors occur at the boundary. These take over when going to 4225 data points, and make the errors in the interior hardly visible. This is the first hint that the behavior near the boundary needs special treatment.

Now Figures 9 (*FigWF289Ban45Fct65)*, 10 (*FigWF4225Ban45)*) show residuals computed with matrices of bandwidth 45. The results are better, of course, but the message is the same.

For even larger values of M we refrain from providing plots of residuals. Instead, we evaluate the error on a fine grid. Table 3 (*TabNonstat)* on page 23 shows the maximum errors for cases with fixed support radius δ

$N \setminus \delta$	0.03125	0.0625	0.125	0.25	0.5	1	2	4	8
9	*	*	*	*	12.1754	5.4808	5.5436	5.8102	5.9030
25	*	*	*	10.2176	1.1995	0.8186	0.6902	0.6889	0.7073
81	*	*	11.5563	1.1013	0.4668	0.3621	0.3570	0.3584	0.3587
289	*	11.7369	0.8148	0.4606	0.1175	0.0397	0.0241	0.0224	0.0226
1089	11.6653	0.7812	0.4783	0.1158	-	-	-	-	-
4225	0.7791	0.4561	-	-	-	-	-	-	-
16641	-	-	-	-	-	-	-	-	-
66049	-	-	-	-	-	-	-	-	-

Table 3: Errors for interpolation of Franke's function, Nonstationary Case (*TabNonstat*)

* Errors too large due to extremely small supports used,

\- Workspace exhausted or non-sparse matrix.

(*nonstationary case*), as far as the computations were numerically feasible. Convergence along columns is clearly visible, but the scope is still severely limited by computational restrictions.

If the support radius is kept strictly proportional to the fill distance (this is called the fully *stationary* case), then the bandwidth B is constant along

Figure 1: Franke's function
(*FigFranke33*)

Figure 2: Reconstruction of Franke's function from thin-plate splines on $M = 81$ points
(*FigTPS81Fct33*)

Figure 3: Residuals for thin-plate splines on $M = 81$ points
(*FigTPS81Res*)

Figure 4: Recovery using Wendland's C^2 function with support radius 2 on $M = 81$ points
(*FigWF81Supp2Fct33*)

Figure 5: Residuals using Wendland's C^2 function with support radius 2 on $M = 81$ points
(*FigWF81Res2*)

Figure 6: Recovery using Wendland's C^2 function with bandwidth 21 on $M = 289$ points
(*FigWF289Ban21Fct65*)

Figure 7: Residuals using Wendland's C^2 function with bandwidth 21 on $M = 289$ points
(*FigWF289Ban21*)

Figure 8: Residuals using Wendland's C^2 function with bandwidth 21 on $M = 4225$ points
(*FigWF4225Ban21*)

Figure 9: Recovery using Wendland's C^2 function with bandwidth 45 on $M = 289$ points
(*FigWF289Ban45Fct65*)

Figure 10: Residuals using Wendland's C^2 function with bandwidth 45 on $M = 4225$ points
(*FigWF4225Ban45*)

columns in Table 4 (*TabStat)* on page 25, but there is no convergence along

$N \setminus B$	1	5	9	13	21	25	29	37	45
9	12.1754	8.1801	5.4801	5.3389	5.3521	5.3770	5.4083	5.4830	5.5436
25	10.2176	4.6070	1.1993	0.9549	0.9209	0.8995	0.8719	0.8400	0.8186
81	11.5563	4.8475	1.1003	0.8840	0.7236	0.6820	0.6316	0.5368	0.4668
289	11.7369	4.5695	0.8148	0.7554	0.7670	0.7190	0.6606	0.5457	0.4606
1089	11.6653	4.4424	0.7812	0.7831	0.7924	0.7432	0.6838	0.5661	0.4783
4225	11.7024	4.4322	0.7791	0.7733	0.7566	0.7099	0.6529	0.5416	0.4561
16641	11.7109	4.4292	0.7786	0.7119	0.7577	0.6994	0.6578	0.5461	-
66049	12.9205	4.4283	-	-	-	-	-	-	-

Table 4: Errors for interpolation of Franke's function, Stationary Case (*TabStat*)

N number of data points

B number of points per support

\- Workspace exhausted

columns, while the scope is greatly enlarged. Convergence occurs along lines with negative slope in this table, but the minimum attainable error still is quite large. The condition is roughly constant in each column, such that the overall numerical cost is approximately proportional to M.

We now recalculate the columns of Table 4 (*TabStat)* by taking successive residuals as we proceed along each column, working at fixed bandwidth and fixed condition, thus with $\mathcal{O}(M)$ overall computational complexity (see Table 5 (*TabMulti)* on page 26). This multilevel approach now decreases the error significantly and seems to have at least a linear convergence along columns. More information on the numerical behavior of the multilevel approach can be found in [13](*floater-iske:95-1*) [14](*floater-iske:96-1*). Here, we support the results of Table 5 (*TabMulti)* by some additional plots of multilevel interpolants to Franke's function. Figure 11 (*FigWF289Ban21MLFig)* shows the multilevel reconstruction with bandwidth 21 after four levels with 9, 25, 81, and 289 data points. The residuals are in Figure 12 (*FigWF289Ban21ML)* and should be compared with Figure 7 (*FigWF289Ban21)* with the same bandwidth on 289 points, using a single step.

To visualize the smoothing effect of the multilevel method, we pick a drastic example by choosing a very small bandwidth of 5. The reader will realize

$N \setminus B$	1	5	9	13	21	25	29	37	45
9	12.0412	8.1801	5.4801	5.3389	5.3521	5.3770	5.4048	5.4830	5.5436
25	7.6972	2.5971	0.9328	0.7840	0.7016	0.6943	0.6842	0.6808	0.6i845
81	5.9089	0.9172	0.4223	0.3820	0.3565	0.3571	0.3595	0.3680	0.3735
289	4.4449	0.2927	0.0680	0.0518	0.0352	0.0332	0.0314	0.0303	0.0288
1089	3.3053	0.0867	0.0256	0.0187	0.0120	0.0112	0.0105	0.0098	0.0092
4225	2.4589	0.0320	0.0090	0.0064	0.0039	0.0036	0.0034	0.0031	0.0029
16641	1.7481	0.0118	0.0034	0.0023	0.0013	0.0011	0.0011	0.0009	0.0008
66049	1.3085	0.0053	-	-	-	-	-	-	-

Table 5: Errors for interpolation of Franke's function, Stationary Case, Interpolation of residuals
(*TabMulti*)

N number of data points

B number of points per support

\- Workspace exhausted

that this method will be feasible even for gigantic data sets. Figures 13 (*FigWF9Ban5MLFig)* 14 (*FigWF25Ban5MLFig)* 15 (*FigWF81Ban5MLFig)* 16 (*FigWF289Ban5MLFig)* 17 (*FigWF4225Ban5MLFig)* show reconstruction from $M = 9,\ 25,\ 81,\ 289,$ and 4225 points. The extremely small bandwidth of 5 does not have a serious influence on the quality on a 3×3 data set, but the spiky reproduction in the medium range introduces wiggles that are ironed out by increasing data density.

Of course, one should take larger supports in the intermediate range and use a bandwidth larger that 5 to produce optimal results, but the above sequence is picked to illustrate what happens qualitatively if the computational restrictions force to work with very small bandwidth. The actual errors can be read off the second column of Table 5 (*TabMulti)*.

To prove statements about the convergence rate and the condition of such calculations will be main goal of this text.

3 Hilbert Space Theory

(*SecHST*) Here we start with the basic theoretical foundations and proceed top-down. First, we pose the problem of recovery of elements of Hilbert

Figure 11: Recovery using Wendland's C^2 function with bandwidth 21 on $M = 289$ points, multilevel method
(*FigWF289Ban21MLFig*)

Figure 12: Residuals using Wendland's C^2 function with bandwidth 21 on $M = 289$ points, multilevel method
(*FigWF289Ban21ML*)

Figure 13: Recovery using Wendland's C^2 function with bandwidth 5 on $M = 9$ points, multilevel method
(*FigWF9Ban5MLFig*)

Figure 14: Recovery using Wendland's C^2 function with bandwidth 5 on $M = 25$ points, multilevel method
(*FigWF25Ban5MLFig*)

Figure 15: Recovery using Wendland's C^2 function with bandwidth 5 on $M = 81$ points, multilevel method
(*FigWF81Ban5MLFig*)

Figure 16: Recovery using Wendland's C^2 function with bandwidth 5 on $M = 289$ points, multilevel method
(*FigWF289Ban5MLFig*)

Figure 17: Recovery using Wendland's C^2 function with bandwidth 5 on $M = 4225$ points, multilevel method
(*FigWF4225Ban5MLFig*)

spaces in a very general sense. It turns out that optimal recovery is necessarily linked to the use of conditionally positive definite functions. Conversely, each conditionally positive definite function allows to define a "native" Hilbert space in which it serves to solve an optimal recovery problem. We study the error and the condition of the recovery process and prove the Uncertainty Relation in general. Altogether, this section is intended to contain all theoretical results that can be proven without resort to (Fourier) transforms and which hold for general domains. This implies that the more sophisticated results for special cases are found in later sections.

3.1 Optimal Recovery in Hilbert Spaces

3.1.1 Optimal Recovery Problems

(*subsecORP*) Assume that we want to reconstruct a function f defined on some domain Ω from M pieces of information concerning f. These may for instance be function values $f(x_j),\ 1 \leq j \leq M$ in case of classical Lagrange interpolation, or inner products $(f, p_j)_{L_2},\ 1 \leq j \leq M$ for L_2 approximation. In both cases the information consists of the value of a linear functional λ_j applied to f, and in the second case the function f is assumed to lie in a space with an inner product $(\cdot, \cdot)$ that serves to give a specific representation $\lambda_j(f) = (f, p_j)$ to the functionals in question.

To incorporate the second case, we thus assume that there is a space $\mathcal{F}$ of functions and a space $\mathcal{L}$ of functionals such that $\lambda(f)$ is the application of the functional $\lambda \in \mathcal{L}$ to the function $f \in \mathcal{F}$. The space $\mathcal{F}$ is supposed to carry an inner product $(\cdot, \cdot)_{\mathcal{F}}$, and the functionals $\lambda \in \mathcal{L}$ are supposed to be continuous with respect to this inner product, i.e.,

$$|\lambda(f)| \leq \|\lambda\|_{\mathcal{L}} \|f\|_{\mathcal{F}}$$

for all $\lambda \in \mathcal{L},\ f \in \mathcal{F}$, where the norm of functionals is defined as usual:

$$\|\lambda\|_{\mathcal{L}} := \sup_{\|f\|_{\mathcal{F}} \neq 0} \frac{|\lambda(f)|}{\|f\|_{\mathcal{F}}} < \infty.$$

We now assume that we want to recover an element f from the space $\mathcal{F}$ using the M real values

(*fj*)

$$\gamma_j = \lambda_j(f),\ 1 \leq j \leq M \tag{3.1.1}$$

of M linear functionals $\lambda_1, \ldots, \lambda_M$ that are continuous on $\mathcal{F}$. Furthermore, we assume the linear functionals $\lambda_1, \ldots, \lambda_M$ to be linearly independent in $\mathcal{L}$, which means that the information is not redundant.

Then there will usually be many elements $f \in \mathcal{F}$ that satisfy the equations (3.1.1, *fj*), which may now be viewed as generalized interpolation conditions. If f solves (3.1.1, *fj*) and if there is some element $v \in \mathcal{F}$ that satisfies the homogeneous conditions

$$0 = \lambda_j(v),\ 1 \leq j \leq M,$$

than all elements $f_\alpha := f + \alpha v$ for arbitrary $\alpha \in I\!R$ will solve (3.1.1, *fj*), too. These elements can have arbitrarily large norms, if v is not identically zero. To exclude solutions with extremely large norms one thus asks for elements $f^* \in \mathcal{F}$ that solve (3.1.1, *fj*) and minimize the norm $\|\cdot\|_{\mathcal{F}}$ under all other solutions. That is, the element f^* solves the **optimal recovery problem**

($ORPF$)

$$\|f^*\|_{\mathcal{F}} = \min_{\substack{f \in \mathcal{F} \\ f_j = \lambda_j(f)}} \|f\|_{\mathcal{F}} \tag{3.1.2}$$

in the space $\mathcal{F}$.

If we pursue this general setting further, we shall finally see that under mild additional assumptions there is a positive definite function that serves to solve the optimal recovery problem. But then we have lost the conditionally positive definite functions of positive order. Thus we try a fresh start that slightly generalizes the above recovery problem.

Instead of a space $\mathcal{F}$ with an inner product, we only assume there is a linear space $\mathcal{G}$ over $I\!R$ with a positive semidefinite bilinear form

$$(\cdot,\cdot)_{\mathcal{G}}\ : \mathcal{G} \times \mathcal{G} \to I\!R.$$

Then $|g|^2_{\mathcal{G}} = (g,g)_{\mathcal{G}}$ defines a seminorm $|\cdot|_{\mathcal{G}}$ on $\mathcal{G}$, and we assume that the nullspace

$$\mathcal{P} := \{\ g \in \mathcal{G}\ :\ |g|_{\mathcal{G}} = 0\ \}$$

has a finite dimension $q \geq 0$ and is spanned by a basis $p_1, \ldots, p_q$. As in (3.1.1, *fj*) we assume that we want to recover an element g from the space $\mathcal{G}$ using the M real values

(*gj*)

$$\gamma_j = \lambda_j(g),\ 1 \leq j \leq M \tag{3.1.3}$$

of M linear functionals $\lambda_1, \dots, \lambda_M$. But we would run into problems if we would simply assume continuity of these functionals with respect to the seminorm, because this would restrict us to functionals that vanish on $\mathcal{P}$. Postponing the precise assumptions on the functionals, we can now pose the **generalized optimal recovery problem**

(*ORP*)

$$|g^*|_{\mathcal{G}} = \min_{\substack{g \in \mathcal{G} \\ g_j = \lambda_j(g)}} |g|_{\mathcal{G}} \tag{3.1.4}$$

in the space $\mathcal{G}$.

3.1.2 Projection onto the Nullspace

(*SecHSP*) To discuss the solvability of the optimal recovery problem 3.1.4 (*ORP*) in a very general way, we need some more information on the space $\mathcal{G}$ and ist finite-dimensional subspace $\mathcal{P}$. It simplifies later arguments to have a simple way of projecting an element $g \in \mathcal{G}$ onto an element of $\mathcal{P}$. In standard applications, this projection will be an interpolation or an approximation by a low-order polynomial. Such a linear projector $\Pi_{\mathcal{P}}$ from $\mathcal{G}$ onto $\mathcal{P}$ can be defined in many different ways. Here we simply assume that there are q linear functionals $\pi_1, \dots, \pi_q$ on $\mathcal{G}$ that are linearly independent over $\mathcal{P}$, i.e. the $q \times q$ matrix P with entries $\pi_k(p_j)$ is nonsingular. Then the projector can be represented as

(*DefPN*)

$$\Pi_{\mathcal{P}}(g) := \sum_{j=1}^{q} \pi_j(g) p_j. \tag{3.1.5}$$

By a change of basis in either the p_j or the π_j one can assume that the linear functionals $\pi_j(g)$ satisfy the system

$$\sum_{j=1}^{q} \pi_j(g)\pi_k(p_j) = \pi_k(g), \ 1 \le k \le q.$$

This is just another way of saying

$$\pi_k(\Pi_{\mathcal{P}}(g)) = \pi_k(g), \ 1 \le k \le q, \ g \in \mathcal{G},$$

and it has the consequence that $\Pi_{\mathcal{P}}(p) = p$ for all $p \in \mathcal{P}$, because of $\pi_j(p_k) = \delta_{jk}$.

Having $\Pi_{\mathcal{P}}$ at hand, we now form $R_{\mathcal{P}}(g) := g - \Pi_{\mathcal{P}}(g)$ for all $g \in \mathcal{G}$. For subsequent use we note that the bilinear form on $\mathcal{G}$ can now be rewritten as

(*Rsp*)

$$(f, g)_{\mathcal{G}} = (R_{\mathcal{P}}(f), R_{\mathcal{P}}(g))_{\mathcal{G}}, \quad f, g \in \mathcal{G}. \tag{3.1.6}$$

The decomposition of an arbitrary element $g \in \mathcal{G}$ as

$$g = \Pi_{\mathcal{P}}(g) + R_{\mathcal{P}}(g)$$

implies that the decomposition

(*gdec*)

$$\mathcal{G} = \mathcal{P} + R_{\mathcal{P}}(\mathcal{G}) \tag{3.1.7}$$

is a direct sum, since $R_{\mathcal{P}}(g) \in \mathcal{P}$ implies $g = \Pi_{\mathcal{P}}(g) + R_{\mathcal{P}}(g) \in \mathcal{P}$ and thus $R_{\mathcal{P}}(g) = 0$. Furthermore, the bilinear form $(\cdot, \cdot)_{\mathcal{F}}$ now is positive definite on $R_{\mathcal{P}}(\mathcal{G})$.

3.1.3 Golomb-Weinberger Technique

There is a way to avoid the explicit construction of the projector $\Pi_{\mathcal{P}}$ by simply modifying the bilinear form. If $p_1, \ldots, p_q$ are a basis of $\mathcal{P}$, one can define an inner product

$$\left(\sum_{j=1}^{q} \alpha_j p_j, \sum_{k=1}^{q} \beta_k p_k \right)_{\mathcal{P}} = \sum_{j,k=1}^{q} \alpha_j \beta_k$$

on $\mathcal{P}$ and replace the bilinear form $(\cdot, \cdot)_{\mathcal{G}}$ by the inner product

$$(\cdot, \cdot) := (R_{\mathcal{P}}(\cdot), R_{\mathcal{P}}(\cdot))_{\mathcal{G}} + (\Pi_{\mathcal{P}}(\cdot), (\Pi_{\mathcal{P}}(\cdot))_{\mathcal{P}}.$$

This does not require an explicit representation for the projector, but it implicitly uses the projector to split an element of $\mathcal{G}$ into two parts that fit into the new inner product. We do not pursue this technique further, though it sometimes facilitates certain arguments. It dates back to Golomb and Weinberger [16](*golomb-weinberger:59-1*), and the relation to the technique used later by Duchon and Madych/Nelson is described in [2](*Light-Wayne:96-1*).

3.1.4 Hilbert Space Completion

(*SecHSC*) We now complete the space $R_{\mathcal{P}}(\mathcal{G})$ in the usual way to form a Hilbert space $\mathcal{F}$, taking us back to the setting that we started from, and where

$$(\cdot,\cdot)_{\mathcal{F}} := (R_{\mathcal{P}}(\cdot), R_{\mathcal{P}}(\cdot))_{\mathcal{G}}$$

is the inner product. This completion works via Cauchy sequences modulo null sequences, and it allows all continuous mappings on $R_{\mathcal{P}}(\mathcal{G})$ to be extended to the completion. See Theorem 12.2.11 (*HSCT*) for details. We now define the closure of $\mathcal{G}$ as the direct sum of $\mathcal{P}$ with the closure $\mathcal{F}$ of $R_{\mathcal{P}}(\mathcal{G})$. Then the decomposition (3.1.7, *gdec*) extends to the closures, and if we denote the closure of $\mathcal{G}$ by $\mathcal{G}$ again, we get

(*GPF1*)

$$\mathcal{G} = \mathcal{P} + \mathcal{F}. \tag{3.1.8}$$

Thus we finally see that it makes no difference to start right away with a space $\mathcal{G}$ that allows a decomposition (3.1.7, *gdec*) such that (3.1.6, *Rsp*) is a scalar product on the Hilbert space $\mathcal{F} := R_{\mathcal{P}}(\mathcal{G})$ that has $\mathcal{P}$ as its nullspace.

We finish this section by checking the proper form of admissible functionals for recovery. If λ is just any functional on $\mathcal{G}$, it defines a functional $\lambda - \lambda\Pi_{\mathcal{P}} = \lambda R_{\mathcal{P}}$ by

(*lrest*)

$$g \mapsto \lambda(g) - \lambda(\Pi_{\mathcal{P}}(g)) = \lambda(R_{\mathcal{P}}(g)), \quad g \in \mathcal{G}, \tag{3.1.9}$$

and this functional is a good candidate for being continuous with respect to the seminorm $|\cdot|_{\mathcal{G}}$, because it vanishes on $\mathcal{P}$. We thus consider all functionals λ on $\mathcal{G}$ such that $\lambda - \lambda\Pi_{\mathcal{P}}$ is continuous, and we denote the space of these functionals by $\mathcal{G}^*$. By (3.1.9, *lrest*), for each $\lambda \in \mathcal{G}^*$ the functional $\lambda - \lambda \circ \Pi_{\mathcal{P}}$ is continuous on the Hilbert space $\mathcal{F} = R_{\mathcal{P}}(\mathcal{G})$, and by the Riesz theorem 12.2.14 (*RieszT*) there is an element $g_\lambda \in \mathcal{G}$ such that the identity

(*lrep*)

$$\lambda(g) - \lambda(\Pi_{\mathcal{P}}(g)) = \lambda(R_{\mathcal{P}}(g)) = (g, g_\lambda)_{\mathcal{G}} \tag{3.1.10}$$

holds for all $\lambda \in \mathcal{G}^*$ and all $g \in \mathcal{G}$. We shall use this identity in the more convenient form

$$\lambda(g) = \lambda(\Pi_{\mathcal{P}}(g)) + (g, g_\lambda)_{\mathcal{G}}$$

and note that g_λ is uniquely defined modulo $\mathcal{P}$, while $R_{\mathcal{P}}(g_\lambda)$ is unique. The functionals from (3.1.9, *lrest*) vanish on $\mathcal{P}$ and they form the dual $\mathcal{F}^*$ of $\mathcal{F}$. If one defines $\Pi^*_{\mathcal{P}}(\lambda) := \lambda \circ \Pi_{\mathcal{P}}$ and $\mathcal{P}^* = \Pi^*_{\mathcal{P}}(\mathcal{G}^*)$, then there are decompositions

$$\lambda = \Pi^*_{\mathcal{P}}(\lambda) + (\cdot, g_\lambda)_{\mathcal{G}}$$

$$\mathcal{G}^* = \mathcal{P}^* + \mathcal{F}^*$$

that correspond to those of $g \in \mathcal{G}$ and $\mathcal{G}$ itself.

We finally remark that the detour via the completion is unnecessary, if (3.1.10, *lrep*) is used as a hypothesis, not as a consequence. But we wanted to show that (3.1.10, *lrep*) does not need any extra assumptions.

3.1.5 Solutions of Optimal Recovery Problems

(*Necessity*) We now can return to the problem (3.1.4, *ORP*) of optimal recovery. The given functionals λ_j are assumed to be in $\mathcal{G}^*$. Then they satisfy (3.1.10, *lrep*) and introduce elements $g_j := g_{\lambda_j} \in \mathcal{G},\ 1 \le j \le M$ in the sense

(*lrepj*)

$$\lambda_j(g) - \lambda_j(\Pi_{\mathcal{P}}(g)) = \lambda_j(R_{\mathcal{P}}(g)) = (g, g_j)_{\mathcal{G}},\ g \in \mathcal{G}. \tag{3.1.11}$$

These elements are not unique, and we could make them unique by defining $g_j := R_{\mathcal{P}}(g_{\lambda_j}),\ 1 \le j \le M$, but the following results do not require this uniqueness. We now can characterize the solutions of the recovery problem:

Theorem 3.1.12 *(ORT1) Any solution g^* of the optimal recovery problem (3.1.4, ORP) with functionals $\lambda_1, \dots, \lambda_M \in \mathcal{G}^*$ satisfying (3.1.11, lrepj) has the form*

(grep)

$$g^* = \sum_{j=1}^{M} \alpha_j g_j + \sum_{i=1}^{q} \beta_i p_i \tag{3.1.13}$$

where the coefficients satisfy the linear system

(EQsys3)

$$\begin{array}{rcrcll} \displaystyle\sum_{j=1}^{M} \alpha_j (g_k, g_j)_{\mathcal{G}} & + & \displaystyle\sum_{i=1}^{q} \beta_i \lambda_k(p_i) & = & \gamma_k, & 1 \le k \le M \\ \displaystyle\sum_{j=1}^{M} \alpha_j \lambda_j(p_i) & + & 0 & = & 0, & 1 \le i \le q. \end{array} \tag{3.1.14}$$

and any solution of the above system solves the optimal recovery problem. However, the representation (3.1.13, grep) is not necessarily unique,

Note how similar (3.1.14, *EQsys3*) and (1.7.2, *EQsys2*) are, and note that we postpone the discussion of the solvability of (3.1.14, *EQsys3*).

Proof: We start by noting that g^* is a solution of (3.1.4, *ORP*) if and only if it satisfies the variational equation

(*charmin*)

$$(g^*, v)_{\mathcal{G}} = 0 \text{ for all } v \in \mathcal{G} \text{ with } \lambda_j(v) = 0,\ 1 \le j \le M. \tag{3.1.15}$$

This follows from Corollary 12.2.7 (*BAC*) in section 12.2 (*SecHSB*).

If $g^* \in \mathcal{G}$ satisfies (3.1.14, *EQsys3*) and $v \in \mathcal{G}$ satisfies the homogeneous conditions $\lambda_j(v) = 0,\ 1 \le j \le M$, then

$$\begin{array}{rcl} (g^*, v)_{\mathcal{G}} & = & \sum\limits_{j=1}^{M} \alpha_j (g_j, v)_{\mathcal{G}} + \sum\limits_{i=1}^{q} \beta_i (p_i, v)_{\mathcal{G}} \\ & = & \sum\limits_{j=1}^{M} \alpha_j \left(\lambda_j(v) - \lambda_j(\Pi_{\mathcal{P}}(v))\right) \\ & = & -\sum\limits_{j=1}^{M} \alpha_j \lambda_j(\Pi_{\mathcal{P}}(v)) \\ & = & 0 \end{array}$$

and g^* satisfies (3.1.15, *charmin*) and solves (3.1.4, *ORP*).

To prove the converse, we note that (3.1.15, *charmin*) implies the existence of $\alpha_1, \ldots, \alpha_M \in I\!R$ such that

(*charmin2*)

$$(g^*, v)_{\mathcal{G}} = \sum_{j=1}^{M} \alpha_j \lambda_j(v) \tag{3.1.16}$$

for all $v \in \mathcal{G}$. In fact, the linear map $v \mapsto (g^*, v)_{\mathcal{G}}$ vanishes on the kernel of the linear map $v \mapsto (\lambda_1(v), \ldots, \lambda_M(v))^T \in I\!R^M$ with finite-dimensional range and thus factorizes over the range of this mapping. See the proof of Corollary 12.2.7 (*BAC*) for this argument. But now (3.1.16, *charmin2*) implies

(*charmin3*)

$$(g^*, v)_{\mathcal{G}} = \sum_{j=1}^{M} \alpha_j \left(\lambda_j(\Pi_{\mathcal{P}}(v)) + (g_j, v)_{\mathcal{G}}\right) \tag{3.1.17}$$

and specialization to $v \in \mathcal{P}$ implies the second set of equations in (3.1.14, *EQsys3*). Then (3.1.17, *charmin3*) can be rewritten in the form

$$\left(g^* - \sum_{j=1}^{M} \alpha_j g_j, v\right)_{\mathcal{G}} = 0 \text{ for all } v \in \mathcal{G}$$

and this implies the representation (3.1.13, *grep*) of g^*. The interpolation conditions finally furnish the first set of equations in (3.1.14, *EQsys3*). $\square$

The system (3.1.14, *EQsys3*) looks terrible at first sight, because neither the functions g_j nor their inner products $(g_j, g_k)_{\mathcal{G}}$ are readily available from the given functionals λ_j. But we shall see in (3.2.14, *gjkrep*) that there is a conditionally positive definite function Φ such that

$$(g_j, g_k)_{\mathcal{G}} = \lambda_j^x \lambda_k^y \Phi(x, y)$$

holds for the elements of the matrix in (3.1.14, *EQsys3*), making an easy access to these elements possible, if Φ is explicitly known. In particular, if $\lambda_j(f) = f(x_j)$, then

$$(g_j, g_k)_{\mathcal{G}} = \Phi(x_j, x_k)$$

as we used in (1.7.2, *EQsys2*) in a slightly more special situation.

We now look at solvability of the system (3.1.14, *EQsys3*) in the shorthand form

(*BDef2*)

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \gamma \\ 0 \end{pmatrix} \tag{3.1.18}$$

generalizing (1.7.3, *BDef*). If vectors $\alpha \in I\!R^M$ and $\beta \in I\!R^q$ satisfy the homogeneous system, then

$$\begin{array}{rcccl} A\alpha & + & P\beta & = & 0 \\ P^T\alpha & + & 0 & = & 0 \end{array}$$

imply

$$\begin{array}{rcl} \alpha^T A \alpha & = & 0 \\ P^T\alpha & = & 0. \end{array}$$

Since the matrix A is a Gramian for the elements $g_1, \ldots, g_M$, it is positive semidefinite and we have

$$\alpha^T A \alpha = \left| \sum_{j=1}^M \alpha_j g_j \right|_{\mathcal{G}}^2$$

Thus the element $\sum_{j=1}^M \alpha_j g_j$ of $\mathcal{G}$ must be in $\mathcal{P}$ and the linear combination $\sum_{j=1}^M \alpha_j \lambda_j$ of functionals is zero due to $P^T\alpha = 0$ and

$$\sum_{j=1}^M \alpha_j \lambda_j(v) = \sum_{j=1}^M \alpha_j(\lambda_j(\Pi_{\mathcal{P}}(v)) + (g_j, v)_{\mathcal{G}}) = 0 + \left(\sum_{j=1}^M \alpha_j g_j, v \right)_{\mathcal{G}} = 0$$

for all $v \in \mathcal{G}$. But we assumed the linear functionals $\lambda_1, \ldots, \lambda_M$ to be linearly independent over $\mathcal{G}$. This implies $\alpha = 0$ and we are left with $P\beta = 0$. There is no way to deduce $\beta = 0$ from this in general, and consequently we have to add injectivity of P to our hypotheses, if we want to assure unique solvability of (3.1.14, *EQsys3*). We summarize:

Theorem 3.1.19 *(ORT2) There is a unique solution to the optimal recovery problem (3.1.4, ORP) if the $M \times q$ matrix P with entries*

$$\lambda_j(p_i),\ 1 \le j \le M,\ 1 \le i \le q$$

is injective. This condition means that the only element $p \in \mathcal{P}$ with vanishing data $\lambda_1(p), \ldots, \lambda_M(p)$ must be the zero element. □

It should be clear by now that we finally want to show how the system (1.7.2, *EQsys2*) is a special case of (3.1.14, *EQsys3*) and how a conditionally positive definite function Φ can arise in the above Hilbert space setting. We shall take point evaluation functionals $\lambda_x(v) := (v - \Pi_{\mathcal{P}}(v))(x)$ if the abstract elements $v \in \mathcal{G}$ can be interpreted as functions on some domain Ω containing the points x, and use the elements $g_x := g_{\lambda_x} \in \mathcal{G}$ from (3.1.10, *lrep*) to define a generalized conditionally positive definite function with $\mathcal{P}$ generalizing $I\!P_m^d$ by

$$\Phi(x, y) := (g_x, g_y)_{\mathcal{G}},\ x, y \in \Omega.$$

The details will be specified in Theorem 3.2.17 (*CPDNeccT)*.

Theorems 3.1.12 (*ORT1)* and 3.1.19 (*ORT2)* show that optimal recovery in the fairly general sense of (3.1.4, *ORP*) necessarily leads to solutions of the special form (3.1.13, *grep*) and linear systems (3.1.14, *EQsys3*). This is why the techniques of section 1.4 (*SubSectDDSpaces)* are a quite natural and general way to access recovery problems.

3.1.6 Related Problems

(*SecRP*) There is an equivalent dual reformulation of the above recovery problem. Instead of reconstructing some $g \in \mathcal{G}$ from the information $\gamma_j = \lambda_j(g),\ 1 \le j \le M$ one can ask for a functional $\lambda^* \in \mathcal{G}^*$ of minimal seminorm in $\mathcal{G}^*$ that satisfies the equations

$$\lambda^*(g_j) = \gamma_j,\ 1 \le j \le M$$

for a set of linearly independent elements $g_1, \ldots, g_M \in \mathcal{G}$. For this the dual bilinear form on functionals in $\mathcal{G}^*$ can be defined as

(*DefDualBil*)

$$(\lambda,\mu)_{\mathcal{G}^*} := (g_\lambda, g_\mu)_{\mathcal{G}} = \lambda(g_\mu) - \lambda(\Pi_{\mathcal{P}}(g_\mu)) = \mu(g_\lambda) - \mu(\Pi_{\mathcal{P}}(g_\lambda)). \quad (3.1.20)$$

The additional property required for uniqueness now is that the $M \times q$ matrix P with entries

$$\pi_i(g_j),\ 1 \le j \le M,\ 1 \le i \le q$$

is injective. This condition means that the zero is the only element in the span of $g_1, \ldots, g_M$ that projects via $\Pi_{\mathcal{P}}$ to zero in $\mathcal{P}$. We leave details to the readers as an exercise. There is a full duality if one replaces λ_j by g_j and π_i by p_i

Another related optimal recovery problem consists in finding an element $g^* \in \mathcal{G}$ with minimal seminorm $|g^*|_{\mathcal{G}}$ such that

(*scaleq*)

$$\begin{aligned} (g^*, g_j)_{\mathcal{G}} &= \gamma_j,\ 1 \le j \le M \\ \Pi_{\mathcal{P}}(g^*) &= 0, \end{aligned} \quad (3.1.21)$$

where we again assume that the functions g_j represent linear independent functionals λ_j in the sense of (3.1.11, *lrepj*). The difference is that the data now are not taking notice of additional functions from $\mathcal{P}$, such that the second condition of (3.1.21, *scaleq*) is necessary to remove the nonuniqueness of g^* modulo $\mathcal{P}$. Furthermore, one can assume

(*picond*)

$$\Pi_{\mathcal{P}}(g_j) = 0,\ 1 \le j \le M \quad (3.1.22)$$

without loss of generality.

Theorem 3.1.23 *(ORT3) Under the additional assumptions*

(spancond)

$$\sum_{j=1}^{M} \alpha_j g_j \in \mathcal{P} \textit{ implies } \alpha_j = 0,\ 1 \le j \le M \quad (3.1.24)$$

and (3.1.22, picond), the above optimal recovery problem with conditions (3.1.21, scaleq) has a unique solution g^* of the form

(grep2)

$$g^* = \sum_{j=1}^{M} \alpha_j g_j \quad (3.1.25)$$

where the coefficients satisfy the linear system

(EQsys4)

$$\sum_{j=1}^{M}\alpha_j(g_k,g_j)_{\mathcal{G}} \;=\; \gamma_k,\ 1\le k\le M. \qquad (3.1.26)$$

Proof: The equivalent variational equation here is

$$(g^*,v)_{\mathcal{G}}=0 \text{ for all } v\in\mathcal{G} \text{ with } \Pi_{\mathcal{P}}(v)=0 \text{ and } (v,g_j)_{\mathcal{G}}=0,\ 1\le j\le M.$$

This transforms into

$$(g^*,v)_{\mathcal{G}}=(\sum_{j=1}^{M}\alpha_j g_j, v)_{\mathcal{G}}$$

for all $v\in\mathcal{G}$. This is satisfied if (3.1.25, *grep2*) holds. To prove the converse, we conclude that the variational equation implies that the difference of both sides in (3.1.25, *grep2*) lies in $\mathcal{P}$. But application of $\Pi_{\mathcal{P}}$ turns the difference into zero, proving necessity of (3.1.25, *grep2*).

To prove nonsingularity of the system (3.1.26, *EQsys4*) we proceed similarly as in the proof of Theorem 3.1.19 (*ORT2*), but use (3.1.24, *spancond*) instead of linear independence of the functionals λ_j. □

Note that (3.1.24, *spancond*) is more restrictive than to assume linear independence of the functionals λ_j, as required for Theorem 3.1.12 (*ORT1*). This is why Theorem 3.1.23 (*ORT3*) has positive definiteness of the matrix $((g_i,g_j)_{\mathcal{G}})_{i,j}$, while Theorems 3.1.12 (*ORT1*) and 3.1.19 (*ORT2*) need the enlarged matrix. Furthermore, the functionals $\mu_j := (\cdot,g_j)_{\mathcal{G}}$ that implicitly arise in Theorem 3.1.23 (*ORT3*) have the additional property $\mu_j(\mathcal{P})=\{0\}$, and this property is not shared by the functionals λ_j in the previous theorems. In case of $\mathcal{P}=\{0\}$ there is no difference at all.

We now consider the best approximation problem

(*BAP*)

$$\inf_{\lambda\in\Lambda}|\mu-\lambda|_{\mathcal{G}^*} \qquad (3.1.27)$$

for a given functional $\mu\in\mathcal{G}^*$ by functionals in

(*DefL*)

$$\Lambda := \text{span}\,\{\lambda_1,\dots,\lambda_M\}\subset\mathcal{G}^*. \qquad (3.1.28)$$

The usual theory of approximation in spaces with inner products or bilinear forms yields the normal equations

$$(\mu, \lambda_j)_{\mathcal{G}^*} = \sum_{k=1}^{M} \alpha_k(\mu)(\lambda_k, \lambda_j)_{\mathcal{G}^*} = \sum_{k=1}^{M} \alpha_k(\mu)(g_k, g_j)_{\mathcal{G}}$$

with a coefficient matrix as in (3.1.26, *EQsys4*), and the optimal value of (3.1.27, *BAP*) is given by

(*BAPN*)

$$\begin{aligned} \inf_{\lambda\in\Lambda} |\mu - \lambda|^2_{\mathcal{G}^*} &= |\mu - \sum_{k=1}^{M} \alpha_k(\mu)\lambda_k|^2_{\mathcal{G}^*} \\ &= (\mu, \mu)_{\mathcal{G}^*} - 2\sum_{k=1}^{M} \alpha_k(\mu)(\lambda_k, \mu)_{\mathcal{G}^*} \\ &\quad + \sum_{j,k=1}^{M} \alpha_j(\mu)\alpha_k(\mu)(\lambda_j, \lambda_k)_{\mathcal{G}^*}. \end{aligned} \tag{3.1.29}$$

3.1.7 Properties of Optimal Recoveries

Assume that we used the method of section 3.1.5 (*Necessity*) to recover an element $g \in \mathcal{G}$ by some element g^* that satisfies

(*ljg*)

$$\lambda_j(g) = \lambda_j(g^*),\ 1 \le j \le M \tag{3.1.30}$$

for a set of linearly independent functionals $\lambda_1, \ldots, \lambda_M$ with representers g_j in the sense of

$$\lambda_j(v) = \lambda_j(\Pi_{\mathcal{P}}(v)) + (v, g_j)_{\mathcal{G}},\ v \in \mathcal{G}.$$

Assume further that the sufficient condition for uniqueness holds, as given in Theorem 3.1.19 (*ORT2*), and that we normalized the functions g_j to satisfy $g_j = R_{\mathcal{P}} g_j$ or $\Pi_{\mathcal{P}} g_j = 0$.

Since any element $g^* = p \in \mathcal{P}$ satisfies (3.1.15, *charmin*), we get

Theorem 3.1.31 *(PolRepT1) The optimal recovery process reproduces elements of $\mathcal{P}$.* □

Corollary 3.1.32 *(PolRepCol) If g^* is the unique optimal recovery of g, then $\Pi_{\mathcal{P}}(g - g^*) = 0$.*

Proof: If $p \in \mathcal{P}$ is arbitrary, then clearly $(g+p)^* = g^*+p$ due to uniqueness. The recovery process thus acts separately on the two parts of $\mathcal{G} = \mathcal{P} + R_{\mathcal{P}}(\mathcal{G})$ with values in the respective parts of $\mathcal{S} = \mathcal{P} + R_{\mathcal{P}}(\mathcal{S})$. But then $(R_{\mathcal{P}}g)^* = R_{\mathcal{P}}(g^*)$ holds and

$$R_{\mathcal{P}}(g^*) = (R_{\mathcal{P}}g)^* = (g - \Pi_{\mathcal{P}}g)^* = g^* - \Pi_{\mathcal{P}}g$$

implies $\Pi_{\mathcal{P}}g^* = \Pi_{\mathcal{P}}g$. $\square$

Turning to orthogonality relations, we have

$$(g_j, g - g^*)_{\mathcal{G}} + \lambda_j \Pi_{\mathcal{P}}(g - g^*) = 0, \ 1 \le j \le M$$

and for each element s from the space

(*DefS*)

$$\mathcal{S} = \left\{ \sum_{j=1}^{M} \alpha_j g_j + \sum_{k=1}^{q} \beta_k p_k \ : \ \sum_{j=1}^{M} \alpha_j \lambda_j(\mathcal{P}) = \{0\} \right\} \tag{3.1.33}$$

we get the orthogonality

(*EqOrtho*)

$$(s, g - g^*)_{\mathcal{G}} = 0 \tag{3.1.34}$$

by summation. But this means that g^* is a best approximation to g from $\mathcal{S}$:

Theorem 3.1.35 *(ORTBA) The solution g^* of the optimal recovery problem (3.1.4, ORP) for data from some element $g \in \mathcal{G}$ is a best approximation to g from the space $\mathcal{S}$ of (3.1.33, DefS) in the sense*

$$|g - g^*|_{\mathcal{G}} = \min_{s \in \mathcal{S}} |g - s|_{\mathcal{G}}.$$

$\square$

Equation (3.1.34, *EqOrtho*) easily generalizes to

Theorem 3.1.36 *(OrtTh) The orthogonal complement of the subspace (3.1.33, DefS) of $\mathcal{G}$ is $\mathcal{P}$ plus the space of all elements $v \in \mathcal{G}$ that have $\lambda_j(v) = 0, \ 1 \le j \le M$.*

Proof: The variational equation (3.1.15, *charmin*) shows that the orthogonal complement must contain the elements in question. Now let $g \in \mathcal{G}$ be an element in the orthogonal complement of (3.1.33, *DefS*) and form its optimal recovery g^*. Then use (3.1.15, *charmin*) and othogonality of g to g^* to get

$$(g - g^*, g - g^*)_{\mathcal{G}} = (g, g)_{\mathcal{G}} - (g, g^*)_{\mathcal{G}} - (g^*, g - g^*)_{\mathcal{G}} = (g, g)_{\mathcal{G}}.$$

But since (3.1.15, *charmin*) always implies orthogonality of $g - g^*$ to g^*, we have the Pythagorean law

(*Pyth*)

$$(g, g)_{\mathcal{G}} = (g - g^*, g - g^*)_{\mathcal{G}} + (g^*, g^*)_{\mathcal{G}} \tag{3.1.37}$$

which leads to $g^* \in \mathcal{P}$ in this special situation. Thus $g = g^* + g - g^*$ is of the required form. □

We now proceed towards the prototype of an error bound. We use the space (3.1.28, *DefL*) of functionals and (3.1.30, *ljg*) to get $\lambda(g - g^*) = 0$ for all $\lambda \in \Lambda$. Now take any $\mu \in \mathcal{G}^*$ and form

$$\begin{array}{rcl} \mu(g - g^*) & = & (\mu - \lambda)(g - g^*) \\ & \leq & |(\mu - \lambda)\Pi_{\mathcal{P}}(g - g^*)| + |(g_\mu - g_\lambda, g - g^*)_{\mathcal{G}}| \\ & \leq & |\mu - \lambda|_{\mathcal{G}^*}|g - g^*|_{\mathcal{G}}, \end{array}$$

using Corollary 3.1.32 (*PolRepCol)*.

Theorem 3.1.38 *(ORTFA) [1](DNW) The solution g^* of the optimal recovery problem (3.1.4, ORP) for data from some element $g \in \mathcal{G}$ satisfies the error bound*

(Eq2inf)

$$|\mu(g - g^*)| \leq \inf_{\lambda \in \Lambda} |\mu - \lambda|_{\mathcal{G}^*} \inf_{s \in \mathcal{S}} |g - s|_{\mathcal{G}} \tag{3.1.39}$$

for any functional $\mu \in \mathcal{G}^*$. □

The crucial factor in the error bound (3.1.39, *Eq2inf*) is the **generalized optimal power function**

(*GPDef*)

$$P(\mu) := P_\Lambda(\mu) := \inf_{\lambda \in \Lambda} |\mu - \lambda|_{\mathcal{G}^*} \tag{3.1.40}$$

with Λ from (3.1.28, *DefL*). If the functionals λ_j are "near" to μ, this quantity should be rather small, and we shall prove specific bounds later in 5.5 (*hrhodef)*. This is made possible by the representation for $P(\mu)$ that follows readily from (3.1.27, *BAP*) and (3.1.29, *BAPN*), and which will also be useful in section 4.6 (*URT)*.

3.1.8 Remarks

The theory of optimal recovery starts with the early paper of Golomb and Weinberger [16](*golomb-weinberger:59-1*), while reproducing kernel Hilbert spaces are much older (see e.g. the textbook by Meschkowski [26](*meschkowski:62-1*)). A milestone was the theory of optimal recovery in the sense of Micchelli, Rivlin, and Winograd ([28](*micchelli-rivlin:77-1*) [29](*micchelli-rivlin:78-1*) [30](*micchelli-rivlin:84-1*) [31](*micchelli-et-al:76-1*)), while the current extension into the direction of information-based complexity is in [6](*bojanov-wozniakowski:92-1*).

3.2 Spaces of Functions

(*SecSF*) In order to arrive at conditionally positive **functions**, we now have to specialize our results on optimal recovery to the case of optimal recovery of functions.

3.2.1 From Hilbert Spaces to Positive Definite Functions

(*SecHSPDF*) We now specialize to a Hilbert space $\mathcal{F}$ of functions on some domain Ω that we do not restrict. But since classical functions are objects that allow the action of specific linear functionals

(*deltadef*)

$$\delta_x \;:\; g \mapsto g(x),\; g \in \mathcal{F},\; x \in \Omega \tag{3.2.1}$$

called point-evaluation functionals, we assume that the above functionals δ_x are in $\mathcal{F}^*$ and thus continuous on $\mathcal{F}$. Then one can invoke the Riesz representation theorem 12.2.14 (*RieszT*) to get a function $g_{\delta_x} \in \mathcal{F}$ for each $x \in \Omega$ such that

(*DRKF*)

$$g(x) = \delta_x(g) = (g, g_{\delta_x})_{\mathcal{F}} \tag{3.2.2}$$

holds for all $g \in \mathcal{F}$, $x \in \Omega$. We now define a function

$$\Phi \;:\; \Omega \times \Omega \to I\!R,\; \Phi(x,y) := (g_{\delta_x}, g_{\delta_y}),\; x, y \in \Omega$$

and get

Theorem 3.2.3 *(PDFT1) If the point evaluation functionals in a Hilbert space $\mathcal{F}$ of functions on some domain Ω are continuous, then the space has a* reproducing kernel function Φ *with the following properties:*

1. $\Phi : \Omega \times \Omega \to I\!R$,

2. $\Phi(x,\cdot) = \Phi(\cdot,x) \in \mathcal{F}$ *for all* $x \in \Omega$,

3. $g(x) = (g, \Phi(x,\cdot))_{\mathcal{F}}$ *for all* $g \in \mathcal{F},\ x \in \Omega$.

Proof: By definition and (3.2.2, *DRKF*),

$$\begin{array}{rcccl} g_{\delta_y}(x) & = & (g_{\delta_y}, g_{\delta_x})_{\mathcal{F}} & = & \Phi(y,x) \\ (g_{\delta_y}, g_{\delta_x})_{\mathcal{F}} & = & (g_{\delta_x}, g_{\delta_y})_{\mathcal{F}} & = & \Phi(x,y) = \Phi(y,x) \end{array}$$

for all $x, y \in \Omega$, proving all of the assertions. □

We now compare this with Definition 1.5.3 (*DPD)* from section 1.5 (*subsecEIA)* on page 11 which we restate here for convenience:

Definition 3.2.4 *A real-valued function*

$$\Phi : \Omega \times \Omega \to I\!R$$

is a **positive definite function** *on* Ω, *iff for any choice of finite subsets* $X = \{x_1, \ldots, x_M\} \subseteq \Omega$ *of* M *different points the matrix*

$$A_{X,\Phi} = (\Phi(x_k, x_j))_{1 \le j,k \le M}$$

is positive definite.

To test the function Φ from Theorem 3.2.3 (*PDFT1)* for positive definiteness, consider a finite subset $X = \{x_1, \ldots, x_M\} \subseteq \Omega$ of M different points and take an arbitrary vector $\alpha \in I\!R^M$. Then

$$\alpha^T A_{X,\Phi} \alpha = \sum_{j,k=1}^{M} \alpha_j \alpha_k \Phi(x_k, x_j) = \left(\sum_{j=1}^{M} \alpha_j g_{x_j}, \sum_{k=1}^{M} \alpha_k g_{x_k} \right)_{\mathcal{F}} = \|\sum_{j=1}^{M} \alpha_j g_{x_j}\|_{\mathcal{F}}^2$$

implies that the matrix $A_{X,\Phi}$ always is positive semidefinite, because it is the Gramian of the functions $g_{x_j},\ 1 \le j \le M$. It is positive definite if and only if these functions are linearly independent in $\mathcal{F}$. Furthermore, is is easy to see from (3.2.2, *DRKF*) that the functions $g_{x_j},\ 1 \le j \le M$ are linearly dependent if and only if the point evaluation functionals $\delta_{x_j},\ 1 \le j \le M$ are linearly dependent in the dual space $\mathcal{F}$. Another simple exercise is to show equivalence of the linear independence of $\delta_{x_j},\ 1 \le j \le M$ with each of the following notions:

Definition 3.2.5 *(DFSP) A space $\mathcal{F}$ of functions on some domain Ω has the* **finite separation property**, *if for all finite subsets $X = \{x_1, \ldots, x_M\} \subseteq \Omega$ of M different points there are M functions $g_1, \ldots, g_M \in \mathcal{F}$ that separate the points in $X = \{x_1, \ldots, x_M\}$, i.e.*

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

Definition 3.2.6 *(DFIP) A space $\mathcal{F}$ of functions on some domain Ω has the* **finite interpolation property**, *if for all finite subsets $X = \{x_1, \ldots, x_M\} \subseteq \Omega$ of M different points and all vectors $\alpha \in I\!R^M$ there is a function $g \in \mathcal{F}$, depending on $X = \{x_1, \ldots, x_M\}$ and α, such that*

$$g(x_k) = \alpha_k,\ 1 \leq k \leq M.$$

We combine this into a result that proves the setting in 1.5 (*subsecEIA)* to occur naturally in fairly general situations:

Theorem 3.2.7 *Let $\mathcal{F}$ be a space of real-valued functions on some domain Ω, and assume*

1. *$\mathcal{F}$ is a Hilbert space over $I\!R$,*
2. *the point evaluation functionals (3.2.1, deltadef) are continuous on $\mathcal{F}$,*
3. *$\mathcal{F}$ has the finite interpolation or the finite separation property.*

Then $\mathcal{F}$ is a reproducing kernel Hilbert space, and its kernel function Φ : $\Omega \times \Omega$ is a positive definite function. □

3.2.2 Generalization towards Conditionally Positive Definite Functions

(*SecGCPDF*) We now return to the slightly more general setting of section 3.1.1 (*subsecORP)*. The continuous linear functionals now have to vanish on the kernel $\mathcal{P}$ of the bilinear form $(\cdot,\cdot)_{\mathcal{G}}$, and this is not a usual property of point evaluation functionals. But we can resort to the functionals

(*deltadef2*)

$$\delta_{x,\mathcal{P}} := \delta_x - \delta_x(\Pi_{\mathcal{P}}) \tag{3.2.8}$$

that will vanish on $\mathcal{P}$ for all $x \in \Omega$. We thus should require the functionals $\delta_{x,\mathcal{P}}$ from (3.2.8, *deltadef2*) to be continuous with respect to the bilinear form $(\cdot,\cdot)_{\mathcal{G}}$. This is the same as to assume that the point evaluation functionals δ_x are in $\mathcal{G}^*$, and then we can use (3.1.10, *lrep*) to get the generalization

(*DRKF2*)

$$\delta_{x,\mathcal{P}}(g) = g(x) - (\Pi_{\mathcal{P}}(g))(x) = (g, g_{\delta_x})_{\mathcal{G}} \tag{3.2.9}$$

of (3.2.2, *DRKF*) for all $g \in \mathcal{G}$, $x \in \Omega$. This is a special form of (3.1.10, *lrep*) on page 32 and yields the Taylor-type formula

(*Taylor*)

$$g(x) = (\Pi_{\mathcal{P}}(g))(x) + (g, g_{\delta_{x,\mathcal{P}}})_{\mathcal{G}} \tag{3.2.10}$$

for all $g \in \mathcal{G}$, $x \in \Omega$. We now define

(*DefPhiGen*)

$$\Phi \;:\; \Omega \times \Omega \to \mathbb{R},\; \Phi(x,y) := (g_{\delta_{x,\mathcal{P}}}, g_{\delta_{y,\mathcal{P}}})_{\mathcal{G}},\; x, y \in \Omega \tag{3.2.11}$$

and get

Theorem 3.2.12 *(CPDFT1) If the functionals (3.2.8,* deltadef2*) for a space $\mathcal{G}$ of functions on some domain Ω are continuous with respect to the bilinear form $(\cdot,\cdot)_{\mathcal{G}}$ with finite-dimensional kernel $\mathcal{P}$ and projector $\Pi_{\mathcal{P}} \;:\; \mathcal{G} \to \mathcal{P}$, then the space has a* reproducing kernel function Φ *with the following properties:*

1. $\Phi \;:\; \Omega \times \Omega \to \mathbb{R}$,
2. $\Phi(x,\cdot) = \Phi(\cdot,x) \in \mathcal{G}$ *for all* $x \in \Omega$,
3. $\Pi_{\mathcal{P}}\Phi(x,\cdot) = \Pi_{\mathcal{P}}\Phi(\cdot,x) = 0$ *for all* $x \in \Omega$,
4. $\Phi(x,y) = (\Phi(x,\cdot), \Phi(y,\cdot))_{\mathcal{G}}$ *for all* $x, y \in \Omega$
5. $g(x) = \Pi_{\mathcal{P}}(g)(x) + (g, \Phi(x,\cdot))_{\mathcal{G}}$ *for all* $g \in \mathcal{G},\; x \in \Omega$.

Proof: We proceed exactly as in Theorem 3.2.3 (*PDFT1)* and get

(*PhiRep2*)

$$\Phi(x,y) = (g_{\delta_{x,\mathcal{P}}}, g_{\delta_{y,\mathcal{P}}})_{\mathcal{G}} = g_{\delta_{x,\mathcal{P}}}(y) - (\Pi_{\mathcal{P}} g_{\delta_{x,\mathcal{P}}})(y). \tag{3.2.13}$$

This proves properties 2 and 3, while 1 holds by definition. Putting the above identity into (3.2.10, *Taylor*) and (3.2.11, *DefPhiGen*) yields the fourth and fifth property. □

We shall see later that the well-known conditionally positive definite functions fail to satisfy some of these properties, but there is a fairly standard process that shows how to get the properties by slight modifications. We shall comment on this when we consider the construction of native Hilbert spaces from given conditionally positive definite functions in section 3.3 (*SecNS*).

The identity (3.1.10, *lrep*) on page 32 introduced a representing function $g_\lambda \in \mathcal{G}$ for each functional $\lambda \in \mathcal{G}^*$. This was used in (3.1.11, *lrepj*) to derive the system (3.1.14, *EQsys3*) for solving the recovery problem. To bring this into line with the system (1.7.2, *EQsys2*) on page 15, we use (3.2.13, *PhiRep2*) to form

$$\begin{array}{rcl} \lambda^y\Phi(x,y) & = & \lambda(g_{\delta_{x,\mathcal{P}}}) - \lambda\Pi_{\mathcal{P}}g_{\delta_{x,\mathcal{P}}} \\ & = & (g_\lambda, g_{\delta_{x,\mathcal{P}}})_{\mathcal{G}} = g_\lambda(x) - (\Pi_{\mathcal{P}}g_\lambda)(x) \end{array}$$

and get

$$g_\lambda = \Pi_{\mathcal{P}}g_\lambda + \lambda^y\Phi(\cdot,y)$$

for all $\lambda \in \mathcal{G}^*$. Since g_λ is nonunique modulo functions from $\mathcal{P}$, we even can omit the first summand and use the above equation as a definition for g_λ. With a second functional $\mu \in \mathcal{G}^*$ we can write

$$\begin{array}{rcl} \mu^x\lambda^y\Phi(x,y) & = & \mu g_\lambda - \mu\Pi_{\mathcal{P}}g_\lambda \\ & = & \mu\Pi_{\mathcal{P}}g_\lambda + (\mu,\lambda)_{\mathcal{G}^*} - \mu\Pi_{\mathcal{P}}g_\lambda \\ & = & (\mu,\lambda)_{\mathcal{G}^*}. \end{array}$$

This proves

(*gjkrep*)

$$(g_j, g_k)_{\mathcal{G}} = (\lambda_j, \lambda_k)_{\mathcal{G}^*} = \lambda_j^x\lambda_k^y\Phi(x,y) \tag{3.2.14}$$

for the elements of the matrix in (3.1.14, *EQsys3*).

We now want to move towards conditionally positive definite functions, but we still have to replace polynomials in Definition 1.6.2 (*DCPD*) on page 13:

Definition 3.2.15 *(DCPD2) A real-valued function*

$$\Phi : \Omega \times \Omega \to I\!R$$

is a **conditionally positive definite function** *with respect to a finite-dimensional space $\mathcal{P}$ of functions on Ω, iff for any choice of finite subsets $X = \{x_1, \ldots, x_M\} \subseteq \Omega$ of M different points the value*

$$\alpha^T A_{X,\Phi}\alpha := \sum_{j,k=1}^{M} \alpha_j\alpha_k\Phi(x_j,x_k)$$

of the quadratic form (1.6.1, QFdef) is positive, provided that the vector $\alpha = (\alpha_1, \ldots, \alpha_M) \in I\!R^M \setminus \{0\}$ has the additional property

(CPDef2)

$$\sum_{j=1}^{M} \alpha_j p(x_j) = 0 \tag{3.2.16}$$

for all $p \in \mathcal{P}$.

Theorem 3.2.17 *(CPDNeccT) Let $\mathcal{G}$ be a space of real-valued functions on some domain Ω, and assume*

1. *$\mathcal{G}$ has a real-valued symmetric bilinear form $(\cdot,\cdot)_{\mathcal{G}}$ with a finite dimensional kernel $\mathcal{P}$ and corresponding projector $\Pi_{\mathcal{P}}$,*
2. *the point evaluation functionals (3.2.8,* deltadef2*) are continuous with respect to the bilinear form,*
3. *$\mathcal{G}$ has the finite interpolation or the finite separation property.*

Then $\mathcal{G}$ has a reproducing kernel in the sense of Theorem 3.2.12 (CPDFT1), and its kernel function $\Phi\ :\ \Omega \times \Omega$ is a conditionally positive definite function with respect to $\mathcal{P}$.

Proof: Again, we consider a finite subset $X = \{x_1, \ldots, x_M\} \subseteq \Omega$ of M different points, but now we take a vector $\alpha \in I\!\!R^M$ with (3.2.16, *CPDef2*). Then we can repeat the steps of the proof of Theorem 3.2.3 (*PDFT1)* to see that the matrix $A_{X,\Phi}$ is positive semidefinite. To prove definiteness, we now assume that

(*inP*)

$$\sum_{j=1}^{M} \alpha_j g_{\delta_{x_j},\mathcal{P}} \in \mathcal{P} \tag{3.2.18}$$

holds and have to prove that α is zero. But (3.2.16, *CPDef2*) and (3.2.18, *inP*) imply via (3.2.9, *DRKF2*) that the point evaluation functionals δ_{x_j}, $1 \leq j \leq M$ are linearly dependent. $\square$

We see that conditionally positive definite functions arise necessarily whenever optimal recovery of functions from a space $\mathcal{G}$ with a bilinear form is attempted. The coefficient matrix of the major part of the linear system has elements of Gramian form $(g_j, g_k)_{\mathcal{G}}$, even if the recovery is carried out in more general (non-function-) spaces. This means that positive (semi-) definiteness is the natural condition to ask for, and there is no reason to replace it by nonsingularity.

3.2.3 Sobolev and Beppo-Levi Spaces

We now want to exhibit some special cases where we can start from a space $\mathcal{G}$ with bilinear form and arrive at a conditionally positive definite function. The most usual bilinear form defined on functions is the L_2 inner product

$$(f,g)_{L_2(\Omega)} := \int_{x\in\Omega} f(x)g(x)dx$$

However, point evaluation functionals are not continuous with respect to this inner product. This is easy to see when looking at the evaluation at zero of functions of the form $f_\alpha(x) := \exp(-\alpha\|x\|_2^2)$ for large positive α. The $L_2(\mathbb{R}^d)$ inner products tend to zero for $\alpha \to \infty$, while the value at zero is always one. Thus there is no positive constant C such that

$$|\lambda(f_\alpha)| \leq C\|f_\alpha\|_{L_2(\mathbb{R}^d)}$$

holds. As a warm-up for similar calculations occurring in later sections of the text, let us do the evaluation of the inner product. It suffices to take $\beta = 2\alpha$ and calculate the integral

$$\int_{x\in\Omega} \exp(-\beta\|x\|_2^2)dx = \mathrm{vol}\,(S^{d-1})\int_0^\infty r^{d-1}\exp(-\beta r^2)dr$$

by going over to polar coordinates and integrating over the scaled unit sphere $S^{d-1} \subset \mathbb{R}^d$. Its surface area (or its $d-1$-dimensional volume) is $\mathrm{vol}\,(S^{d-1}) = 2\pi^{(d-1)/2}/?\,((d-1)/2)$ due to (12.3.4, *VolS*). The rest follows from substitution and the definition (12.3.1, *GammaDef*) of the Gamma function:

$$\begin{array}{rcl}\displaystyle\int_0^\infty r^{d-1}\exp(-\beta r^2)dr & = & \frac{1}{2\beta}\displaystyle\int_0^\infty (\frac{t}{\beta})^{d/2-1}\exp(-t)dt \\ & = & \frac{1}{2}\beta^{-d/2}?\,(d/2).\end{array}$$

If the reader has difficulties with this, it is time to work through part 12.3 (*SecSFT*) of the appendix.

To make point evaluation functionals continuous, we require a stronger bilinear form than just the L_2 inner product. And the above discussion shows that problems may get worse with increasing space dimension.

The usual trick is to introduce derivatives into the bilinear form. In particular, take a multiindex $\alpha \in \mathbb{Z}_{\geq 0}^d$ and define f^α as the multivariate derivative of order α of some function f. For a fixed integer $m \geq 0$, assemble all derivatives with $|\alpha| := \|\alpha\|_1 = m$ into a positive semidefinite bilinear form

$$(f,g)_m := \int_\Omega \sum_{|\alpha|=m} \binom{m}{\alpha} f^\alpha(x) g^\alpha(x) dx$$

on all functions that are at least in $C^m(\Omega)$. Here, we used the multivariate version of

$$\binom{m}{\alpha} := \frac{m!}{\alpha_1! \ldots, \alpha_d!} \text{ with } |\alpha| = m.$$

For simply connected domains Ω with a nonzero interior in $I\!R^d$ the nullspace of the bilinear form will then coincide with the space $\mathcal{P} = I\!P_m^d$ of polynomials of order m on $I\!R^d$. To do this, we need that a C^m function on Ω with vanishing derivatives of order m must necessarily be a polynomial, and this works nicely in the interior of Ω by application of the multivariate Taylor formula. The boundary does not count for the integral, and the polynomial is unique, if we do not have multiple components of the domain.

However, we still have to check the continuity of point-evaluation functionals $\delta_{x,\mathcal{P}}$ in the sense of (3.2.8, *deltadef2*) on page 44. The construction of a suitable projector $\Pi_\mathcal{P}$ to the nullspace $\mathcal{P} = I\!P_m^D$ will be given in Lemma 5.4.3 (*LemPIG*) on page 115 for use in a different context, but it is actually no big deal. Much more serious is the proof of the fact that $m > d/2$ is necessary and sufficient for continuity of the point-evaluation functionals. This is called the **Sobolev inequality**, but its proof is delayed to 12.6 (*SecSob*).

If we assume $m > d/2$ and start with of $\mathcal{G} = C^m(\Omega)$ in the sense of section 3.2.2 (*SecGCPDF*), we still have to form the Hilbert space completion and to derive the functions $g_{\delta_{x,\mathcal{P}}}$ that occur in (3.2.9, *DRKF2*) and allow to define a normalized conditionally positive definite function Φ via (3.2.11, *DefPhiGen*). To do these things on the full space $I\!R^d$ will later turn out to be much easier than to use a compact domain Ω. To avoid problems with nonexistence of $\|f\|_m$, we restrict ourselves to the subspace of $C^m(I\!R^d)$ of functions with bounded seminorm $|\cdot|_m$. The resulting completed space $\mathcal{G}$ with the bilinear form $(\cdot,\cdot)_m$ is called the **Beppo-Levi space** of order m on $I\!R^d$. For readers without a background in partial differential equations it will probably be a surprise to hear that the resulting Φ then precisely is the normalization of the conditionally positive definite radial function $\phi(r) = r^{2m-d}$ for d odd and $\phi(r) = r^{2m-d} \log r$ for d even.

We give a brief and sloppy "physicist-style" explanation for this and do the strict proof the other way round: we later construct the space from the conditionally positive definite function along the lines of the next section. The informal technique just takes (3.2.9, *DRKF2*) for granted and rewrites it in the form

$$\begin{aligned} \delta_{x,\mathcal{P}}(g) &= \int_\Omega \sum_{|\alpha|=m} \binom{m}{\alpha} g^\alpha(y) g^\alpha_{\delta_{x,\mathcal{P}}}(y)dy \\ &= (-1)^m \int_\Omega g(y) \sum_{|\alpha|=m} \binom{m}{\alpha} g^{2\alpha}_{\delta_{x,\mathcal{P}}}(y)dy \end{aligned}$$

if boundary terms are neglected. Thus, in the sense of linear partial differential equations, the function $g_{\delta_{x,\mathcal{P}}}$ must (up to a sign) be a fundamental solution corresponding to the differential operator

$$g \mapsto (-1)^m \sum_{|\alpha|=m} \binom{m}{\alpha} g^{2\alpha}$$

which (by a simple inductive proof) coincides with the m-th power $(-1)^m \Delta^m$ of the negative Laplacian

$$\Delta(f) := \sum_{j=1}^d \frac{\partial^2 f}{\partial x_j^2}.$$

This is the hidden reason for the $\binom{m}{\alpha}$ factors in the definition of the bilinear form. The corresponding fundamental solutions are well-known and must be radial due to the radial symmetry of the Laplacian. Using the radial form of the Laplacian, they can be calculated explicitly, and they always are either of the form r^γ or $r^\gamma \log r$. The boundary conditions, when evaluated properly, force to take the solution with maximal smoothness in zero or with minimal decay at infinity, and this is the radial function given above.

The case $d = 2$ requires $m > d/2 = 1$, and the minimal possible m leads to $m = 2$ and $\phi(r) = r^2 \log r$. The corresponding differential operator is Δ^2, describing the surfaces formed by thin plates under external forces or constraints. This is where **thin-plate splines** have their name, and the original approach by Duchon started from the partial differential equation background of these functions. The other cases are fundamental solutions of the iterated Laplacian, and since solutions of the plain Laplacian are called **harmonic functions**, the radial functions of the form $\phi(r) = r^\beta$ for $\beta \notin 2\mathbb{Z}$ or $\phi(r) = r^\beta \log r$ for $\beta \in 2\mathbb{Z}$ are called **polyharmonic functions**. The transition to non-integer values of β is possible via Fourier transforms and will be done in general later.

Looking back at the seminorm $|\cdot|_m$ induced by the bilinear form $(\cdot,\cdot)_m$, we see that the optimal recovery problem attempts to pick a function with least weighted mean square of all derivatives of order m. This is somewhat like an energy minimization in case $m = 2$, but $m = 2$ is admissible only in spaces of dimension up to $d = 3$.

Another even more important space arises when all derivatives up to order m are summed up to generate a new bilinear form

$$((f,g))_m := \sum_{j=0}^{m} \sum_{|\alpha|=j} \int_\Omega \frac{\partial^\alpha f}{\partial x^\alpha} \frac{\partial^\alpha g}{\partial x^\alpha} dx$$

This is positive definite and defines via completion a Hilbert space $W_2^m(\Omega)$ called **Sobolev space** of order m. Again, the point evaluation functionals are continuous only if $m > d/2$ holds. Using Fourier transforms, the special case $\Omega = I\!R^d$ can be treated explicitly and yields a positive definite radial basis function

$$\phi(r) = r^{m-d/2} K_{m-d/2}(r)$$

up to a factor depending on m and d, where K_ν is the Bessel or Macdonald function defined in (12.3.22, *KnuDef*). The power of r cancels the singularity of $K_{m-d/2}$ at zero exactly, since the asymptotics near zero are given by (12.3.23, *KnuAsyZero*).

These radial basis functions look strange, but they arise very naturally, Since the Bessel functions K_ν have exponential decay towards infinity due to (12.3.24, *KnuAsyInf*), the translates of $\phi(\|x\|_2)$ lead to virtually band-limited interpolation matrices. The evaluation of such functions is easily possible by calling standard subroutine packages.

If one considers other (equivalent) inner products on Sobolev spaces, the associated positive definite functions Φ will change. Naively, we would not expect these changes to be substantial, but surprisingly there is an equivalent inner product that generates a compactly supported radial basis function. We shall see this when we check the functions introduced by Wendland in [46](*wendland:95-1*).

3.2.4 Invariance Principles

(*SecIP*) The preceding discussion showed that conditionally positive definite functions associated to function spaces on $I\!R^d$ often come out to be radial. We shall now look at this phenomenon in more detail.

Assume that the domain Ω allows a group $\mathcal{T}$ of geometric transformations, and that the bilinear form $(\cdot,\cdot)_{\mathcal{G}}$ of the space $\mathcal{G}$ is invariant under transformations from $\mathcal{T}$. By this we mean the properties

(GInv)

$$\begin{array}{rcl} g \circ T & \in & \mathcal{G} \\ (f \circ T, g \circ T)_{\mathcal{G}} & = & (f, g)_{\mathcal{G}} \\ (\Pi_{\mathcal{P}} g) \circ T & = & \Pi_{\mathcal{P}}(g \circ T) \end{array} \tag{3.2.19}$$

for all $T \in \mathcal{T}$ and all $f, g \in \mathcal{G}$. Then there are two ways to interpret the action of a functional δ_{Tx} for $x \in \Omega$ and $T \in \mathcal{T}$:

$$\begin{array}{rclcl} \delta_{Tx}(g) & = & g(Tx) & = & (\Pi_{\mathcal{P}} g)(Tx) + (g, g_{\delta_{Tx}})_{\mathcal{G}} \\ & = & (g \circ T)(x) & = & (\Pi_{\mathcal{P}}(g \circ T))(x) + (g \circ T, g_{\delta_x})_{\mathcal{G}} \\ & & & = & (\Pi_{\mathcal{P}} g)(Tx) + (g \circ T, g_{\delta_x} \circ T^{-1} \circ T)_{\mathcal{G}} \\ & & & = & (\Pi_{\mathcal{P}} g)(Tx) + (g, g_{\delta_x} \circ T^{-1})_{\mathcal{G}} \end{array}$$

and this proves

$$g_{\delta_{Tx}} - g_{\delta_x} \circ T^{-1} \in \mathcal{P}$$

for all $g \in \mathcal{G}$, $T \in \mathcal{T}$. But this can be inserted into the definition of Φ to get

$$\Phi(Tx, Ty) = (g_{\delta_{Tx}}, g_{\delta_{Ty}})_{\mathcal{G}} = (g_{\delta_x} \circ T^{-1}, g_{\delta_y} \circ T^{-1})_{\mathcal{G}} = (g_{\delta_x}, g_{\delta_y})_{\mathcal{G}} = \Phi(x, y)$$

for all $x, y \in \Omega$. We thus have

Theorem 3.2.20 *(InvT1) Let $\mathcal{G}$ and Φ satisfy the assumptions of Theorem 3.2.12 (CPDFT1). If the domain Ω allows a group $\mathcal{T}$ of transformations that leave the bilinear form $(\cdot,\cdot)_{\mathcal{G}}$ on $\mathcal{G}$ invariant in the sense of (3.2.19, GInv), then Φ is invariant under $\mathcal{T}$ in the sense*

(PhiInv)

$$\Phi(x, y) = \Phi(Tx, Ty) \tag{3.2.21}$$

for all $x, y \in \Omega$, $T \in \mathcal{T}$. □

Corollary 3.2.22 *If the domain Ω has a fixed element denoted by x_0, and if for all $x \in \Omega$ there is a transformation $T_x \in \mathcal{T}$ with $T_x(x) = x_0$, then Φ takes the form*

(Phi1arg)

$$\Phi(x, y) = \Phi(T_y(x), x_0) \tag{3.2.23}$$

such that one of the two arguments of Φ is redundant.

We now consider some examples of domains with groups of transformations, and we always assume the invariance requirements of Theorem 3.2.20 (*InvT1)* to be satisfied.

Example 3.2.24 *If Ω is itself a group with neutral element 1, then*

$$\Phi(x, y) = \Phi(y^{-1}x, 1)$$

for all $x, y \in \Omega$.

Example 3.2.25 *If $\Omega = \mathbb{R}^d$ with the group of translations, then*

(PhiDiff)

$$\Phi(x, y) = \Phi(y - x, 0) = \Phi(x - y, 0) \tag{3.2.26}$$

for all $x, y \in \mathbb{R}^d$.

Example 3.2.27 *If $\Omega = \mathbb{R}^d$ with the group of Euclidean rigid-body transformations (i.e. translations and rotations), then Φ is a radial function*

$$\Phi(x, y) = \phi(\|y - x\|_2)$$

for all $x, y \in \mathbb{R}^d$, where $\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}$.

Proof: First use the translations of the previous case to write $\Phi(x, y) = \Phi(x - y, 0)$, and then rotate $x - y$ to a fixed unit vector in $\mathbb{R}^d$ multiplied by $\|x - y\|_2$. Then we are left with a scalar function of $\|x - y\|_2$. □

We note the remarkable fact that conditionally positive definite radial basis functions always occur in optimal recovery problems on $\mathbb{R}^d$ for functions from spaces that carry a bilinear form with Euclidean invariance.

Example 3.2.28 *If $\Omega = S^{d-1} \subset \mathbb{R}^d$ is the $(d-1)$-sphere, i.e. the surface of the unit ball in $\mathbb{R}^d$, then rotational invariance implies that Φ is* **zonal**, *i.e.*

$$\Phi(x, y) = \phi(x^T y)$$

for all $x, y \in S^{d-1}$, where $\phi : [0, 1] \to \mathbb{R}$.

In this case the function Φ can be written as a scalar function of the angle between the two arguments, or the cosine of this angle.

Example 3.2.29 *If $\Omega = \mathbb{R}^d$ and if the group $\mathcal{T}$ is $\mathbb{Z}^d$ under addition, then $\mathcal{G}$ is a shift-invariant space (see [7](*boor-et-al:94-2*)), and Φ is fully determined by its values on $\mathbb{R}^d_{\geq 0} \times \mathbb{R}^d_{\geq 0}$.*

In this case, pick T to shift $\lfloor \min(x,y) \rfloor$ to the origin, using minimum and $\lfloor \cdot \rfloor$ coordinatewise.

Example 3.2.30 *If $\Omega = [-\pi,\pi]^d$, if the space $\mathcal{G}$ consists of d-variate 2π-periodic functions, and if the bilinear form is invariant under coordinatewise real-valued shifts, then we are in a fully periodic setting and $\Phi(x,y)$ has the form (3.2.26, PhiDiff) with a 2π-periodic first argument.*

3.2.5 Remarks

The monograph [4](*atteia:92-1*) also explores the relation between reproducing kernel Hilbert spaces and associated recovery problems. This section used parts of [42](*schaback:96-1*).

3.3 Native Spaces

(*SecNS*) The previous sections have shown that each Hilbert space setting of a recovery problem leads to a specific conditionally positive definite function acting as a reproducing kernel. We now turn this upside down: for each conditionally positive definite function Φ there is a Hilbert space (called the **native space**) with reproducing kernel Φ, and we need as much information as possible about this space. The construction of such a space seems to be a quite academic question, but it isn't. The main reason is that it is much more easy to construct useful conditionally positive definite functions than to find certain Hilbert spaces. Thus it often happens that one starts with a conditionally positive definite function, not with a Hilbert space. But it is necessary to know the Hilbert space in order to assess the optimality properties of the reconstruction process, and thus we cannot ignore the construction of the native space.

Furthermore, if a conditionally positive definite function Φ is constructed without any relation to a Hilbert space, the latter can be theoretically defined and nicely used to investigate the recovery quality of Φ. And there is a third reason: no matter how we arrived at some conditionally positive definite function Φ, we might want to change it somehow, e.g.: by scaling into $\Phi_\delta(\cdot) = \Phi(\cdot/\delta)$. Then we have to calculate the native space for Φ_δ from scratch in order to compare it to the native space for Φ.

3.3.1 From Conditionally Positive Definite Functions to Hilbert Spaces

Now let Φ be a conditionally positive definite function on some domain Ω with respect to some finite-dimensional space $\mathcal{P}$ in the sense of Definition

3.2.15 (*DCPD2)* on page 46. We have to construct the space $\mathcal{G}$ occurring the preceding sections, and its associated bilinear form with nullspace $\mathcal{P}$. Since there is no other tool available than the definition of conditionally positive definite functions, we first have to work with finitely supported functionals

(*Deflxma*)

$$\lambda_{X,M,\alpha} \; : \; f \mapsto \sum_{j=1}^{M} \alpha_j f(x_j) \tag{3.3.1}$$

for arbitrary subsets $X = \{x_1, \ldots, x_M\} \subset \Omega$ of M distinct points, where the coefficient vector $\alpha \in I\!R^M$ satisfies (3.2.16, *CPDef2*), i.e. the above functional is zero on the space $\mathcal{P}$. We thus define $\mathcal{P}_\Omega^-$ to be the set containing all of these functionals. To turn $\mathcal{P}_\Omega^-$ into a vector space over $I\!R$, we use the obvious multiplication by scalars and define the sum of $\lambda_{X,M,\alpha}$ and $\lambda_{Y,N,\beta}$ as $\lambda_{Z,L,\gamma}$ with $Z = \{z_1, \ldots, z_L\}$ and

$$\begin{array}{rcll} Z & = & X \cup Y & \\ L & = & \text{card}\,(Z) & \\ \gamma_\ell & = & \alpha_j & \text{if} \quad z_\ell = x_j \in X \setminus (X \cap Y) \\ \gamma_\ell & = & \beta_k & \text{if} \quad z_\ell = y_k \in Y \setminus (X \cap Y) \\ \gamma_\ell & = & \alpha_j + \beta_k & \text{if} \quad z_\ell = x_j = y_k \in X \cap Y. \end{array}$$

This definition makes sure that

$$\lambda_{X,M,\alpha}(f) + \lambda_{Y,N,\beta}(f) = \lambda_{Z,L,\gamma}(f)$$

holds for each function f on Ω, and thus the sum satisfies (3.2.16, *CPDef2*). The usual laws for vector spaces are satisfied, and we now define a bilinear form on $\mathcal{P}_\Omega^-$ by

(*DefBil*)

$$(\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta})_\Phi := \sum_{j=1}^{M} \sum_{k=1}^{N} \alpha_j \beta_k \Phi(x_j, y_k). \tag{3.3.2}$$

Since Φ is positive definite with respect to $\mathcal{P}$, we even have positive definiteness of the bilinear form on $\mathcal{P}_\Omega^-$, and $\mathcal{P}_\Omega^-$ is a pre-Hilbert space with the inner product $(\cdot, \cdot)_\Phi$ introduced by Φ. Note that the vector space $\mathcal{P}_\Omega^-$ is only dependent on Ω and $\mathcal{P}$, not on Φ itself, but the inner product on $\mathcal{P}_\Omega^-$ depends on Φ, as we indicate by our notation.

We now can define the **native space** $\mathcal{G}$ with respect to Φ to consist of all functions on Ω on which all functionals from $\mathcal{P}_\Omega^-$ are continuous:

(*calgdef*)

$$\mathcal{G} := \left\{ f \ : \ \Omega \to I\!R, \ |\lambda(f)| \leq C_f \|\lambda\|_\Phi \text{ for all } \lambda \in \mathcal{P}_\Omega^- \right\}. \tag{3.3.3}$$

It is immediately clear that $\mathcal{P}$ is a subset of $\mathcal{G}$, but it is neither clear nor true (in general) that the functions $\Phi(x, \cdot)$ are in $\mathcal{G}$. Furthermore, we still need a bilinear form on $\mathcal{G}$ that has $\mathcal{P}$ as its nullspace. To do this, we first define the map

(*Fmapdef*)

$$F \ : \ \mathcal{P}_\Omega^- \to \mathcal{G}, \ F(\lambda_{X,M,\alpha}) = \sum_{j=1}^{M} \alpha_j \Phi(x_j, \cdot) \tag{3.3.4}$$

and have to make sure that the image is indeed in $\mathcal{G}$. But this follows from the very important identity

(*lmF*)

$$\lambda_{Y,N,\beta}(F(\lambda_{X,M,\alpha})) = (\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta})_\Phi = \lambda_{X,M,\alpha}(F(\lambda_{Y,N,\beta})) \tag{3.3.5}$$

for all $\lambda_{X,M,\alpha}, \ \lambda_{Y,N,\beta} \in \mathcal{P}_\Omega^-$. Then we define $\mathcal{F}_0 := F(\mathcal{P}_\Omega^-)$ and assert

Lemma 3.3.6 *The sum $\mathcal{P} + \mathcal{F}_0$ is direct, and the map F is bijective.*

Proof: Indeed if $F(\lambda) = p \in \mathcal{P}$, then for all $\mu \in \mathcal{P}_\Omega^-$ we have $\mu(F(\lambda)) = \mu(p) = (\lambda, \mu)_\Phi = 0$ due to (3.3.5, *lmF*), proving both assertions at the same time. □

In the above proof we used shorthand notation for functionals in $\mathcal{P}_\Omega^-$, and we shall only return to the full notation if absolutely necessary.

We now can define an inner product on $\mathcal{F}_0$ via F, turning F into an isometry and $\mathcal{F}_0$ into a pre-Hilbert space:

$$(F(\lambda), F(\mu))_\Phi := (\lambda, \mu)_\Phi$$

for all $\lambda, \mu \in \mathcal{P}_\Omega^-$. We used the same notation for the inner product, since there will be no confusion between spaces of functions and functionals, respectively.

The next step is to go over to Hilbert space completions of $\mathcal{P}_\Omega^-$ and $\mathcal{F}_0$ in the sense of Theorem 12.2.11 (*HSCT*). Then we get a continuous extension of the isometry F to the completions for free, and we denote this map again by F. The completion of $\mathcal{F}_0$ will be denoted by $\mathcal{F}$, and our final goal is to prove the validity of a direct sum like

(*GPF2*)

$$\mathcal{G} = \mathcal{P} + \mathcal{F} \tag{3.3.7}$$

to recover (3.1.8, *GPF1*) on page 32. But this is a hard task since we do not know that the elements of the completion $\mathcal{F}$ of $F(\mathcal{P}_\Omega^-)$ are functions on Ω at all, let alone that they lie in $\mathcal{G}$. However, we know that an abstract element f of $\mathcal{F}$ allows the action of all functionals $\lambda_{X,M,\alpha} \in \mathcal{P}_\Omega^-$, since (3.3.5, *lmF*) yields

(*lfgeneral*)

$$\lambda_{X,M,\alpha}(f) = (\lambda_{X,M,\alpha}, F^{-1}(f))_\Phi. \tag{3.3.8}$$

This immediately implies a proper definition of function values for f in case of $\mathcal{P} = \{0\}$, since we can define

(*lfsimple*)

$$f(x) := \lambda_{\{x\},1,1}(f) \tag{3.3.9}$$

for all $x \in \Omega$. This definition is consistent with what we know for functions in $\mathcal{F}_0$, and we could proceed to prove (3.3.7, *GPF2*). But we need a little detour for the case $\mathcal{P} \neq \{0\}$, since the above point evaluation functionals are not in $\mathcal{P}_\Omega^-$. To facilitate this, we again require a projector $\Pi_\mathcal{P}$ onto $\mathcal{P}$ as in section 3.1.2 (*SecHSP*) on page 30. We could copy this definition, but since we are in a space of functions now, we want to give a specific construction that can be expressed in terms of function values.

To get such a special projector, we shall assume the existence of a subset

$$\Xi = \{\xi_1, \ldots, \xi_r\} \subseteq \Omega$$

which is nondegenerate with respect to $\mathcal{P}$ and assume without loss of generality that Ξ has a minimal number r of distinct points. Then there is a standard argument from linear algebra that allows to conclude that r equals the dimension of $\mathcal{P}$. In fact, the map

$$p \mapsto (p(\xi_1), \ldots, p(\xi_r))^T \in I\!R^r$$

is injective and we have $q := \dim\ \mathcal{P} \le r$. If $p_1, \ldots, p_q$ form a basis of $\mathcal{P}$, we can write down the injective $r \times q$ matrix

(*Prq*)

$$P := (p_k(\xi_j))_{1 \le j \le r,\, 1 \le k \le q} \tag{3.3.10}$$

and pick a subset of rows that generate a submatrix of maximal row rank. If this were a proper subset, we could reduce r by going over to a subset of Ξ. Thus P has maximal row rank r. But then we must have $q = r$, because there cannot be r linearly independent vectors in a space of dimension $q < r$.

This shows that we can assume $r = q = \dim \mathcal{P}$ and nonsingularity of the $q \times q$ matrix P of (3.3.10, *Prq*). We use this to go over to a Lagrange-type basis of $\mathcal{P}$ with respect to Ξ which we again denote by $p_1, \ldots, p_q$. Then P is the identity matrix and we can write every function $p \in \mathcal{P}$ as

(*PRq2*)

$$p(\cdot) = \sum_{j=1}^{q} p(\xi_j) p_j(\cdot). \tag{3.3.11}$$

This now yields the explicit form of a projector $\Pi_{\mathcal{P}}$ onto $\mathcal{P}$ as

$$\Pi_{\mathcal{P}}(f)(\cdot) := \sum_{j=1}^{q} f(\xi_j) p_j(\cdot)$$

for all functions that are at least defined on Ξ. The projector has the additional property

$$(f - \Pi_{\mathcal{P}} f)(\Xi) = \{0\}$$

for all functions f that are defined on Ξ, because of $\delta_{\xi_j,\Xi} = 0,\ 1 \le j \le q$. Note that $\pi_j(f) = f(\xi_j)$ holds if we compare (3.3.11, *PRq2*) with (3.1.5, *DefPN*).

So the projector is well-defined, but we cannot use it right away, since we first need nice functionals in $\mathcal{P}_\Omega^\perp$. But such functionals come from the projector via

(*deltagen*)

$$\delta_{x,\Xi}(f) := f(x) - (\Pi_{\mathcal{P}}(f))(x) = f(x) - \sum_{j=1}^{q} f(\xi_j) p_j(x) \tag{3.3.12}$$

for all $x \in \Omega$ and they annihilate $\mathcal{P}$, as required.

The notation $\delta_{x,\mathcal{P}}$ from (3.2.8, *deltadef2*) is very similar, but there will be no possible confusion. Similar variations of point evaluation functionals will occur later. These functionals are useful to prove an intermediate result that will be of some use later:

Lemma 3.3.13 *(SuffPol) If the action of all functionals λ from $\mathcal{P}_\Omega^\perp$ is zero on a given function f from $\mathcal{G}$, then f coincides with a function from $\mathcal{P}$ on Ω.*

Proof. : Just take the functionals $\delta_{x,\Xi}$ for all $x \in \Omega$, and look at

$$0 = \delta_{x,\Xi}(f) = f(x) - (\Pi_{\mathcal{P}}(f))(x).$$

□

We now could generalize (3.3.9, *lfsimple*) using the above functionals in (3.3.8, *lfgeneral*):

(*lf2*)

$$f(x) := (\delta_{x,\Xi}, F^{-1}(f))_\Phi,\ x \in \Omega,\ f \in \mathcal{F}. \tag{3.3.14}$$

This assigns specific function values to the abstract element of the closure $\mathcal{F}$ of $\mathcal{F}_0$. The assignment has the consequence that $f(\Xi) = \{0\}$ due to $\delta_{\xi_j,\Xi} = 0,\ 1 \le j \le q$, and thus it is rather an assignment of values to $f - \Pi_{\mathcal{P}} f$ than to f itself. We thus avoid this complication and define a mapping

$$R_{\mathcal{P}}\ :\ \mathcal{F} \to \mathcal{G}$$

by

(*lf3*)

$$(R_{\mathcal{P}} f)(x) := (\delta_{x,\Xi}, F^{-1}(f))_\Phi,\ x \in \Omega,\ f \in \mathcal{F}. \tag{3.3.15}$$

We have to show that this maps into $\mathcal{G}$, and for this we have to evaluate

$$\begin{aligned} \lambda_{X,M,\alpha}(R_{\mathcal{P}}(f)) &= \sum_{j=1}^{M} \alpha_j (\delta_{x_j,\Xi}, F^{-1}(f))_\Phi \\ &= \left(\sum_{j=1}^{M} \alpha_j \delta_{x_j,\Xi}, F^{-1}(f) \right)_\Phi . \end{aligned}$$

Now the functional in the bilinear form boils down to

$$\begin{aligned} \sum_{j=1}^{M} \alpha_j \delta_{x_j,\Xi}(f) &= \sum_{j=1}^{M} \alpha_j \left(f(x_j) - \sum_{k=1}^{q} f(\xi_k) p_k(x_j) \right) \\ &= \sum_{j=1}^{M} \alpha_j f(x_j) - \sum_{k=1}^{q} f(\xi_k) \sum_{j=1}^{M} \alpha_j p_k(x_j) \\ &= \sum_{j=1}^{M} \alpha_j f(x_j) - 0 \\ &= \lambda_{X,M,\alpha}(f), \end{aligned}$$

and we end up with

(*RfDef*)

$$\lambda_{X,M,\alpha}(R_{\mathcal{P}}(f)) = (\lambda_{X,M,\alpha}, F^{-1}(f))_\Phi \tag{3.3.16}$$

which proves $R_{\mathcal{P}}(f) \in \mathcal{G}$.

Theorem 3.3.17 *(GPFT2) The spaces $\mathcal{P}$, $\mathcal{G}$, and $\mathcal{F}$ of functions on Ω form a direct sum*

$$\mathcal{G} = \mathcal{P} + R_{\mathcal{P}}(\mathcal{F}),$$

and $R_{\mathcal{P}}$ defined by (3.3.15, lf3) is an isometry between $\mathcal{F}$ and $R_{\mathcal{P}}(\mathcal{F}) \subseteq \mathcal{G}$. The inner products on $\mathcal{F}$ and $R_{\mathcal{P}}(\mathcal{F})$ introduce a bilinear form

$$(g,h)_{\mathcal{G}} := (R_{\mathcal{P}}^{-1}(g - \Pi_{\mathcal{P}} g), R_{\mathcal{P}}^{-1}(h - \Pi_{\mathcal{P}} h))_{\Phi}$$

with nullspace $\mathcal{P}$ on $\mathcal{G}$.

Proof: The intersection of $\mathcal{P}$ and $R_{\mathcal{P}}(\mathcal{F})$ is zero, because the second space consists of functions vanishing on Ξ, and the only such function in the first space is the zero function. Thus the sum is direct, and we have to show that the sum fills all of $\mathcal{G}$. Before we do that, we take a look at the mapping $R_{\mathcal{P}}$ and check the topology of $\mathcal{G}$. Each function f in $\mathcal{G}$ has the well-defined norm

$$\|f\|_{\mathcal{G}} := \sup_{\lambda \in \mathcal{P}_{\Omega}^{\perp} \setminus \{0\}} \frac{|\lambda(f)|}{\|\lambda\|_{\Phi}},$$

and the identity (3.3.16, *RfDef*) immediately yields

$$\|R_{\mathcal{P}}(f)\|_{\mathcal{G}} = \|F^{-1}(f)\|_{\Phi} = \|f\|_{\Phi}$$

for all $f \in \mathcal{F}$. Thus $R_{\mathcal{P}}$ is isometric, and $R_{\mathcal{P}}(\mathcal{F})$ is the closure of $R_{\mathcal{P}}(\mathcal{F}_0)$ in $\mathcal{G}$.

We now proceed to show that $\mathcal{P} + R_{\mathcal{P}}(\mathcal{F})$ fills all of $\mathcal{G}$, and we shall construct the inverse of $R_{\mathcal{P}}$. Take an arbitrary function $f \in \mathcal{G}$ and define a functional L_f on the space $\mathcal{P}_{\Omega}^{-}$ by

$$L_f(\lambda) := \lambda(f),\ \lambda \in \mathcal{P}_{\Omega}^{-}.$$

This functional is continuous on $\mathcal{P}_{\Omega}^{-}$ because f is in $\mathcal{G}$, and it has a continuous extension to the closure of $\mathcal{P}_{\Omega}^{-}$ which is a space isomorphic to the Hilbert space $\mathcal{F}$. We thus invoke the Riesz representation theorem 12.2.14 (*RieszT*) to get an element $S(f) \in \mathcal{F}$ with

$$L_f(\lambda) = \lambda(f) = (\lambda, F^{-1}(S(f)))_{\Phi} = (F(\lambda), S(f))_{\Phi} \text{ for all } \lambda \in \mathcal{P}_{\Omega}^{-}.$$

Using (3.3.16, *RfDef*), this turns into

$$\lambda_{X,M,\alpha}(R_{\mathcal{P}} S(f)) = (\lambda_{X,M,\alpha}, F^{-1} S f)_{\Phi} = \lambda_{X,M,\alpha}(f)$$

and Lemma 3.3.13 (*SuffPol)* implies that $f - R_{\mathcal{P}}S(f)$ coincides with a function from $\mathcal{P}$ on Ω, and since $\Pi_{\mathcal{P}}R_{\mathcal{P}}$ is the zero mapping, we see that

$$f = \Pi_{\mathcal{P}}f + R_{\mathcal{P}}Sf$$

holds for all f in $\mathcal{G}$, proving that the direct sum fills all of $\mathcal{G}$. The statement on the bilinear form is straightforward to prove. □

To write down a more explicit representation of the functions from $\mathcal{G}$, we apply F to $\delta_{x,\Xi}$ and get

$$F(\delta_{x,\Xi})(\cdot) = \Phi(x,\cdot) - \sum_{k=1}^{q} \Phi(\xi_k,\cdot)p_k(x) \in \mathcal{G}.$$

Then (3.3.15, *lf3*) and Theorem 3.3.17 (*GPFT2)* imply the representation
(*Taylor2*)

$$f(x) \quad = \quad \sum_{j=1}^{q} f(\xi_j)p_j(x) + \left(\Phi(x,\cdot) - \sum_{j=1}^{q}\Phi(\xi_j,\cdot)p_j(x), Sf(\cdot)\right)_{\mathcal{G}}, \qquad (3.3.18)$$

but note that the sum in the first argument of the bilinear form cannot easily be taken out, because $\Phi(x,\cdot)$ may not be in $\mathcal{G}$. The same problem prevents us from concluding that Φ serves as a reproducing kernel in the strong sense of Theorem 3.2.12 (*CPDFT1)*. A good candidate, however, is the readily available function
(*EqPsiDef*)

$$\begin{aligned} \Psi(x,y) &:= (\delta_{x,\Xi}, \delta_{y,\Xi})_{\Phi} \\ &= (F(\delta_{x,\Xi}), F(\delta_{y,\Xi}))_{\Phi}, \end{aligned} \qquad (3.3.19)$$

because then (3.3.18, *Taylor2*) yields

$$\begin{aligned} F(\delta_{y,\Xi})(x) &= (\Pi_{\mathcal{P}}F(\delta_{y,\Xi}))(x) + (F(\delta_{x,\Xi}), F(\delta_{y,\Xi}))_{\Phi}\,, \\ &= (\Pi_{\mathcal{P}}F(\delta_{y,\Xi}))(x) + \Psi(x,y) \end{aligned}$$

such that

$$\Psi(x,y) = F(\delta_{y,\Xi})(x) - (\Pi_{\mathcal{P}}F(\delta_{y,\Xi}))(x)$$

holds, proving that $\Psi(\cdot,x)$ is indeed in $\mathcal{G}$ and satisfies $\Pi_{\mathcal{P}}(\Psi(\cdot,y)) = 0$ for all $y \in \Omega$. The above identity can now be put into (3.3.18, *Taylor2*) to get (3.2.10, *Taylor*) via

(*Taylor4*)

$$
\begin{aligned}
f(x) &= (\Pi_{\mathcal{P}} f)(x) + (\Psi(x,\cdot) + \Pi_{\mathcal{P}} F(\delta_{x,\Xi}))(\cdot), f(\cdot))_{\Phi}, \\
&= (\Pi_{\mathcal{P}} f)(x) + (\Psi(x,\cdot), f(\cdot))_{\Phi}.
\end{aligned} \tag{3.3.20}
$$

Thus the function Ψ satisfies all we need for Theorem 3.2.12 (*CPDFT1)*, but we still have to look at its relation to the original function Φ:

(*DefSymmPsi*)

$$
\begin{aligned}
\Psi(x,y) &= F(\delta_{x,\Xi})(y) - (\Pi_{\mathcal{P}} F(\delta_{x,\Xi}))(y) \\
&= \Phi(x,y) - \sum_{j=1}^{q} \Phi(\xi_j,y) p_j(x) - (\Pi_{\mathcal{P}}(\Phi(x,\cdot) - \sum_{j=1}^{q} \Phi(\xi_j,\cdot) p_j(x)))(y) \\
&= \Phi(x,y) - \sum_{j=1}^{q} \Phi(\xi_j,y) p_j(x) - \sum_{k=1}^{q} \Phi(x,\xi_k) p_k(y) + \sum_{j,k=1}^{q} \Phi(\xi_k,\xi_j) p_k(x) p_j(y).
\end{aligned} \tag{3.3.21}
$$

Inspection of this equation and comparison with (3.3.2, *DefBil*) implies that Φ and Ψ generate the same bilinear form for the definition of the native space. Thus Ψ is also conditionally positive definite and the native spaces generated by Φ and Ψ coincide.

From (3.3.20, *Taylor4*) one can deduce that all functions from $\mathcal{G}$ are continuous on Ω, provided that Φ and the functions in $\mathcal{P}$ are continuous. In fact, we get

$$
\begin{aligned}
|f(x) - f(y)| &= |(\Pi_{\mathcal{P}} f)(x) - (\Pi_{\mathcal{P}} f)(y) + (\Psi(x,\cdot) - \Psi(y,\cdot), f(\cdot))_{\Phi}| \\
&\le |(\Pi_{\mathcal{P}} f)(x) - (\Pi_{\mathcal{P}} f)(y)| + |(\Psi(x,\cdot) - \Psi(y,\cdot), f(\cdot))_{\Phi}| \\
&\le \sum_{j=1}^{Q} |f(\xi_j)| p_j(x) - p_j(y)| + \|\Psi(x,\cdot) - \Psi(y,\cdot)\|_{\Phi} \|f(\cdot)\|_{\Phi}
\end{aligned}
$$

and we can expand $\|\Psi(x,\cdot) - \Psi(y,\cdot)\|_{\Phi}^2$ as

$$\|\Psi(x,\cdot) - \Psi(y,\cdot)\|_{\Phi}^2 = \Psi(x,x) - 2\Psi(x,y) + \Psi(y,y).$$

Since all quantities now are continuous for $y \to x$, we are finished.

We can now add up the results of this section:

Theorem 3.3.22 *(CPDSuffT) Let Φ be a conditionally positive definite function on some domain Ω with a finite-dimensional nullspace $\mathcal{P}$ of functions on Ω that allows an interpolatory projector*

$$(\Pi_{\mathcal{P}} f)(\cdot) = \sum_{j=1}^{q} f(\xi_j) p_j(\cdot)$$

where $p_1, \ldots, p_q$ are a basis of $\mathcal{P}$ and $\xi_1, \ldots, \xi_q$ form a $\mathcal{P}$-nondegenerate subset Ξ of Ω. Then there is a **native space** *$\mathcal{G}$ for Φ carrying a bilinear form with nullspace $\mathcal{P}$, and having the function Ψ as defined in (3.3.21, DefSymmPsi) as a reproducing kernel in the sense of Theorem 3.2.17 (CPDNeccT). The native space is formed by adding a Hilbert space to $\mathcal{P}$. The functions in the native space are continuous if Φ and the functions in $\mathcal{P}$ are continuous.* □

The transition from a conditionally positive definite function Φ to the function Ψ with (3.3.21, *DefSymmPsi*) will be called **normalization** in the sequel. We note that the normalized function Ψ can also be defined if the projector is not interpolatory, but rather of the more general form (3.1.5, *DefPN*).

3.3.2 Normalization of conditionally positive definite functions

(*PhiNormalization*) With the notation of the preceding section it is fairly easy to describe the reduction of a conditionally positive definite function to an unconditionally positive definite function. This process coincides with the normalization by (3.3.21, *DefSymmPsi*).

Theorem 3.3.23 *(RedCPDFT) Let Φ be a conditionally positive definite function with respect to the nullspace $\mathcal{P}$ of the bilinear form on $\mathcal{G}$, and let the projector $\Pi_{\mathcal{P}}$ onto $\mathcal{P}$ be interpolatory with a minimal $\mathcal{P}$-nondegenerate set $\Xi = \{\xi_1, \ldots, \xi_q\}$ of points of Ω. Then the normalized function Ψ defined as in (3.3.21, DefSymmPsi) is unconditionally positive definite on $\Omega \setminus \Xi$.*

Proof: Consider a finite subset $X = \{x_1, \ldots, x_M\}$ of $\Omega \setminus \Xi$ and an arbitrary coefficient vector $\alpha \in I\!R^M$. Then the functional

$$\begin{aligned}
\left(\sum_{j=1}^{M} \alpha_j \delta_{x_j,\Xi}\right)(f) &= \sum_{j=1}^{M} \alpha_j f(x_j) - \sum_{j=1}^{M} \alpha_j \left(\sum_{k=1}^{q} f(\xi_k) p_k(x_j)\right) \\
&= \sum_{j=1}^{M} \alpha_j f(x_j) - \sum_{k=1}^{q} f(\xi_k) \left(\sum_{j=1}^{M} \alpha_j p_k(x_j)\right)
\end{aligned}$$

necessarily vanishes on $\mathcal{P}$ and is in $\mathcal{P}_\Omega^-$. Applying the conditional positive definiteness of Φ for this functional yields positivity of

$$\alpha^T A_{X,\Psi} \alpha$$

unless the coefficients of the above functional are zero, which implies that α is zero. □

By some simple linear algebra techniques the above normalization method can be shown to be equivalent to the method described in 10.2 (*Red2)* on page 173. To see this we give some hints, but suppress details of the full argument. Starting in section 10.2 (*Red2)* with a $\mathcal{P} = I\!P_m^d$-nondegenerate set $X = \{x_1, \ldots, x_M\}$, we can renumber the points and assume that $\Xi = \{x_1 \ldots, x_q\} = \{\xi_1 \ldots, \xi_q\}$ holds. Furthermore, if we pick the right basis in $I\!P_m^d$, the matrix S in (10.2.1, *Dec2*) has the elements $p_j(x_k),\ k = q+1, \ldots, M$. But then the matrix occurring in (10.2.4, *RedSys3*) precisely describes how to form the elements $\Psi(x_j, x_k)$ for $j, k = q+1, \ldots, M$ via the normalization formula (3.3.21, *DefSymmPsi*).

The function Ψ vanishes whenever one of its arguments is in Ξ. This is reflected in the above argument, since Ψ is responsible for reconstruction on $X \setminus \Xi = \{x_{q+1}, \ldots, x_M\}$.

3.3.3 Characterization of Native Spaces

(*SecCNS*) The native space associated to each conditionally positive definite function Φ is a rather abstract object, and it would be nice to know precisely which functions are in the space and which are not. This is a nontrivial task, since the only available information to start with is the conditional positive definiteness of Φ. Using transforms, we can give some results in section 6.1 (*SecCNST)*. But there are some simple things that we can do right now.

We first want to know how smooth the functions in native spaces are. Since we have the representation (3.3.20, *Taylor4*) that allows any $\lambda \in \mathcal{P}_\Omega^-$ to be applied to f with result $(\lambda^x \Psi(x, \cdot), f(\cdot))_\Phi$, we check the functionals in $\mathcal{P}_\Omega^-$. Assume that λ is a linear functional that

1. we can safely and independently apply to both arguments of Φ (or Ψ, for convenience),
2. that vanishes on $\mathcal{P}$, and
3. can be approximated by functionals from $\mathcal{P}_\Omega^-$.

Then we assert that $\lambda \in \mathcal{F}^*$. More precisely, we have

Theorem 3.3.24 *Assume a general linear functional λ and a sequence $\{\lambda_n\}_n \subset \mathcal{P}_\Omega^-$ to satisfy*

$$\begin{array}{rcl} \lambda^x \lambda^y \Phi(x, y) & & \textit{exists} \\ \lim\limits_{n\to\infty} \lambda_n^x \mu^y \Phi(x, y) & = & \lambda^x \mu^y \Phi(x, y) \textit{ for all } \mu \in \mathcal{P}_\Omega^- \\ \lim\limits_{n\to\infty} \lambda_n^x \lambda_n^y \Phi(x, y) & = & \lambda^x \lambda^y \Phi(x, y) \\ \lambda(\mathcal{P}) & = & \{0\}. \end{array}$$

Then λ acts on functions in $\mathcal{G}$ like a functional in $\mathcal{F}^*$, the closure of $\mathcal{P}_\Omega^-$.

Proof: We first note that $\{\|\lambda_n\|_\Phi\}_n$ is a Cauchy sequence, because it converges in $I\!R$. Then we use the standard technique of the proof of Theorem 12.2.11 (*HSCT*) to conclude that $\{\lambda_n\}_n$ is a Cauchy sequence in $\mathcal{P}_\Omega^-$. Thus it has a limit $\mu \in \mathcal{F}^*$, and we can prove

$$(\lambda - \mu)^x \delta_{y,\Xi}^z \Phi(x,z) = \lim_{n\to\infty} \lambda_n^x \delta_{y,\Xi}^z \Phi(x,z) - \lim_{n\to\infty} \lambda_n^x \delta_{y,\Xi}^z \Phi(x,z) = 0$$

for all $y \in \Omega$. This implies $\lambda^x \Phi(x,y) = \mu^x \Phi(x,y) + p(y)$ with some $p \in \mathcal{P}$. Using an arbitrary $\rho \in \mathcal{P}_\Omega^-$ we get $\lambda^x \rho^y \Phi(x,y) = \mu^x \rho^y \Phi(x,y)$. Thus λ and μ generate the same functional on $\mathcal{P}_\Omega^-$ and can be identified as functionals in the closure $\mathcal{F}^*$. $\square$

This coarse result can be applied to functionals that are point-evaluation functionals of derivatives, and which are approximated by finite difference functionals. It shows that Gaussians and multiquadrics generate native spaces of infinitely differentiable functions, while non-smooth conditionally positive definite functions Φ generate spaces of roughly half the smoothness of Φ. But note that the above approach does not cover smoothness of derivatives, just their pointwise existence.

3.3.4 Remarks

The association of a Hilbert space to each conditionally positive definite function dates back to Madych and Nelson ([20](*madych-nelson:83-1*) [21](*madych-nelson:88-1*) [22](*madych-nelson:89-1*) [23](*madych-nelson:90-1*)).

3.4 Standardized Notation

(*SecSN*) The previous sections showed that it does not matter whether we start our theory from optimal recovery in spaces of functions with a bilinear form or from any given conditionally positive definite function. The only difference was that in the first case we constructed a normalized conditionally positive definite function from the given bilinear form, while in the second the given conditionally positive definite function Φ may not be normalized, though its normalization Ψ will generate the same bilinear form as Φ. From now on we want to be independent from the starting point, and thus we collect the following facts that hold in both cases:

1. $\Phi : \Omega \times \Omega \to I\!R$ is a conditionally positive definite function on some domain Ω with respect to some nullspace $\mathcal{P}$ of finite dimension q.

2. There is a positive semidefinite bilinear form $(\cdot,\cdot)_\Phi$ on a space $\mathcal{G}$ of functions on Ω with nullspace $\mathcal{P}$.

3. The nullspace $\mathcal{P}$ has a basis $p_1,\dots,p_q$ such that with certain linear functionals $\pi_1,\dots,\pi_q$ on $\mathcal{G}$ the projector $\Pi_\mathcal{P}$ from $\mathcal{G}$ onto $\mathcal{P}$ is well-defined via
$$\Pi_\mathcal{P} g = \sum_{k=1}^{q} \pi_k(g) p_k \text{ for all } g \in \mathcal{G}.$$

4. For each $x \in \Omega$ the linear functionals
$$\delta_{x,\mathcal{P}} \ : \ g \mapsto g(x) - (\Pi_\mathcal{P} g)(x)$$
are continuous with respect to $(\cdot,\cdot)_\Phi$ and the Taylor-type reconstruction formula
$$\delta_{x,\mathcal{P}}(g) = (\delta^y_{x,\mathcal{P}}\Phi(y,\cdot), g(\cdot))_\Phi$$
holds for all $g \in \mathcal{G}$, $x \in \Omega$.

5. The space $\mathcal{G}$ can be decomposed into a direct sum $\mathcal{G} = \mathcal{P} + \mathcal{F}$ such that $\mathcal{F}$ is a Hilbert space with inner product $(\cdot,\cdot)_\Phi$.

6. If functionals $\lambda_{X,M,\alpha}$ are defined as
$$\lambda_{X,M,\alpha} \ : \ f \mapsto \sum_{j=1}^{M} \alpha_j f(x_j)$$
for sets $X = \{x_1,\dots,x_M\} \subset \Omega$ and vectors $\alpha \in I\!\!R^M$ for arbitrary values of $M \geq q$, then one can define an inner product
$$(\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta})_\Phi := \sum_{j=1}^{M}\sum_{k=1}^{N} \alpha_j\beta_k\Phi(x_j,y_k)$$
on all such functionals that vanish on $\mathcal{P}$. The set $\mathcal{P}_\Phi^-$ of all of these functionals then is an inner product space.

7. The space $\mathcal{G}$ is the largest space of functions on Ω such that all functionals in $\mathcal{P}_\Phi^-$ are continuous with respect to the norm induced by $(\cdot,\cdot)_\Phi$ on $\mathcal{P}_\Phi^-$.

8. The closure of $\mathcal{P}_\Phi^-$ under $(\cdot,\cdot)_\Phi$ is the dual $\mathcal{F}^*$ of $\mathcal{F}$, and the map $F \ : \ \mathcal{F}^* \to \mathcal{F}$ provides the Riesz correspondence between functionals and functions.

9. The action of F is related to Φ via
$$F(\lambda)(\cdot) = \lambda^y \Phi(y, \cdot)$$
for all $\lambda \in \mathcal{F}^*$. This is evident in case of $\lambda = \lambda_{X,M,\alpha} \in \mathcal{P}_\Phi^-$ and has to be read as a definition of the right-hand side for general λ.

10. The dual $\mathcal{G}^*$ of $\mathcal{G}$ consists of functionals that are sums of a linear functional on $\mathcal{P}$ and a linear functional in $\mathcal{F}^*$.

11. For each $\lambda \in \mathcal{G}^*$ we have $\lambda - \lambda\Pi_{\mathcal{P}} \in \mathcal{F}^*$ and
$$\begin{array}{rcl} \lambda(g) & = & \lambda\Pi_{\mathcal{P}}g + (F(\lambda - \lambda\Pi_{\mathcal{P}}), g)_\Phi \\ & = & \lambda\Pi_{\mathcal{P}}g + (\lambda - \lambda\Pi_{\mathcal{P}}, F^{-1}(g - \Pi_{\mathcal{P}}g))_\Phi \\ & = & \lambda\Pi_{\mathcal{P}}g + ((\lambda - \lambda\Pi_{\mathcal{P}})^y\Phi(y, \cdot), g)_\Phi \end{array}$$
for all $g \in \mathcal{G}$.

12. The normalization $\Psi_{\mathcal{P}}$ of Φ is defined via
$$\Psi_{\mathcal{P}}(x, y) = (\delta_{x,\mathcal{P}}, \delta_{y,\mathcal{P}})_\Phi = \delta^u_{x,\mathcal{P}}\delta^v_{y,\mathcal{P}}\Phi(u, v)$$
for all $x, y \in \Omega$. It has the properties described in Theorem 3.2.12 (*CPDFT1)*.

13. The unique solution g^* of the optimal recovery problem for data $\lambda_j(g)$ with $g \in \mathcal{G}$ and $\lambda_j \in \mathcal{G}^*$ represented by functions $g_j \in \mathcal{G}$ is of the form (3.1.13, *grep*) with coefficients satisfying erefEQsys3.

14. The solution is orthogonal to all functions from $\mathcal{G}$ which are a sum of functions from $\mathcal{P}$ with functions v such that $\lambda_j(v) = 0$ for all j $1 \leq j \leq M$.

3.5 Restrictions, Extensions, and Infinite Problems

(*SecREIP*) It will turn out later that the recovery of functions via conditionally positive definite basis functions involves certain “natural” boundary conditions like those of “natural” cubic splines (see 3.6 (*SecCSOneV*)). To study these, we look at the construction of native spaces for recovery of functions on subsets Ω_0 of Ω. Another reason is that in case of a given Hilbert space setting for functions on some large domain Ω, we do not know how the native spaces for smaller domains Ω_0 are related to the given Hilbert space of functions on Ω.

To this end, let us assume that Φ is a conditionally positive definite function on some domain Ω, while we do reconstruction on data given on a subdomain Ω_0. Of course, the reconstructing functions based on finitely many data in Ω_0 can be extended to all of Ω, but it needs proof that *all* functions f from a native space based on Ω_0 have a canonical extension to Ω. Each of these extensions solves a reconstruction problem on Ω with possibly infinitely many data, namely the data of f on the subset Ω_0. Then we study the orthogonal complement of the extensions, and the result leads us to the general treatment of recovery problems with infinitely many data.

3.5.1 Restriction Mapping

Assume that Ω_0 is a subset of Ω. Then we can restrict Φ to Ω_0 and carry out the whole construction of a native space based on Ω_0 instead of Ω. We add the subscripts Ω or Ω_0 in this discussion to distinguish between the construction of native spaces with respect to Ω or Ω_0. The reader should be aware that Ω can be infinite, e.g.: $\Omega = I\!R^d$, while Ω_0 often will be the bounded domain that we actually work on. However, for the definition of the projector $\Pi_{\mathcal{P}}$ and the functionals $\delta_{x,\Xi}$ we use a subset Ξ of $\Omega_0 \subset \Omega$, such that the projector and the functionals are the same for both cases. Unfortunately, earlier writeups of the following arguments contained numerous traps that had to be eliminated later, and thus we take a very formal approach here, starting with the explicit use of a restriction map r_0 that takes functions defined on Ω to functions defined on Ω_0. After normalization, which also is the same in both cases since we use the same functionals, we arrive at two versions of the representation (3.3.20, *Taylor4*), namely

(*Taylor4b*)

$$\begin{array}{rclcll} g(x) & = & (\Pi_{\mathcal{P}} g)(x) & + & (\Psi(x,\cdot), g(\cdot))_{\Phi,\Omega}\,, & x \in \Omega,\ g \in \mathcal{G}_\Omega, \\ f(x) & = & (r_0 \Pi_{\mathcal{P}} f)(x) & + & ((r_0^y(\Psi(x,y)))(\cdot), f(\cdot))_{\Phi,\Omega_0} & x \in \Omega_0,\ f \in \mathcal{G}_{\Omega_0}, \end{array} \tag{3.5.1}$$

which illustrate the use of the restriction map r_0. Note that we cannot identify the two bilinear forms without additional information, since they were generated via continuous extension from bilinear forms acting on objects that depended on the domain. This applies to the bilinear forms acting on functions; the corresponding bilinear forms defined on functionals will just depend on Φ but not on the domain, since they are obtained by continuous extension of (3.3.2, *DefBil*) on page 55.

The normalization Ψ of Φ is defined on all of Ω, and thus the second equation of (3.5.1, *Taylor4b*) can possibly be used for $x \in \Omega \setminus \Omega_0$, too. However, this would require

(*restrPsi*)

$$(r_0^y(\Psi(x,y)))(\cdot) \in \mathcal{G}_{\Omega_0} \tag{3.5.2}$$

for all $x \in \Omega$, which is not trivial.

We thus proceed more carefully and repeat the construction of native spaces to some detail. Consider the general definition of $\mathcal{G}$ in (3.3.3, *calgdef*). This leads to

$$\begin{array}{rcl} \mathcal{G}_\Omega & := & \left\{g \;:\; \Omega \to I\!R,\ |\lambda(g)| \le C_g \|\lambda\|_\Phi \text{ for all } \lambda \in \mathcal{P}_\Omega^-\right\} \\ \mathcal{G}_{\Omega_0} & := & \left\{f \;:\; \Omega_0 \to I\!R,\ |\lambda(f)| \le C_f \|\lambda\|_\Phi \text{ for all } \lambda \in \mathcal{P}_{\Omega\,0}^-\right\}, \end{array}$$

where we note with relief that

$$\|\lambda\|_\Phi^2 = \lambda^x \lambda^y \Phi(x,y)$$

does not depend on the domain. Since each functional $\lambda \in \mathcal{P}_{\Omega\,0}^-$ induces a functional $\lambda \circ r_0$ in $\mathcal{P}_\Omega^-$, we can define an intermediate space

$$\mathcal{G}_0 := \left\{g \;:\; \Omega \to I\!R,\ |\lambda(r_0(g))| \le C_g \|\lambda\|_\Phi \text{ for all } \lambda \in \mathcal{P}_{\Omega\,0}^-\right\}.$$

with $\mathcal{G}_\Omega \subseteq \mathcal{G}_0$. But then the restriction $r_0(g)$ of any function g in $\mathcal{G}_0$ to Ω_0 clearly is in $\mathcal{G}_{\Omega_0}$. We abbreviate this fact by

$$r_0(\mathcal{G}_\Omega) \subseteq r_0(\mathcal{G}_0) \subseteq \mathcal{G}_{\Omega_0}$$

and note for later use that

(*restrnorm*)

$$|r_0(g)|_{\Omega_0} \le |g|_\Omega \tag{3.5.3}$$

holds for all $g \in \mathcal{G}_\Omega$, since the minimal constant C_g that is good for g in the definition of $\mathcal{G}$ will also work for $r_0(g)$ in the definition of $\mathcal{G}_{\Omega_0}$. Altogether, this discussion showed that there are no problems with (3.5.2, *restrPsi*), since $\Psi(x,\cdot)$ is in $\mathcal{G}_\Omega$ for all $x \in \Omega$.

3.5.2 Extension Mapping

We now can extend the interpretation of the second equation in (3.5.1, *Taylor4b*) by defining an extension $e^0(f)$ of a function $f \in \mathcal{G}_{\Omega_0}$ by

(*extendef*)

$$e^0(f)(x) := (\Pi_{\mathcal{P}} f)(x) + ((r_0^y(\Psi(x,y))(\cdot)), f(\cdot))_{\Phi,\Omega_0} \tag{3.5.4}$$

for all $x \in \Omega$. We assert that this function is in $\mathcal{G}_\Omega$. To this end we have to apply an arbitrary functional $\lambda_{X,M,\alpha} \in \mathcal{P}_\Omega^-$ to get

$$\begin{array}{rcl} |\lambda_{X,M,\alpha} e^0(f)| & = & \left| \left((r_0^y(\lambda_{X,M,\alpha}^x \Psi(x,y)))(\cdot), f(\cdot) \right)_{\Phi,\Omega_0} \right| \\ & \leq & |f|_{\Phi,\Omega_0} |r_0^y(\lambda_{X,M,\alpha}^x \Psi(x,y))|_{\Phi,\Omega_0}. \\ & \leq & |f|_{\Phi,\Omega_0} |(\lambda_{X,M,\alpha}^x \Psi(x,\cdot))|_{\Phi,\Omega} \\ & \leq & |f|_{\Phi,\Omega_0} \|\lambda_{X,M,\alpha}\|_\Phi \end{array}$$

due to (3.5.3, *restrnorm*). Since now $e^0(f)$ is in $\mathcal{G}_\Omega$, we can compare (3.5.4, *extendef*) with the application of (3.3.20, *Taylor4*) for both $e^0(f)$ and f:

(*Taylor4a*)

$$\begin{array}{rcrcll} e^0(f)(x) & = & (\Pi_\mathcal{P} e^0(f))(x) & + & (\Psi(x,\cdot), e^0(f)(\cdot))_{\Phi,\Omega}, & x \in \Omega,\ f \in \mathcal{G}_{\Omega_0}, \\ f(x) & = & (r_0 \Pi_\mathcal{P} f)(x) & + & ((r_0^y(\Psi(x,y)))(\cdot), f(\cdot))_{\Phi,\Omega_0} & x \in \Omega_0,\ f \in \mathcal{G}_{\Omega_0}. \end{array} \tag{3.5.5}$$

The second comparison implies $f(x) = e^0(f)(x)$ for all $x \in \Omega_0$, as expected, and the first yields

$$((r_0^y \Psi(x,y))(\cdot), f(\cdot))_{\Phi,\Omega_0} = \left(\Psi(x,\cdot), e^0(f)(\cdot) \right)_{\Phi,\Omega}$$

for all $x \in \Omega$. If we apply arbitrary functionals from $\mathcal{P}_\Omega^-$ with respect to x to this equation and use continuity, we see that

$$(r_0(g), f)_{\Phi,\Omega_0} = \left(g, e^0(f) \right)_{\Phi,\Omega}$$

holds for all $f \in \mathcal{G}_{\Omega_0}$, $g \in \mathcal{G}_\Omega$. But since we already know that $r_0 e^0$ is the identity, the mapping e^0 must be isometric. This easily follows from setting $g = e^0(h)$ in the above equation.

This altogether yields an extension theorem which was first observed in full generality by Iske [19](*iske:94-2*).

Theorem 3.5.6 *(ExtTh1) A function f from a native space on a domain Ω_0 always has a canonical extension $e^0(f)$ to the largest domain Ω on which the generating basis function Φ is conditionally positive definite with respect to some finite-dimensional space $\mathcal{P}$. The extension is furnished by the Taylor-type reproduction formula (3.5.4, extendef). The extension map e^0 is an isometric operator from $\mathcal{G}_{\Omega_0}$ into $\mathcal{G}_\Omega$.* □

3.5.3 Properties of the Extension

We now look somewhat more closely at the extension $e^0(f)$ of a function $f \in \mathcal{P}_{\Omega\,0}^-$.

Theorem 3.5.7 *(ExtTh2) The canonical extension $e^0(f)$ of a function $f \in \mathcal{G}_{\Omega_0}$ to the larger domain Ω is a solution of the optimal recovery problem posed in $\mathcal{G}_\Omega$ for functions g satisfying*

$$\lambda(r_0(g)) = \lambda(f) \text{ for all } \lambda \in \mathcal{P}^{-}_{\Omega\,0}$$

or

$$\delta_{x,\Xi}(r_0(g)) = \delta_{x,\Xi}(f) \text{ for all } x \in \Omega_0.$$

Any two solutions differ by a function in $\mathcal{P}$, and for $\Pi_{\mathcal{P},\Omega_0}(f)$ fixed, the solution is unique. The orthogonal complement of $e^0(\mathcal{G}_{\Omega_0})$ in $\mathcal{G}_\Omega$ consists of all functions that coincide with a function from $\mathcal{P}$ on Ω_0.

Proof: We show that $e^0(f)$ satisfies the necessary and sufficient variational equation

$$(e^0(f), v)_\Omega = 0 \text{ for all } v \in \mathcal{G} \text{ with } \lambda(r_0(v)) = 0 \text{ for all } \lambda \in \mathcal{P}^{-}_{\Omega\,0}.$$

The condition on v is equivalent to $r_0(v) \in \mathcal{P}$ due to Lemma 3.3.13 (*SuffPol)*, and it is equivalent to the same condition restricted to all functionals of the form $\delta_{x,\Xi}$. But since

$$(e^0(f), v)_\Omega = (f, r_0(v))_{\Omega_0}$$

holds for all $v \in \mathcal{G}$, we have that $e^0(f)$ solves the recovery problem. The rest is standard. □

Corollary 3.5.8 *(DensCor) Let Φ and the functions from $\mathcal{P}$ be continuous. For $\Omega \subset I\!R^d$ and Ω_0 dense in Ω, the embedding $e^0(\mathcal{G}_{\Omega_0})$ of $\mathcal{G}_{\Omega_0}$ is dense in $\mathcal{G}_\Omega$.*

Proof: Let v be in the orthogonal complement of $e^0(\mathcal{G}_{\Omega_0})$ in $\mathcal{G}_\Omega$. Then v coincides with a function p from $\mathcal{P}$ on Ω_0. By continuity as following from Theorem 3.3.22 (*CPDSuffT)*, $v = p$ holds on all of Ω. □

3.5.4 Infinite Problems

We take an increasing sequence of $\mathcal{P}$-nondegenerate data sets $X_M := \{x_1, \dots, x_M\} \subset \Omega$ for $M = Q, Q+1, \dots$ and denote the interpolant to data from some $f \in \mathcal{G}_\Omega$ on the set X_M by f_M. The usual orthogonality property, as induced by (3.1.15, *charmin*), implies both

$$\begin{aligned} (f - f_N, f_N)_{\Phi,\Omega} &= 0 \\ (f_M - f_N, f_N)_{\Phi,\Omega} &= 0 \end{aligned}$$

for all $M \geq N \geq Q$. This implies the Pythagorean laws

$$\begin{array}{rcl} |f - f_N|^2_{\Phi,\Omega} + |f_N|^2_{\Phi,\Omega} & = & |f|^2_{\Phi,\Omega} \\ |f_M - f_N|^2_{\Phi,\Omega} + |f_N|^2_{\Phi,\Omega} & = & |f_M|^2_{\Phi,\Omega}. \end{array}$$

Thus the sequence $\{|f_N|^2_{\Phi,\Omega}\}_N$ converges monotonically to some value bounded by $|f|^2_{\Phi,\Omega}$. It necessarily is a Cauchy sequence, and thus $\{f_M\}_M$ and $\{f_M - \Pi_{\mathcal{P}} f_M\}_M$ are Cauchy sequences. The latter lies in a Hilbert space and thus is convergent to a function that we can write as $f_\infty - \Pi_{\mathcal{P}} f_\infty$ with some $f_\infty \in \mathcal{G}_\Omega$, and where the term $\Pi_{\mathcal{P}} f_\infty$ is at our disposal. We fix it to be identical to $\Pi_{\mathcal{P}} f$. Since all of the interpolation functionals are in $\mathcal{G}^*_\Omega$, and since we can conclude from Corollary 3.1.32 (*PolRepCol*) that

$$(\Pi_{\mathcal{P}} f_M)(x_j) = (\Pi_{\mathcal{P}} f)(x_j) \text{ for all } M \geq j,$$

we see that

$$(f_\infty - \Pi_{\mathcal{P}} f)(x_j) = \lim_{M\to\infty} (f_M - \Pi_{\mathcal{P}} f_M)(x_j) = f(x_j) - (\Pi_{\mathcal{P}} f)(x_j)$$

such that $f(x_j)$ and $f_\infty(x_j)$ agree for all j.

Theorem 3.5.9 *(IIT) For functions f in the native space $\mathcal{G}_\Omega$ corresponding to a conditionally positive definite function Φ on Ω, one can solve all recovery problems based on countably many $\mathcal{P}$-nondegenerate data in the form of functionals $\lambda_j \in \mathcal{G}^*_\Omega$ for $j \in I\!N$. In case of Lagrange data, the solution coincides with the function $f^0 := e^0 r_0(f)$ if Ω_0 is the set of all data locations.*

Proof: The first assertion easily generalizes from the case of Lagrange data to general functionals. To prove the second, we know that both f^0 and f_∞ satisfy all interpolation conditions. Since we have

$$|f_M|_{\mathcal{G}_\Omega} \leq |f_\infty|_{\mathcal{G}_\Omega},\ |f_M|_{\mathcal{G}_\Omega} \leq |f^0|_{\mathcal{G}_\Omega}$$

because f_M is based on less data that the other recovery functions, and since $|f_M|_{\mathcal{G}_\Omega}$ converges to $|f_\infty|_{\mathcal{G}_\Omega}$, we get

$$|f_\infty|_{\mathcal{G}_\Omega} \leq |f^0|_{\mathcal{G}_\Omega}.$$

But the seminorm of f^0 is minimal under all other recovery functions, proving that f_∞ also solves the total recovery problem on Ω_0. Two solutions differ by a function of $\mathcal{P}$, but since the data are $\mathcal{P}$-nondegenerate, they coincide. $\square$

We could proceed from here towards orthogonal expansions of recovery functions, but we shall delay these things for section 4.4 (*SecRecCon*).

3.5.5 The Connection to L_2 via Convolution

(*SecCLC*) In this section we require the subset $\Omega_0 \subseteq \Omega$ to be compact and assume continuity of Φ and functions in $\mathcal{P}$. We then define the bilinear form (*Norm2*)

$$(f,g)_{2,\Omega_0} := \int_{\Omega_0} \delta_{x,\mathcal{P}} r_0 f \cdot \delta_{x,\mathcal{P}} r_0 g dx = (\delta_{\cdot,\mathcal{P}} r_0 f, \delta_{\cdot,\mathcal{P}} r_0 g)_{L_2(\Omega_0)} \qquad (3.5.10)$$

for all $f,g \in \mathcal{G}_\Omega$. Because of

$$\begin{array}{rcl} (\delta_{x,\mathcal{P}} r_0 f)^2 & = & (\delta^y_{x,\mathcal{P}} \Phi(y,\cdot), r_0 f)^2_{\Phi,\Omega_0} \\ & \leq & \|\delta_{x,\mathcal{P}}\|^2_{\Phi,\Omega_0} \|r_0 f\|^2_{\Phi,\Omega_0} \\ & = & \Psi(x,x)\|r_0 f\|^2_{\Phi,\Omega_0} \\ & \leq & \Psi(x,x)\|f\|^2_{\Phi_\Omega} \end{array}$$

and since $\Psi_\mathcal{P}(x,y) = (\delta_{x,\mathcal{P}}, \delta_{y,\mathcal{P}})_\Phi$ is the normalization of Φ with respect to $\mathcal{P}$, we can use its continuity and get that the above function is continuous. Thus we can integrate it over Ω_0 to see that the bilinear form (3.5.10, *Norm2*) is well-defined and continuous with respect to the bilinear form in $\mathcal{G}_{\Omega_0}$ and $\mathcal{G}_\Omega$:

$$\begin{array}{rcl} |(f,g)_{2,\Omega_0}| & \leq & \|r_0 f\|_{2,\Omega_0} \|r_0 g\|_{2,\Omega_0} \\ & \leq & \|\Psi_\mathcal{P}(x,x)\|^2_{L_2(\Omega_0)} \|r_0 f\|_{\Phi,\Omega_0} \|r_0 g\|_{\Phi,\Omega_0}, \\ & \leq & \|\Psi_\mathcal{P}(x,x)\|^2_{L_2(\Omega_0)} \|f\|_{\Phi,\Omega} \|g\|_{\Phi,\Omega}, \end{array}$$

The inner product $(\cdot,\cdot)_{2,\Omega_0}$ coincides with the inner product of $L_2(\Omega_0)$ applied to functions from the Hilbert subspace $\mathcal{F}_\Omega$ after restriction to Ω_0, i.e.:

$$(f,g)_{2,\Omega_0} = (r_0 f, r_0 g)_{L_2(\Omega_0)} \text{ for all } f,g \in \mathcal{F}_\Omega.$$

This proves that the restriction mapping r_0 can also be viewed as a continuous mapping from $\mathcal{F}_\Omega$ into $L_2(\Omega_0)$. If we denote bilinear forms carefully, we do not have to distinguish between r_0 as a map into $L_2(\Omega_0)$ or as a map into $\mathcal{G}_{\Omega_0}$.

We now proceed to construct the adjoint $c^0 : L_2(\Omega_0) \to \mathcal{F}_\Omega$ of r_0 considered as a mapping into $L_2(\Omega_0)$. For any $g \in L_2(\Omega_0)$ we can consider the continuous linear functional $f \mapsto (r_0 f, g)_{2,\Omega_0}$ on the Hilbert space $\mathcal{F}_\Omega$. Then there is a function $c^0 g \in \mathcal{F}_\Omega$ such that

(*smapdef*)

$$(r_0 f, g)_{L_2(\Omega_0)} = (f, c^0 g)_{\Phi,\Omega} \text{ for all } f \in \mathcal{F}_\Omega,\ g \in L_2(\Omega_0). \qquad (3.5.11)$$

The image of c^0 in $\mathcal{F}_\Omega$ is orthogonal to the kernel of r_0, and thus it coincides with $\mathcal{F}_\Omega \cap e^0(\mathcal{G}_{\Omega_0})$. Furthermore, it follows from (3.5.11, *smapdef*) that $c^0 r_0$ is a nonnegative self-adjoint operator on $\mathcal{F}_\Omega$. But there is a more convenient way to represent c^0 that sheds some light on the functions in $\mathcal{G}_\Omega$:

Theorem 3.5.12 *(ConvTh) The map c^0 takes any function $v \in L_2(\Omega_0)$ into the generalized convolution*

$$(v * \Psi)(x) := \int_{\Omega_0} v(y)\Psi(x,y)dy,\ x \in \Omega.$$

Proof: Let us first check that these convolutions are in $\mathcal{G}_\Omega$. The integral is well-defined because both integrands are in $L_2(\Omega_0)$. Take any $\lambda_{X,M,\alpha} \in \mathcal{P}_\Omega^-$ and form

$$\begin{aligned}
|\lambda_{X,M,\alpha}(v * \Psi)| &= |\lambda^x_{X,M,\alpha}\int_{\Omega_0} v(y)\Psi(x,y)dy| \\
&= |\int_{\Omega_0} v(y)\lambda^x_{X,M,\alpha}\Psi(x,y)dy| \\
&= |(v, r_0 F_{\lambda_{X,M,\alpha}})_{L_2(\Omega_0)}| \\
&\leq \|v\|_{L_2(\Omega_0)}\|r_0 F_{\lambda_{X,M,\alpha}}\|_{L_2(\Omega_0)} \\
&\leq \|\Psi_\mathcal{P}(x,x)\|_{L_2(\Omega_0)}\|v\|_{L_2(\Omega_0)}\|\lambda_{X,M,\alpha}\|_{\Phi,\Omega}.
\end{aligned}$$

Thus we are in $\mathcal{G}_\Omega$, and we see immediately by application of $\Pi_\mathcal{P}$ that the convolution lies in $\mathcal{F}_\Omega$, since we used Ψ instead of Φ.

Let us do the same to $c^0 v$ and compare. Then

$$\begin{aligned}
\lambda_{X,M,\alpha}(c^0 v) &= (F_{\lambda_{X,M,\alpha}}, c^0 v)_{\Phi,\Omega} \\
&= (r^0 F_{\lambda_{X,M,\alpha}}, v)_{L_2(\Omega_0)}
\end{aligned}$$

shows that

$$\lambda_{X,M,\alpha}(c^0 v - v * \Psi) = 0$$

for all functionals $\lambda_{X,M,\alpha} \in \mathcal{P}_\Omega^-$. Then the two functions can only differ by a function from $\mathcal{P}$, but since both are in $\mathcal{F}$, they agree everywhere. □

Corollary 3.5.13 *(ConvThCor) The ranges of the extension map e^0 and the convolution map c^0 have the same closure in $\mathcal{F}_\Omega$.*

Proof: In fact, if $g \in \mathcal{F}$ is orthogonal to all $c^0(v)$ for $v \in L_2(\Omega_0)$, then

$$0 = (g, c^0 v)_{\Phi,\Omega} = (r_0 g, v)_{L_2(\Omega_0)}$$

implies that g vanishes on Ω_0. The corresponding condition

$$0 = (g, e^0 v)_{\Phi,\Omega} = (r_0 g, v)_{\Phi,\Omega_0}$$

for all $v \in \mathcal{F}_{\Omega_0}$ first implies that g coincides with a function from $\mathcal{P}$ in Ω, but since we work on $\mathcal{F}_\Omega$ here, it must be zero on Ω. Thus the orthogonal complements in $\mathcal{F}_\Omega$ of the two ranges coincide. and the closures must coincide. □

There is another way to look at the map c^0. Consider any function $v \in (r_0\mathcal{P})^- L_2(\Omega_0)$ and the corresponding functional

$$\lambda_v \ : \ g \mapsto (r_0 g, v)_{L_2(\Omega_0)}$$

which is continuous on $\mathcal{G}_\Omega$ and lies in $\mathcal{F}_\Omega^*$. Its Riesz representer is $c^0(v)$, because

$$(r_0 g, v)_{L_2(\Omega_0)} = \lambda_v(g) = (g, c^0 v)_{\Phi,\Omega}$$

holds for all $g \in \mathcal{F}$.

3.6 Example: Cubic Splines in One Variable

This section serves to illustrate the construction of the native space and the extension/convolution maps for cubic splines. We consider the radial function $\Phi(x,y) := \phi(\|x-y\|_2)$ with $\phi(r) = r^3$ in $I\!R^1 =: \Omega$, which will turn out to be conditionally positive definite of order $m = 2$ there. Furthermore, we restrict the data to an interval $\Omega_0 = [a,b] \subset I\!R$ later. For $\mathcal{P} = I\!P_2^1$ and any $\lambda_{X,M,\alpha} \in \mathcal{P}_\Omega^-$ according to (3.3.1, *Deflxma*) we have

$$F(\lambda_{X,M,\alpha})(x) = \sum_{j=1}^{M} \alpha_j |x - x_j|^3,$$

due to (3.3.4, *Fmapdef*), and with $|x|^3 = 2x_+^3 - x^3$ we find

$$\begin{array}{rcl} F(\lambda_{X,M,\alpha}) & = & \displaystyle\sum_{j=1}^{M} \alpha_j (2(x-x_j)_+^3 - (x-x_j)^3) \\ & = & 2\displaystyle\sum_{j=1}^{M} \alpha_j (x-x_j)_+^3 - \displaystyle\sum_{j=1}^{M} \alpha_j (x_j^3 - 3x x_j^2) + 0 \end{array}$$

because $\lambda_{X,M,\alpha}$ annihilates linear polynomials. Then

(*Fmaplin*)

$$\frac{d^2}{dx^2} F(\lambda_{X,M,\alpha})(x) = 12 \sum_{j=1}^{M} \alpha_j (x - x_j)_+^1 \qquad (3.6.1)$$

is a piecewise linear function with support in

$$x_1 < x_2 < \ldots < x_N.$$

If two functionals

$$\lambda_{X,M,\alpha}(f) = \sum_{j=1}^{M} \alpha_j f(x_j), \quad \lambda_{Y,N,\beta}(f) = \sum_{k=1}^{N} \beta_k f(y_k)$$

from $\mathcal{P}_\Omega^-$ are given, then

$$(\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta})_\Phi = \sum_{j=1}^{M}\sum_{k=1}^{N} \alpha_j\beta_k|x_j - y_k|^3$$

by definition, and we want to compare this to

$$\begin{array}{rl} & (\frac{d^2}{dx^2}F(\lambda_{X,M,\alpha})(x), \frac{d^2}{dx^2}F(\lambda_{Y,N,\beta})(x))_{L_2(\mathbb{R})} \\ := & \int_{-\infty}^{\infty} \frac{d^2}{dx^2}F(\lambda_{X,M,\alpha})(x)\frac{d^2}{dx^2}F(\lambda_{Y,N,\beta})(x)dx. \\ = & \int_a^b \frac{d^2}{dx^2}F(\lambda_{X,M,\alpha})(x)\frac{d^2}{dx^2}F(\lambda_{Y,N,\beta})(x)dx. \end{array}$$

Using $x_+^1 = (-x)_+^1 + x$ we rewrite $\frac{d^2}{dx^2}F(\lambda_{X,M,\alpha})$ as

$$\begin{array}{rcl} \frac{d^2}{dx^2}F(\lambda_{X,M,\alpha})(x) & = & 12\ \sum\limits_{j=1}^{M}\alpha_j(x_j - x)_+^1 + 12\ \sum\limits_{j=1}^{M}\alpha_j(x - x_j) \\ & = & 12\sum\limits_{j=1}^{M}\alpha_j(x_j - x)_+^1. \end{array}$$

(note the swap of x with x_j) and get

Theorem 3.6.2 *Under the above assumptions,*

(cubqf)

$$\begin{array}{rcl} (\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta})_\Phi & = & \sum\limits_{j=1}^{M}\sum\limits_{k=1}^{N}\alpha_j\beta_k|x_j - y_k|^3 \\ & = & \frac{1}{12}(\frac{d^2}{dx^2}F(\lambda_{X,M,\alpha})(x), \frac{d^2}{dx^2}F(\lambda_{Y,N,\beta})(x))_{L_2(\mathbb{R})} \end{array} \tag{3.6.3}$$

for all $\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta} \in \mathcal{P}_\Omega^-$.

Proof: We use Taylor's formula

$$f(x) = f(a) + (x-a)f'(a) + \int_a^b f''(u)(x-u)_+^1du$$

for functions $f \in C^2[a,b]$ and $a \le x \le b$. Fixing $y \in [a,b]$, we insert $f_y(u) = (y-u)_+^3/3!$ and get

$$\begin{array}{rcl} \frac{(y-x)_+^3}{3!} & = & \frac{(y-a)^3}{3!} - (x-a)\frac{(y-a)^2}{2!} + \int_a^b (y-u)_+^1(x-u)_+^1du \\ & = & \frac{1}{2\cdot 3!}(|y-x|^3 + (y-x)^3). \end{array}$$

We now apply functionals $\lambda_{X,M,\alpha}$ and $\lambda_{Y,N,\beta}$. This yields

$$
\begin{aligned}
& \frac{1}{12}\sum_{j=1}^{M}\sum_{k=1}^{N}\alpha_j\beta_k|x_j-y_k|^3 \;+\; \frac{1}{12}\sum_{j=1}^{M}\sum_{k=1}^{N}\alpha_j\beta_k(y_k-x_j)^3\\
= & \frac{1}{12}\sum_{j=1}^{M}\sum_{k=1}^{N}\alpha_j\beta_k|x_j-y_k|^3 \;+\; 0\\
= & \frac{1}{6}\sum_{j=1}^{M}\sum_{k=1}^{N}\alpha_j\beta_k(y_k-a)^3 \;-\; \frac{1}{2}\sum_{j=1}^{M}\sum_{k=1}^{N}\alpha_j\beta_k(x_j-a)(y_k-a)^2\\
& \;+\; \sum_{j=1}^{M}\sum_{k=1}^{N}\alpha_j\beta_k\int_a^b (y_k-u)_+^1(x_j-u)_+^1du\\
= & \; 0-0+\sum_{j=1}^{M}\sum_{k=1}^{N}\alpha_j\beta_k\int_a^b (y_k-u)_+^1(x_j-u)_+du\\
= & \left(\tfrac{d^2}{dx^2}F(\lambda_{X,M,\alpha}),\tfrac{d^2}{dx^2}F(\lambda_{Y,N,\beta})\right)_{L_2(I\!R)},
\end{aligned}
$$

where the functions $\frac{d^2}{dx^2}F(\lambda_{X,M,\alpha})$ and $\frac{d^2}{dx^2}F(\lambda_{Y,N,\beta})$ are supported in $[x_1,x_M]$ and $[y_1,y_M]$, respectively, such that the L_2 integral could be suitably restricted. □

Corollary 3.6.4 *The function $\Phi(x,y) := \phi(\|x-y\|_2)$ with $\phi(r)=r^3$ is conditionally positive definite of order 2 on $I\!R$.*

Proof: Theorem 3.6.2 (*thcs2*) yields that the quadratic form (3.6.3, *cubqf*) is positive semidefinite. If $\|\lambda_{X,M,\alpha}\|_\Phi$ vanishes, then $\frac{d^2}{dx^2}F(\lambda_{X,M,\alpha}=0$ holds, and the representation (3.6.1, *Fmaplin*) as a piecewise linear function implies that all coefficients α_j must vanish. □

We now use (3.3.3, *calgdef*) for $\Omega_0 := [a,b]$ to define the abstract space

$$\mathcal{G}_{\Omega_0}=\mathcal{G}_{[a,b]}=\{g\;:\;[a,b]\to I\!R\;:\;|\lambda(f)|\le C_f\|\lambda\|_\Phi \text{ for all } \lambda\in\mathcal{P}_{\Omega\,0}^{-}\}$$

and assert that it coincides with the space

$$W_2^2[a,b] := \{\;g\;:\;[a,b]\to I\!R\;:\;g''\in L_2[a,b]\;\}$$

.

Lemma 3.6.5

$$W_2^2[a,b]\subseteq\mathcal{G}_{[a,b]}.$$

Proof. Generalizing Taylor's formula for $f \in W_2^2[a,b]$, we find for all $\lambda_{X,M,\alpha} \in \mathcal{P}_{\Omega\,0}^-$ the identity

$$\begin{aligned}
\lambda_{X,M,\alpha}(f) &= \sum_{j=1}^{M} \alpha_j f(x_j) = 0 + \int_a^b f''(u) \sum_{j=1}^{M} (x_j - u)_+^1 du \\
&= (\tfrac{d^2}{dx^2} f, \tfrac{d^2}{dx^2} F(\lambda_{X,M,\alpha}))_{L_2[a,b]} \\
&\leq \|\tfrac{d^2}{dx^2} f\|_{L_2[a,b]} \cdot \|\tfrac{d^2}{dx^2} F(\lambda_{X,M,\alpha})\|_{L_2[a,b]} \\
&\leq \|\tfrac{d^2}{dx^2} f\|_{L_2[a,b]} \cdot \sqrt{12} \cdot \|\lambda\|_{\Phi}
\end{aligned}$$

□

Lemma 3.6.6

$$\mathcal{G} \subseteq W_2^2[a,b].$$

Proof. Define the subspace

$$\mathcal{F}_0 := \{ \frac{d^2}{dx^2} F(\lambda_{X,M,\alpha}) : \lambda \in \mathcal{P}_{\Omega\,0}^- \}$$

of $L_2[a,b]$. It carries an inner product

$$(\frac{d^2}{dx^2} F(\lambda_{X,M,\alpha}), \frac{d^2}{dx^2} F(\lambda_{Y,N,\beta}))_{L_2[a,b]} = 12(\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta})_{\Phi}$$

constructed from the inner product $(\cdot,\cdot)_\Phi$, and we define $\mathcal{F} := \overline{\mathcal{F}_0}$ to be the L_2 closure of $\mathcal{F}_0$ with respect to $(\ ,\ )_{L_2[a,b]}$. Any $g \in \mathcal{G}_{[a,b]}$ defines a linear functional on $\mathcal{F}_0$ by

$$\frac{d^2}{dx^2} F(\lambda_{X,M,\alpha}) \mapsto \lambda_{X,M,\alpha}(g), \qquad \lambda_{X,M,\alpha} \in \mathcal{P}_{\Omega\,0}^-.$$

Here, we used that the map $\lambda_{X,M,\alpha} \mapsto \frac{d^2}{dx^2} F(\lambda_{X,M,\alpha})$ is one–to–one on $\mathcal{P}_{\Omega\,0}^-$. The above functional is continuous on $\mathcal{F}_0$ by definition of $\mathcal{G}_{[a,b]}$. Thus there is some $h_g \in \mathcal{F}_{[a,b]} = \overline{\mathcal{F}_0} \subseteq L_2[a,b]$ such that

$$\lambda_{X,M,\alpha}(g) = (\frac{d^2}{dx^2} h_g, \frac{d^2}{dx^2} F(\lambda_{X,M,\alpha}))_{L_2[a,b]}$$

for all $\lambda \in \mathcal{P}_{\Omega\,0}^-$, and we tacitly assume $h_g \in W_2^2[a,b]$ (we can start with $\frac{d^2}{dx^2} h_g = f_g \in L_2[a,b]$ and do integration). Now Taylor's formula for h_g yields

$$\lambda_{X,M,\alpha}(h_g) = 0 + (\frac{d^2}{dx^2} h_g, \frac{d^2}{dx^2} F(\lambda_{X,M,\alpha}))_{L_2[a,b]} = \lambda_{X,M,\alpha}(g)$$

for all $\lambda_{X,M,\alpha} \in \mathcal{P}^{-}_{\Omega\,0}$. By the standard argument from Lemma 3.3.13 (*SuffPol*) we see that $g = h_g + p_g$ with a polynomial $p_g \in I\!P_2^1$. □

This finishes the construction of the native space, and we can reconstruct functions from $\mathcal{G}_{[a,b]} = W_2^2[a,b]$ from data at locations $a \leq x_1 < x_2 < \ldots < x_M \leq b$ uniquely by cubic splines of the form

($cubrep$)

$$s(x) = \sum_{j=1}^{M} \alpha_j |x - x_j|^3 + \sum_{k=0}^{1} \beta_k x^k \tag{3.6.7}$$

under the two additional conditions

($cubreprestr$)

$$\sum_{j=1}^{M} \alpha_j x_j^k = 0,\ k = 0, 1. \tag{3.6.8}$$

The representation (3.6.1, *Fmaplin*) shows that these conditions imply linearity of s outside of $[x_1, x_M]$. Thus the solution is a **natural cubic spline**. The above representation extends to all of $I\!R$ and can easily be shown to coincide with the extension defined via the general map e^0.

We now want to show how the natural boundary conditions come out of Theorem 3.5.7 (*ExtTh2*). There we concluded that the extension of the recovery problem is orthogonal to all functions that coincide with a function from $\mathcal{P}$ on the recovery domain Ω_0. The extensions here are valid in $W_2^2(I\!R)$ with bilinear form $(f'', g'')_{L_2(I\!R)}$, and the splines s constructed here must be orthogonal with respect to the above bilinear form to all functions v on $I\!R$ that are linear in $[a, b]$. We take any function $w = v''$ that is in $L_2(I\!R)$ and vanishes on $[a, b]$. Then

$$0 = (s'', v'')_{L_2(I\!R)} = (s'', w)_{L_2(I\!R)} = (s'', w)_{L_2(-\infty,a]} + (s'', w)_{L_2[b,\infty)}$$

implies $s'' = 0$ in $(-\infty, a]$ and $[b, \infty)$, as required.

The next thing is to look at the convolution map s^0 and Theorem 3.5.12 (*ConvTh*) in this case. The explicit evaluation of the normalization Ψ of Φ is possible, but left to the reader. Theorem 3.5.12 (*ConvTh*) asserts that the closures of the ranges of the convolution map s^0 and the extension map e^0 are the same. Let $v \in L_2[a, b]$ be given. We want to show that $v * \Psi$ is in $W_2^2[a, b]$ with $(v * \Psi)" = 0$ outside of $[a, b]$. We use (3.3.21, *DefSymmPsi*) to get

$$(v * \Psi)"(y) = \int_a^b v(x) \frac{d^2}{dy^2} \Psi(x, y) dx = \int_a^b v(x) \frac{d^2}{dy^2} F(\delta_{x,\Xi})(y) dx,$$

where we took derivatives under the integral. But (3.6.1, *Fmaplin*) shows that the second factor vanishes if y is outside of $[a, b]$. To show that $v * \Psi$ is in $W_2^2[a, b]$, compare the above formula with Taylor's formula after inserting $\delta_{x,\Xi}$ for $\Xi = \{a, b\}$.

4 Power Functions and Applications

(*SecPF*) This section introduces the notion of power functions. They associate to each point x of a domain Ω and to each linear quasi-interpolation process $g \mapsto S(g)$ the norm $P(x) = P_{S,\mathcal{G}}(x)$ of the error functional $g \mapsto g(x) - S(g)(x)$ with respect to the space $\mathcal{G}$ of functions g to be considered. Thus they describe the worst-case behaviour of the reconstruction process S at x and are very useful for comparing different reconstruction processes. We illustrate this for some simple examples.

When specialized to the optimal recovery processes considered in this text, their square $P^2(x)$ has a representation as the diagonal of an explicitly available quadratic form $P(x, y)$, which in turn has remarkable properties. In particular, it allows recursive constructions like Newton's interpolation formula, and it can be optimized with respect to placement of centers.

4.1 Power functions

(*SubSecPF*) Assume that we have a quite general process that associates to each function g in a space $\mathcal{G}$ of functions on Ω another function $S(g) \in \mathcal{G}$ such that the map $S\ :\ g \mapsto S(g)$ is linear. The space $\mathcal{G}$ should carry at least a seminorm $|\cdot|$ with nullspace $\mathcal{P}$.

Definition 4.1.1 *The function*

(DefPowfct)

$$P(x) := P_{S,\mathcal{G}}(x) := \sup_{\substack{g \in \mathcal{G} \\ |g| \neq 0}} \frac{|(g - S(g))(x)|}{|g|_{\mathcal{G}}} \in I\!R \cup \{\infty\} \tag{4.1.2}$$

is the **power function** *of S with respect to Φ.*

This is nothing else than the norm of the pointwise error functional if the latter is finite:

$$P(x) := \|\delta_{x,S}\| \text{ with } \delta_{x,S}(g) := g(x) - S(g)(x).$$

It yields the elementary error bound

(EqgSg)

$$|g(x) - S(g)(x)| \leq P(x)|g|, \quad g \in \mathcal{G}, \; x \in \Omega. \tag{4.1.3}$$

If the projection property $S \circ S = S$ holds, then one can insert $g - S(g)$ instead of g into this bound to get

(EqgSg1)

$$|g(x) - S(g)(x)| \leq P(x)|g - S(g)|, \quad g \in \mathcal{G}, \; x \in \Omega, \tag{4.1.4}$$

which often is some improvement over (4.1.3, *EqgSg*), because we frequently have $|g - S(g)| \leq |g|$.

To make the reader somewhat more familiar with the notion of a power function, we recall interpolation by univariate polynomials of order at most n on n distinct points $x_1 < \ldots < x_n$ in $[a, b] \subset I\!R$. The space $\mathcal{G}$ is $C^n[a, b]$ with seminorm $|g|_n := \|g^{(n)}\|_{[a,b],\infty}$, and the interpolant to g will be denoted by $S(g)$. The usual error bound

$$|g(x) - S(g)(x)| \leq \frac{1}{n!} |\prod_{j=1}^{n} (x - x_j)| |g|_n$$

is precisely of the form (4.1.3, *EqgSg*), and the power function is

$$P(x) = \frac{1}{n!} \prod_{j=1}^{n} |x - x_j|,$$

since it is well-known that the error bound is exact.

Power functions can be associated to almost every process of approximation or interpolation, and they enable comparison between different processes S on the same space $\mathcal{G}$ as well as the comparison of the same process S on different spaces $\mathcal{G}$, respectively. Before we give some examples for this, let us give some straightforward alternative representations:

Lemma 4.1.5 *(LemARepPow) The power function $P_{S\mathcal{G}}$ of (4.1.2, DefPowfct) can be written as*

$$P_{S,\mathcal{G}}(x) := \sup_{\substack{g \in \mathcal{G} \\ S(g) = 0}} \frac{g(x)}{|g|_{\mathcal{G}}} \in I\!R \cup \{\infty\} = \sup_{\substack{g \in \mathcal{G} \\ S(g) = 0 \\ \|g\|_{\mathcal{G}} \leq 1}} g(x) \in I\!R \cup \{\infty\},$$

if S is linear, preserves functions in $\mathcal{P}$, and has the projection property $S \circ S = S$. □

4.2 Optimal Recovery Redefined

The notion of a power function allows to define optimal recovery processes in a somewhat different way. Let us fix the space $\mathcal{G}$ with its seminorm $|\cdot|_{\mathcal{G}}$, but let us consider different linear quasi-interpolation processes S. To make these comparable, we assume them to be based on the same information, i.e.: the evaluation of M linear continuous functionals $\lambda_1, \ldots, \lambda_M$ on $\mathcal{G}$. Since we restrict ourselves to linear recovery processes, we assume representations

(*SgRep*)

$$S(g)(x) := \sum_{j=1}^{M} u_{j,S}(x)\lambda_j(g), \tag{4.2.1}$$

where $u_{1,S}, \ldots, u_{M,S}$ are certain functions on Ω that may not necessarily be in $\mathcal{G}$. The recovery processes S just differ in their choice of these functions. In all cases it is reasonable to ask for preservation of the nullspace $\mathcal{P}$ of the seminorm on $\mathcal{G}$ under the recovery process S in the sense

$$S(p) = p \text{ for all } p \in \mathcal{P},$$

and we shall abbreviate this condition by $S_{|\mathcal{P}} = Id$. The main reason for this is that (4.1.3, *EqgSg*) should always hold.

The representation (4.2.1, *SgRep*) can now be considered for a fixed x as a function of the M real-valued quantities $u_{1,S}(x), \ldots, u_{M,S}(x)$. Then the **optimal linear recovery process** at x solves the finite-dimensional minimization problem

(*OLRx*)

$$\inf_{\substack{u_1,\ldots,u_M \in \mathbb{R} \\ \sum u_j(x)\lambda_j(p) = p(x),\ p \in \mathcal{P}}} \sup_{|g|_{\mathcal{G}} \neq 0} \frac{1}{|g|_{\mathcal{G}}} \left| g(x) - \sum_{j=1}^{M} u_j \lambda_j(g) \right|. \tag{4.2.2}$$

If a solution $u_1^*(x), \ldots, u_M^*(x)$ exists for all $x \in \Omega$, one can define the optimal process as

(*OSgRep*)

$$S^*(g)(x) := \sum_{j=1}^{M} u_j^*(x)\lambda_j(g). \tag{4.2.3}$$

It is by no means obvious that the solution, considered as a set of M functions on Ω, lies in the space $\mathcal{G}$. Section 4.3.2 (*SecOPFOR*) will prove that optimal recovery in the sense of 3.1.2 (*ORPF*) in spaces $\mathcal{G}$ with a positive definite bilinear form always is optimal in the above sense, too. The corresponding functions $u_1^*(x),\ldots,u_M^*(x)$ will then be of the form (3.1.13, *grep*) and certainly lie in $\mathcal{P}$. This is in agreement with our expectations, but the example considered in the next section will show that the situation may be much more difficult if we move away from a Hilbert space setting.

4.3 Example: Optimal Interpolation in $W^1_\infty[a,b]$

On the spaces $\mathcal{G} = W^1_\infty[a,b]$ or $\mathcal{G} = C^1[a,b]$ for $-\infty < a < b < \infty$ we have the seminorm $|f'|_\infty$ with the one-dimensional nullspace $\mathcal{P} = I\!P^1_1$ spanned by the constant functions. We now ask for the optimal quasi–interpolant under all representations (4.2.1, *SgRep*) for point evaluation functionals in the points $x_1,\ldots,x_M$ of the mesh $a =: x_0 \le x_1 < \ldots < x_M \le x_{M+1} := b$ under reproduction of constants. This means that we consider $x \in [a,b]$ as fixed and vary the M real numbers $u_1(x),\ldots,u_M(x)$ in

(*SgRepP*)

$$S_{u(x)}(g)(g) := \sum_{j=1}^M u_j(x)g(x_j) \text{ with } \sum_{j=1}^M u_j(x) = 1. \tag{4.3.1}$$

That is, we use Lemma 4.1.5 (*LemARepPow*) and want to solve

$$\begin{aligned}
& \inf_{\substack{u_j(x)\\(4.2.1,SgRep)}} \sup_{|f'|_\infty \ne 0} \frac{1}{|f'|_\infty} \left| f(x) - \sum_{j=1}^N u_j(x) f(x_j) \right| \\
= & \inf_{\substack{u_j(x)\\(4.2.1,SgRep)}} \sup_{|f'|_\infty \le 1} \left| f(x) - \sum_{j=1}^N u_j(x) f(x_j) \right|.
\end{aligned}$$

We start with three lemmas:

Lemma 4.3.2 *Let $a \le x_0 < x_1 < \ldots < x_M \le b$ and*

$$u_0, u_1, \ldots, u_M \in I\!R, \ \sum_{j=0}^M u_j = 0$$

be given. Then

$$\sup_{\substack{f \in W^1_\infty\\|f'|_\infty \le 1}} \left| \sum_{j=0}^M u_j \cdot f(x_j) \right| = \int_a^b \left| \sum_{j=0}^M u_j (x_j - t)^0_+ \right| dt.$$

Proof. We write

$$f(x) = f(a) + \int_a^x f'(t)dt = f(a) + \int_a^b f'(t)(x-t)_+^0 dt$$

and take a function f with $|f'| \leq 1$ to get

$$\begin{aligned} \left|\sum_{j=0}^{M} u_j \cdot f(x_j)\right| &= \left|f(a) \cdot 0 + \int_a^b f'(t) \sum_{j=0}^{M} u_j (x_j - t)_+^0\right| dt \\ &\leq \int_a^b \left|\sum_{j=0}^{M} u_j (x_j - t)_+^0\right| dt \end{aligned}$$

with equality for the special function f with

$$f'(t) = \operatorname{sgn}\left(\sum_{j=0}^{M} u_j (x_j - t)_+^0\right)$$

which clearly is in W_∞^1. □

Note that the proof does not require the ordering of the x_j.

Lemma 4.3.3 *Let $a \leq x_1 < x_2 < \ldots < x_M \leq b$ and $x \in [a, b]$ be given. For $u_1(x), \ldots, u_M(x)$ with*

$$1 = \sum_{j=1}^{M} u_j(x)$$

we have

$$\begin{aligned} &\sup_{\substack{f \in W_\infty^1 \\ \|f'|_\infty \leq 1}} \left| f(x) - \sum_{j=1}^{M} u_j(x) f(x_j) \right| \\ = &\int_a^b \left| (x-t)_+^0 - \sum_{j=1}^{M} u_j(x)(x_j - t)_+^0 \right| dt. \end{aligned}$$

Proof: Use Lemma 1 with $x_0 := x$, $u_0(x) := -1$ and reordering of points. □

Lemma 4.3.4 *(L3W2) Let $a = x_0 \leq x_1 < \ldots < x_M \leq x_{M+1} = b$ and $x \in [a, b]$ be given. Furthermore, assume*

$$1 = \sum_{j=1}^{M} u_j(x), \quad h_j := x_j - x_{j-1}, \quad 1 \leq j \leq M+1$$

and $x \in [x_{k-1}, x_k]$, $1 \le k \le M+1$. Then

$$
\begin{aligned}
&\int_a^b \left| (x-t)_+^0 - \sum_{j=1}^{M} u_j(x)(x_j - t)_+^0 \right| dt \\
=\ & \sum_{1 \le j \le k-1} h_j \left| 1 - \sum_{i \ge j} u_i(x) \right| \\
+\ & (x - x_{k-1}) \left| 1 - \sum_{i \ge k} u_i(x) \right| \\
+\ & (x_k - x) \left| \sum_{i \ge k} u_i(x) \right| \\
+\ & \sum_{M+1 \ge j \ge k+1} h_j \left| \sum_{i \ge j} u_i(x) \right|.
\end{aligned}
$$

Proof: We use the fact that the integrand is piecewise constant with breakpoints only at the x_j and at x. We thus split the integrals into

$$
\left(\sum_{j \le k-1} \int_{x_{j\perp 1}}^{x_j} \right) + \int_{x_{k\perp 1}}^{x} + \int_x^{x_k} + \left(\sum_{j \ge k+1} \int_{x_{j\perp 1}}^{x_j} \right)
$$

$$
\begin{aligned}
= \ & \sum_{j=1}^{k-1} h_j \cdot \left| 1 - \sum_{i \ge j} u_i(x) \right| \\
& + (x - x_{k-1}) \left| 1 - \sum_{i \ge k} u_i(x) \right| \\
& + (x_k - x) \left| \sum_{i \ge k} u_i(x) \right| \\
& + \sum_{j=k+1}^{M+1} h_j \left| \sum_{i \ge j} u_i(x) \right|.
\end{aligned}
$$

□

We now introduce new variables

$$
z_j(x) := \sum_{i \ge j} u_i(x), \qquad u_0(x) := u_{M+1}(x) := 0.
$$

Then $z_1(x) := 1$ for all x and

$$u_j(x) = z_j(x) - z_{j+1}(x), \qquad 1 \leq j \leq M.$$

We have to minimize

$$\sum_{j=2}^{k-1} h_j|1 - z_j(x)|$$

$$+(x - x_{k-1})|1 - z_k(x)| + (x_k - x)|z_k(x)|$$

$$+\sum_{j=k+1}^{M} h_j|z_j(x)|$$

under the constraint $z_1(x) = 1$, which appears only for $k = 1$ (or $x \in [a, x_1]$). In this case, we get the optimal value $x_1 - x$ with all other $z_j(x)$ being zero. Thus

$$u_i(x) = \delta_{j1} \quad \text{for } a \leq x \leq x_1, \;\; 1 \leq j \leq M + 1$$

for $x \in [a, x_1]$.

The case $x_M \leq x \leq b$ can be treated similarly. We thus can assume $2 \leq k \leq M$, and the optimal solution will have the property

$$\begin{array}{rclcrcccl} z_j(x) &=& 1 & \text{for} & 2 & \leq & j & \leq & k-1 \\ z_j(x) &=& 0 & \text{for} & k+1 & \leq & j & \leq & M \end{array}$$

because each single term can be minimized separately. This leaves $z_k(x)$ open and yields

$$\begin{array}{rcl} u_k(x) &=& z_k(x) \\ u_{k-1}(x) &=& 1 - z_k(x), \end{array}$$

the other $u_j(x)$ being zero automatically. We now form

$$\frac{x - x_{k-1}}{x_k - x_{k-1}}\;|1 - z_k(x)| + \frac{x_k - x}{x_k - x_{k-1}}\;|z_k(x)|$$

and set

$$\alpha := \frac{x - x_{k-1}}{x_k - x_{k-1}}\;, \qquad z := z_k(x)$$

to get the minimization of

$$\alpha|1 - z| + (1 - \alpha)|z|$$

for α fixed. Any polygonal function attains its minima at one of its breakpoints, and thus the overall minimum is

$$\begin{array}{rlcl}
\alpha, & \text{attained at} & z=0 & \text{, if } \alpha \leq 1-\alpha \\
1-\alpha, & \text{attained at} & z=1 & \text{, if } \alpha \geq 1-\alpha \\
\alpha = 1-\alpha, & \text{attained at} & 0 \leq z \leq 1 & \text{, if } \alpha = 1-\alpha.
\end{array}$$

This means

$$\begin{array}{rclrcllrcl}
u_k(x) & = & 0, & u_{k-1}(x) & = & 1 & \text{if } x - x_{k-1} & < & x_k - x \\
u_k(x) & = & 1, & u_{k-1}(x) & = & 0 & \text{if } x - x_{k-1} & > & x_k - x \\
u_k(x) & = & \beta, & u_{k-1}(x) & = & 1-\beta & \text{if } x - x_{k-1} & = & x_k - x
\end{array}$$

for any $\beta \in [0,1]$, and the value of the power function is

$$\begin{array}{rll}
x_1 - x & \text{for} & x \in [x_0, x_1], \\
\min(x - x_{k-1}, x_k - x) & \text{for} & x \in [a, x_k],\ 2 \leq k \leq M \\
x - x_M & \text{for} & x \in [x_M, b].
\end{array}$$

Theorem 4.3.5 *(ORWIT) The solution of the optimal recovery problem posed in $W_\infty^1[a,b]$ under the seminorm $|f'|_\infty$ and Lagrange data thus consists of the simple* **next-neighbour-rule**

> *Take the value of the nearest data point, if it is unique, and take some weighted arithmetic mean of the two nearest data values otherwise.*

The interpolant is piecewise constant with breakpoints halfway between the data points. The solution thus is not in W_∞^1. □

If we compare this solution with the classical piecewise linear B–spline interpolant that everybody would expect to be optimal, we use Lemma 4.3.4 (*L3W2*) for $x \in [x_{k-1}, x_k]$ with

$$u_{k-1}(x) = \frac{x - x_{k-1}}{x_k - x_{k-1}} \quad u_k(x) = 1 - u_{k-1}(x) = \frac{x_k - x}{x_k - x_{k-1}}$$

to get the power function

$$\begin{array}{rl}
& (x - x_{k-1})\dfrac{x_k - x}{x_k - x_{k-1}} + (x_k - x)\dfrac{x - x_{k-1}}{x_k - x_{k-1}} \\
= & \dfrac{2}{x_k - x_{k-1}}\,(x - x_{k-1})(x_k - x)
\end{array}$$

for $x \in [x_{k-1}, x_k]$, $2 \le k \le M$. At $x = x_{k-1}, x_k, \frac{1}{2}(x_{k-1} + x_k)$ the two power functions coincide, but the piecewise linear power function belonging to the optimal "nearest-data-interpolant" is pointwise smaller elsewhere (see Figure 18 (*FigPFEx1*)). Note here that the piecewise linear interpolant has a piecewise quadratic power function.

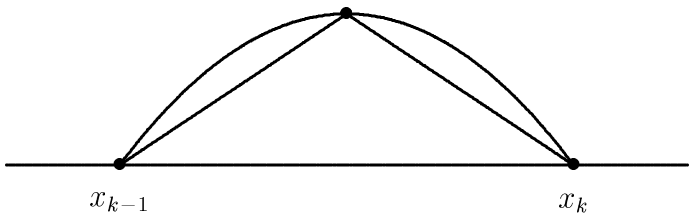

Figure 18: Comparison of Power Functions(*FigPFEx1*)

If we look at the case $k = 1$, we find the two solutions coincide for $a \le x \le x_1$.

Now we can ask for uniform minimization of both solutions on $[a, b]$ with respect to knot placement. Clearly, a minimum value η is attained in the situation of Figure 19 (*FigPFEx2*). Thus

$$2\eta \cdot M = b - a \quad \text{or} \quad \eta = \frac{b-a}{2M}$$

holds for the optimal knot placement. In $W_\infty^1[a, b]$, optimal interpolation on M points thus has a uniformly minimal error of $\frac{b-a}{2M}$, and the error is attained for both the natural piecewise linear spline (which is not pointwise optimal) and the discontinuous pointwise optimal interpolant.

This example shows how power functions allow precise and sometimes unexpected statements about the local behaviour of recovery processes in given spaces for given data.

4.3.1 Representations of Power Functions

(*SecRPF*) We now want to specialize the notion of a power function to the context of optimal recovery in function spaces. We assume the situation

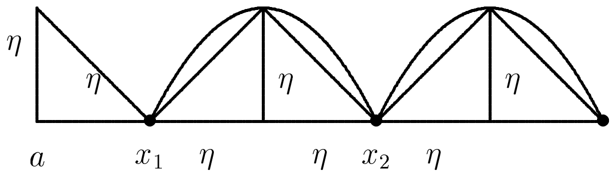

Figure 19: Optimal Knot Placement (*FigPFEx2*)

of Theorem 3.1.19 (*ORT2*) on page 36. That is, there are M linearly independent functionals $\lambda_1, \ldots, \lambda_M$ from $\mathcal{G}^*$ and a unique solution g^* of the optimal recovery problem (3.1.4, *ORP*). But we want to compare g^* with arbitrary other recoveries of g by linear methods that use the information $\lambda_j(g)$, $1 \leq j \leq M$. These have the form

(*GenRec*)

$$S_u(g) = \sum_{j=1}^{M} \lambda_j(g) u_j, \tag{4.3.6}$$

and we assume them to reproduce functions from $\mathcal{P}$. Then for each $x \in \Omega$ there is a functional

$$\delta_{x,u,S} \ : \ g \mapsto g(x) - S_u(g)(x) = g(x) - \sum_{j=1}^{M} \lambda_j(g) u_j(x)$$

vanishing on $\mathcal{P}$. The power function with respect to S_u is then representable via

$$P^2_{S_u,\Phi}(x) = |\delta_{x,u,S}|^2_\Phi.$$

It is now fairly easy to form

$$F(\delta_{x,u,S})(\cdot) = \Phi(x,\cdot) - \sum_{j=1}^{M} \lambda_j^z \Phi(z,\cdot) u_j(x)$$

and the function

(*DefPuxy*)

$$\begin{aligned} P_u(x,y) \ &:= \ (\delta_{x,u,S}, \delta_{y,u,S})_\Phi = \delta_{y,u,S}(F\delta_{x,u,S}) \\ &= \ \Phi(x,y) - \sum_{j=1}^{M} \lambda_j^z \Phi(z,y) u_j(x) \\ &\quad -\sum_{k=1}^{M} \lambda_k^z \Phi(,\cdot,z) u_k(y) + \sum_{j,k=1}^{M} \lambda_j^z \lambda_k^u \Phi(z,u) u_j(x) u_k(y) \end{aligned} \tag{4.3.7}$$

for all $x, y \in \Omega$. The reader will suspect some misuse of notation here, but the function $P_u(x, y)$ has some nice properties that justify this:

Theorem 4.3.8 *(PuT) The function $P_u(\cdot, \cdot)$ defined in (4.3.7, DefPuxy) satsfies*

- $P_u^2(x) = P_u(x, x) = |P_u(x, \cdot)|_\Phi^2$ *for all* $x \in \Omega$,
- $P_u(x, y) \le P_u(x)P_u(y)$ *for all* $x, y \in \Omega$,
- $P_u(x, \cdot)/P_u(\cdot)$ *attains its maximum* $P_u(x)$ *in* Ω *at* x,
- *if* $X = \{x_1, \ldots, x_M\} = \Xi = \{\xi_1, \ldots, \xi_q\}$ *is* $\mathcal{P}$*-nondegenerate and minimal, then* P_u *coincides with the normalization of* Φ *with respect to* Ξ,
- P_u *is another conditionally positive definite function that generates the same native space as* Φ.

Proof: The property $P_u^2(x) = P_u(x, x)$ follows from the definitions of both functions, and

$$|P_u(x, \cdot)|_\Phi = |F(\delta_{x,u,S})|_\Phi = \|\delta_{x,u,S}\|_\Phi$$

implies $|P_u(x, \cdot)|_\Phi = P_u(x)$. The next assertion is a consequence of the Cauchy-Schwarz inequality applied to the definition of $P_u(x, y)$, and together with the first it yields the third. The proof of the final property listed above is the same as for the normalization. □

The merit of (4.3.7, *DefPuxy*) is that it allows to write down the power function in explicit form and under quite general circumstances. This is of paramount importance for deriving error bounds in subsequent sections, and the basic feature is the optimality principle described in the next section.

4.3.2 Optimality of Power Functions of Optimal Recoveries

(*SecOPFOR*) Equation (4.3.7, *DefPuxy*) defines $P_u(x, x) = P_u^2(x)$ for fixed x as a quadratic form of the M real variables $u_j(x)$, $1 \le j \le M$. We now want to minimize this quadratic form with respect to these variables, but we have to consider the restrictions

(*PolRepEq*)

$$\delta_{x,u,S}(p_i) = p_i(x) - \sum_{j=1}^{M} \lambda_j(p_i)u_j(x) = 0, \ 1 \le i \le q \qquad (4.3.9)$$

imposed by reproduction of $\mathcal{P}$. Since $P_u^2(x)$ is nonnegative, the minimization must have a solution, and this solution can be characterized by the usual necessary conditions for quadratic optimization under linear constraints. There must be Lagrange multipliers $\beta_1(x), \ldots, \beta_q(x)$ such that the solution $u_j^*(x)$ of the restricted optimization is a minimum of the unrestricted function

$$P_u^2(x) + \sum_{i=1}^q \beta_i(x) \left(p_i(x) - \sum_{j=1}^M \lambda_j(p_i) u_j(x) \right) = 0$$

of $u_1(x), \ldots, u_M(x)$. Taking the derivative with respect to $u_k(x)$, we get

$$0 = -2\lambda_k^z \Phi(z,x) + 2\sum_{j=1}^M \lambda_j^z \lambda_k^u \Phi(z,u) u_j^*(x) - \sum_{i=1}^q \beta_i(x) \lambda_k(p_i).$$

We can rewrite this together with (4.3.9, *PolRepEq*) to get the system (*PFORSys*)

$$\begin{array}{rcrcll} \sum_{j=1}^M \lambda_j^z \lambda_k^u \Phi(z,u) u_j^*(x) & + & \sum_{i=1}^q \frac{-\beta_i(x)}{2} \lambda_k(p_i) & = & \lambda_k^z \Phi(z,x), & 1 \le k \le M \\ \sum_{j=1}^M \lambda_j(p_i) u_j^*(x) & + & 0 & = & p_i(x), & 1 \le i \le q \end{array} \tag{4.3.10}$$

The coefficient matrix of this system is the same as in (3.1.14, *EQsys3*) on page 33, if we use (3.2.14, *gjkrep*) on page 46. Thus the solution is in the span of the right-hand side, proving that $u_j^*(x) \in \mathcal{S}$, $1 \le j \le M$, as functions of x, but note that the necessary restriction on the $\beta_i(\cdot)$ of (3.1.33, *DefS*) is not satisfied. If we apply λ_k^x to these equations, we see that the conditions

$$\lambda_k^x(u_j(x)) = \delta_{jk},\ 1 \le j,k \le M$$

of interpolation are satsfied together with

$$\lambda_k^x(\beta_i(x)) = 0,\ 1 \le k \le M,\ 1 \le i \le q.$$

Thus we have

Theorem 4.3.11 *(OPFT) The power function $P_{u^*}(x)$ of the optimal recovery problem (3.1.4, ORP) is optimal with respect to u under all power functions $P_u(x)$ of recoveries of the form (4.3.6, GenRec) that reproduce $\mathcal{P}$.*
□

This is in line with the optimality of the generalized optimal power function $P_\Lambda(\mu)$ of (3.1.40, *GPDef*) on page 41. Note that there we used optimal recovery right from the start, but allowed a general functional μ instead of a point evaluation functional δ_x, while in this section we allowed general recoveries, but restricted ourselves to the special functional δ_x. The explicit correspondence is

$$P_\Lambda(\delta_x) = P_{u^*}(x)$$

between these two versions of optimal power functions. For use in the next sections, we rewrite the system (4.3.10, *PFORSys*) in an abbreviated form, omitting the asterisks standing for optimality and writing the free variables as indices:

(*PFORSys2*)

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} u_x \\ v_x \end{pmatrix} = \begin{pmatrix} \Phi_x \\ p_x \end{pmatrix}. \tag{4.3.12}$$

This means

$$\begin{array}{rcccl} Au_x & + & Pv_x & = & \Phi_x \\ P^T u_x & + & 0 & = & p_x \end{array}$$

and we compare with the shorthand form of the P function associated with the square $P^2(x) = P(x,x)$ of the optimal power function:

(*Pdef3*)

$$P(x,y) = \Phi(x,y) - u_x^T \Phi_y - u_y \Phi_x + u_x^T A u_y. \tag{4.3.13}$$

From the equations in (4.3.12, *PFORSys2*) we get

$$u_y^T A u_x + u_y^T P v_x = u_y^T A u_x + p_y^T v_x = u_y^T \Phi_x$$

and insert the result into (4.3.13, *Pdef3*) to arrive at

(*PnewDef*)

$$P(x,y) = \Phi(x,y) - u_x^T \Phi_y - v_x^T p_y. \tag{4.3.14}$$

This is one way of writing P explicitly in terms of the solution vectors of the system (4.3.12, *PFORSys2*). Note that the coefficient matrix of the system is constant, such that both u_x and v_x are linear functions of the right-hand sides Φ_x and v_x. Another simple consequence is the symmetry of the expression

(*QFSym*)

$$(u_y^T, v_y^T) \begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} u_x \\ v_x \end{pmatrix} = u_y^T \Phi_x + v_y^T p_x \tag{4.3.15}$$

with respect to swapping x and y.

4.4 Recursive Constructions

(*SecRecCon*) This section will study the effect of adding data to the recovery problem. It will turn out that there are some easy recursion formulae of Newton type.

4.4.1 Orthogonal Decomposition

(*SecOD*) We now want to add a new functional λ_{M+1} to the set $\Lambda = \{\lambda_1, \ldots, \lambda_M\}$. We shall use a tilde to denote symbols that now depend on $\tilde{\Lambda} := \Lambda \cup \{\lambda_{M+1}\}$ instead of Λ. Our basic result is

Theorem 4.4.1 *If P of (4.3.14, PnewDef) is based on Λ and $\tilde{P}$ is based on $\Lambda \cup \{\lambda_{M+1}\}$, then*

(4p11)

$$\tilde{P}(x, y) = P(x, y) - \frac{\lambda^z_{M+1} P(x, z) \cdot \lambda^z_{M+1} P(y, z)}{\lambda^u_{M+1} \lambda^v_{M+1} P(u, v)} \tag{4.4.2}$$

for all $x, y \in \Omega$.

Proof: We use (4.3.12, *PFORSys2*) and its extended version

(4p21)

$$\begin{pmatrix} A & a_{M+1} & P \\ a^T_{M+1} & \rho & p^T_{M+1} \\ P^T & p_{M+1} & 0 \end{pmatrix} \begin{pmatrix} \tilde{u}_x \\ f_x \\ \tilde{v}_x \end{pmatrix} = \begin{pmatrix} \Phi_x \\ g_x \\ p_x \end{pmatrix}, \tag{4.4.3}$$

where we introduced shorthands for

$$\begin{array}{rclrcl} \rho & = & \lambda^u_{M+1} \lambda^v_{M+1} \Phi(u, v) & g_x & = & \lambda^z_{M+1} \Phi(z, x) \\ \Phi_x & = & (\lambda^z_j \Phi(x, z))^T_{1 \le j \le M} & p_x & = & (p_j(x))^T_{1 \le j \le Q} \\ A & = & (\lambda^u_j \lambda^v_k \Phi(u, v))_{1 \le j, k \le M} & P & = & (\lambda^z_j p_k(z))_{1 \le j \le M, 1 \le k \le Q} \\ a_{M+1} & = & (\lambda^u_j \lambda^v_{M+1} \Phi(u, v))_{1 \le j \le M} & p_{M+1} & = & (\lambda^v_{M+1} p_j(v))_{1 \le j \le Q} \end{array}$$

and kept $\tilde{u}_x, \tilde{v}_x$ at the same size as u_x, v_x. Then

$$\begin{aligned} \begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \tilde{u}_x - u_x \\ \tilde{v}_x - v_x \end{pmatrix} &= -f_x \begin{pmatrix} a_{M+1} \\ p_{M+1} \end{pmatrix} \\ &= -f_x \lambda^z_{M+1} \begin{pmatrix} \Phi_z \\ p_z \end{pmatrix} \\ &= -f_x \lambda^z_{M+1} \begin{pmatrix} A & P \\ p^T & 0 \end{pmatrix} \begin{pmatrix} u_z \\ v_z \end{pmatrix} \end{aligned}$$

if we subtract (4.3.12, *PFORSys2*) from part of (4.4.3, *4p21*) and use (4.3.12, *PFORSys2*) after application of the functional λ_{M+1}. This yields

$$\begin{aligned}\tilde{u}_x &= u_x - f_x \lambda^z_{M+1} u(z)\\ \tilde{v}_x &= v_x - f_x \lambda^z_{M+1} v(z),\end{aligned}$$

and we use the remaining equation of (4.4.3, *4p21*) for

$$\begin{aligned}g_x &= a^T_{M+1}\tilde{u}_x + \rho f_x + p^T_{M+1}\tilde{v}_x\\ &= a^T_{M+1}(u_x - f_x\lambda^z_{M+1}u(z)) + \rho f_x + p^T_{M+1}(v_x - f_x\lambda^z_{M+1}v(z))\\ &= a^T_{M+1}u_x + p^T_{M+1}v_x + f_x(\rho - a^T_{M+1}\lambda^z_{M+1}u(z) - p^T_{M+1}\lambda^z_{M+1}v(z))\end{aligned}$$

and

$$\begin{aligned}f_x &= \frac{g_x - a^T_{M+1}u_x - p^T_{M+1}v_x}{\rho - a^T_{M+1}\lambda^w_{M+1}u(w) - p^T_{M+1}\lambda^w_{M+1}v(w)}\\ &= \frac{\lambda^z_{M+1}(\Phi(x,z) - \Phi^T_z u_x - p(z)^T v_x)}{\lambda^z_{M+1}(\lambda^w_{M+1}\Phi(w,z) - \Phi^T_z\lambda^w_{M+1}u(w) - p^T_z\lambda^w_{M+1}v(w))}\\ &= \frac{\lambda^z_{M+1}P(x,z)}{\lambda^z_{M+1}\lambda^w_{M+1}(\Phi(w,z) - \Phi^T_z u(w) - p^T_z v(w))}\\ &= \frac{\lambda^z_{M+1}P(x,z)}{\lambda^z_{M+1}\lambda^w_{M+1}P(w,z)}.\end{aligned}$$

We still have to evaluate $\tilde{P}$ via

$$\begin{aligned}\tilde{P}(x,y) &= \Phi(x,y) - (\tilde{u}^T_x, f_x)\begin{pmatrix}\Phi_y\\ g_y\end{pmatrix} - \tilde{v}^T_x p_y\\ &= \Phi(x,y) - \tilde{u}^T_x\Phi_y - f_x g_y - \tilde{v}^T_x p_y\\ &= \Phi(x,y) - (u_x - f_x\lambda^z_{M+1}u(z))\Phi_y - f_x g_y\\ &\quad -(v_x - f_x\lambda^z_{M+1}v(z))^T p_y\\ &= P(x,y) - f_x(g_y - \lambda^z_{M+1}u(z)^T\Phi_y - \lambda^z_{M+1}v(z)^T p_y)\\ &= P(x,y) - f_x\lambda^z_{M+1}(\Phi(y,z) - u(z)^T\Phi_y - v(z)^T p_y)\\ &= P(x,y) - f_x\cdot\lambda^z_{M+1}P(y,z)\end{aligned}$$

to prove the assertion. Note that $\lambda_{M+1} \notin \operatorname{span}\Lambda$ implies

$$\lambda^x_{M+1}\lambda^x_{M+1}P(x,x) = \|\lambda^x_{M+1}\delta_{x,u}\|^2_\Phi = \|\lambda_{M+1} - \sum_{j=1}^M \lambda^z_{M+1}u_j(z)\lambda_j\|^2_\Phi > 0.$$

We now use the recursion (4.4.2, *4p11*) of P to construct a sequence of orthogonal functions that serve to solve the interpolation problem directly. To this end we now use the index M to indicate quantities that depend on $\Lambda_M = \{\lambda_1, \dots, \lambda_M\}$, and we assume that $\Lambda_Q \subset \lambda_{Q+1} \subset \dots$ is a strictly increasing sequence of $\mathcal{P}$–nondegenerate sets of functionals. Then we use (4.4.2, *4p11*) in the form

(*4p41*)

$$P_{M+1}(x,y) = P_M(x,y) - \frac{\lambda^z_{M+1}P_M(x,z)\cdot\lambda^z_{M+1}P_M(y,z)}{\lambda^u_{M+1}\lambda^v_{M+1}P_M(u,v)} \tag{4.4.4}$$

for $M \geq Q$ and $x,y \in \Omega$. The recursion starts with P_Q, which in case $M > 0$ is associated to a set Λ_Q on which interpolation by functions from $\mathcal{P}$ is uniquely possible. On such a set the reconstruction takes place within $\mathcal{P}$, and P_Q coincides with the normalization of Φ. For $m = 0$ the functions from $\mathcal{P}$ are not present at all, and we formally use $Q = 0$ and

$$P_0(x,y) := \Phi(x,y).$$

This established the start of the recursion (4.4.4, *4p41*), and we now define functions

$$\begin{aligned} r_{M+1}(x) &:= \lambda^z_{M+1}P_M(x,z) \\ s_{M+1}(x) &:= \frac{r_{M+1}(x)}{\sqrt{\lambda^u_{M+1}\lambda^v_{M+1}P_M(u,v)}} \\ u_{M+1}(x) &:= \frac{r_{M+1}(x)}{\lambda^u_{M+1}\lambda^v_{M+1}P_M(u,v)}\,. \end{aligned}$$

Clearly

$$\begin{aligned} r_{M+1}, s_{M+1}, u_{M+1} &\in S_{M+1} \\ \lambda^x_j r_{M+1}(x) &= \{0\},\ 1 \leq j \leq M \end{aligned}$$

hold, and thus Theorem 3.1.36 (*OrtTh)* implies

$$(r_N, r_M)_\Phi = 0, \qquad Q < N < M.$$

The different normalization of s_{M+1} and u_{M+1} yield

$$\begin{aligned} \|s_{M+1}\|_\Phi &= 1 \\ u_{M+1}(x_{M+1}) &= 1 \end{aligned}$$

to generate orthonormal and Lagrange–type functions.

4.4.2 Recursion of Power Functions

(*SecRecPFu*) The recursion (4.4.2, *4p11*) leads to the formula

(*4p11a*)

$$P_{M+1}^2(x) = P_M^2(x) - s_{M+1}^2(x) \tag{4.4.5}$$

for all $x \in \Omega$, relating two subsequent (squared) power functions. If P_M^2 is a measure of the "energy" of the error for recovery based on M pieces of information, the addition of a new functional λ_{M+1} takes just s_{M+1}^2 ot of the energy. If the basis function Φ has a sharp localization, this will not necessarily lead to a decrease in the L_∞ norm of P_M^2. The inherent L_2 structure of this decomposition of the power function rather suggests to pick λ_{M+1} to maximize the expression

$$\int_\Omega s_{M+1}^2(x)dx = \int_\Omega \frac{\lambda_{M+1}^z P_M^2(x,z)}{\lambda_{M+1}^u \lambda_{M+1}^v P_M(u,v)} dx.$$

Since P_M is continuous and Ω is compact, this extremum exists, though it will be hard to calculate. Anyway, there are lots of interesting research problems opened up by these recursive techniques.

4.4.3 Newton's Formula

We now write the reconstruction g_M of some function f based on data $f_k = \lambda_k(f)$ for functionals from Λ_M in terms of the orthogonal functions $r_{Q+1}, \ldots, r_M$ as

$$g_M(\cdot) = g_Q(\cdot) + \sum_{j=Q+1}^{M} \beta_j r_j(\cdot).$$

This is a **Newton–type interpolation formula**, and we can calculate the generalized divided differences β_j by a simple recursion. In fact, for any $k > Q$ we have

$$f_k = \lambda_k(g_Q) + \sum_{j=Q+1}^{k} \beta_j \lambda_k(r_j)$$

and get the recursions

$$\begin{aligned} \beta_k &= \frac{1}{\lambda_k(r_k)} \left(f_k - \lambda_k(g_Q) - \sum_{j=Q+1}^{k-1} \beta_j \lambda_k(r_j) \right) \\ &= \frac{1}{\lambda_k(r_k)} \left(f_k - \lambda_k(g_{k-1}) \right). \end{aligned}$$

Unfortunately, the recursive method based on Newton's formula is not particularly effective. One could rewrite the formula in terms of the functions u_j to avoid the denominators, but this is no serious improvement.

4.4.4 Kernel Expansion

(*SecHSE*) If we carry the above method out for $M \to \infty$ on a domain $\Omega \subset I\!R^d$ with infinitely many points, we clearly have a pointwise decrease of the functions

$$\ldots \geq P_M^2(x,x) \geq P_{M+1}^2(x,x) \geq \ldots \geq \lim_{M\to\infty} P_M^2(x,x) =: P_\infty^2(x,x) \geq 0$$

for all $x \in \Omega$. If we are in the special case of Lagrange interpolation, where $\lambda_j(f) = f(x_j)$, we can let the points x_M gradually get dense in Ω for $M \to \infty$. Then we can expect that the L_∞ norm of P_M^2 on Ω decreases to zero for $M \to \infty$. Section 5 (*SecEB)* will contain a variety of such results, and uniform Lipschitz continuity of Φ along the diagonal will usually be sufficient (see 5.5 (*hrhodef)* on page 120).

In such cases we get a **kernel expansion**

$$P_Q(x,y) = \sum_{j=Q+1}^{\infty} s_j(x)s_j(y)$$

for the normalized basis function P_Q. This specializes to the series

$$P_Q^2(x) = P_Q(x,x) = \sum_{j=Q+1}^{\infty} s_j(x)^2$$

for the squares of power functions. There are lots of highly interesting open problems along this line of research.

4.4.5 Remarks

Most of the material on power functions as presented in this section seems to be new, though there is some earlier work on recursive constructions of interpolants (see e.g. Mühlbach [3](*muehlbach:??-??*))

4.5 Condition

(*SecCondition*) We now look at the stability of solutions of the systems (1.7.2, *EQsys2*) and (3.1.14, *EQsys3*) written in matrix form as

(*BDef3*)

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}. \tag{4.5.1}$$

which is exactly (1.7.3, *BDef*) or (3.1.18, *BDef2*), but repeated here for convenience. Introducing perturbations of the solution and the right-hand side we get the system

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \alpha + \Delta\alpha \\ \beta + \Delta\beta \end{pmatrix} = \begin{pmatrix} f + \Delta f \\ 0 \end{pmatrix}$$

and can subtract (4.5.1, *BDef3*) to get

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \Delta\alpha \\ \Delta\beta \end{pmatrix} = \begin{pmatrix} \Delta f \\ 0 \end{pmatrix}.$$

This implies

(*DAD*)

$$(\Delta\alpha)^T A(\Delta\alpha) = (\Delta\alpha)^T \Delta f. \tag{4.5.2}$$

Since we have $P^T(\Delta\alpha) = 0$, we know that the above quadratic form is positive definite. Thus there are positive real eigenvalues σ and Σ of the matrix A such that

(*Defsigma*)

$$0 < \sigma := \inf \frac{\alpha^T A\alpha}{\alpha^T\alpha} \leq \sup \frac{\alpha^T A\alpha}{\alpha^T\alpha} =: \Sigma < \infty, \tag{4.5.3}$$

where the sup and inf are extended over all $\alpha \in I\!R^M$ which are nonzero but satisfy $P^T\alpha = 0$. The **condition number** $\kappa(A)$ of A in the Euclidean norm then is the quotient $\kappa(A) = \Sigma/\sigma$, and it appears in the bound

$$\frac{\|\Delta\alpha\|_2}{\|\alpha\|_2} \leq \kappa(A)\frac{\|\Delta f\|_2}{\|f\|_2}$$

that follows from (4.5.2, *DAD*)and the corresponding equation

$$\alpha^T A\alpha = \alpha^T f$$

for the unperturbed quantities. This bound holds for the relative error, while the absolute error is governed by

(*Stab*)

$$\|\Delta\alpha\|_2 \leq \frac{1}{\sigma}\|\Delta f\|_2. \tag{4.5.4}$$

Numerical experiments show that σ can indeed be extremely small, while Σ does not grow too wildly, at least not as wildly as $1/\sigma$. Later theoretical results will support these statements, and thus the study of σ or some lower bounds for it will be of great importance for any assessment of the numerical stability of systems like (4.5.1, *BDef3*).

4.5.1 Remarks

The technique for proving error bounds via power functions goes at least back to Golomb and Weinberger [16](*golomb-weinberger:59-1*) but probably further back to Peano, since the error evaluation of linear functionals by bounding their Peano kernels is very similar. The pointwise optimality principle of Theorem 4.3.11 (*OPFT)* was used by various authors and possibly dates back to Duchon [10](*duchon:76-1*).

4.6 Uncertainty Relation

(*URT*) It would be very desirable to have recovery methods with small errors and good stability. However, these two goals cannot be met at the same time, since there is a connection between them that implies bad stability whenever the a-priori error bound is very small.

Let us look at this connection in a fairly general way. If we try optimal recovery of a function $g \in \mathcal{G}$ from data $\lambda_j(g), 1 \leq j \leq M$ in the setting of section 3.1.1 (*subsecORP)* and bound the error by Theorem 3.1.38 (*ORTFA)* on page 41, then we have to study the generalized optimal power function $P(\mu)$ of (3.1.40, *GPDef*), whose square has the representation (3.1.29, *BAPN*). But this quantity can be written as a value of the quadratic form associated to the matrix

$$A_{\mu,\Lambda} = \begin{pmatrix} (\mu,\mu)_\Phi & (\mu,\lambda_1)_\Phi & \dots & (\mu,\lambda_M)_\Phi \\ (\lambda_1,\mu)_\Phi & (\lambda_1,\lambda_1)_\Phi & \dots & (\lambda_1,\lambda_M)_\Phi \\ \vdots & \vdots & & \vdots \\ (\lambda_M,\mu)_\Phi & (\lambda_M,\lambda_1)_\Phi & \dots & (\lambda_M,\lambda_M)_\Phi \end{pmatrix}$$

with the vector $(1, -\alpha_1(\mu), \dots, -\alpha_M(\mu))^T \in I\!R^{M+1}$. This yields a lower bound

(*UR1*)

$$P^2(\mu) \geq \sigma(A_{\mu,\Lambda}) \left(1 + \|\alpha(\mu)\|_2^2\right) \geq \sigma(A_{\mu,\Lambda}) \qquad (4.6.1)$$

for the power function in terms of the smallest eigenvalue of the matrix. This relates the error analysis to the stability analysis and provides the background for various cases of the Uncertainty Relation. Furthermore, it sets the direction for further progress: we need upper bounds for the power function P and positive lower bounds for the smallest eigenvalue σ. But we should be aware that the two sides of (4.6.1, *UR1*) behave differently as functions of Λ: the right-hand side will vanish, but not the left-hand side, if two functionals from Λ come too close to each other.

4.6.1 The Lagrange Case

We now specialize to the setting of Theorem 4.3.11 (*OPFT)* on page 91 with $X = \{x_1, \ldots, x_M\} \subset I\!R^d$ and $\Lambda = \{\delta_{x_1}, \ldots, \delta_{x_M}\}$. Then we have the matrix

$$A_{x,X} = \begin{pmatrix} \Phi(x,x) & \Phi(x,x_1) & \ldots & \Phi(x,x_M) \\ \Phi(x_1,x) & \Phi(x_1,x_1) & \ldots & \Phi(x_1,x_M) \\ \vdots & \vdots & & \vdots \\ \Phi(x_M,x) & \Phi(x_M,x_1) & \ldots & \Phi(x_M,x_M) \end{pmatrix}$$

and the vector $(1, -u_1^*(x), \ldots, -u_M^*(x))^T \in I\!R^{M+1}$ and get the special form (*UR2*)

$$P_{u^*}^2(x) = P_\Lambda^2(\delta_x) \geq \sigma(A_{x,X}) \left(1 + \sum_{j=1}^{M} |u_j^*(x)|^2\right) \geq \sigma(A_{x,X}) \qquad (4.6.2)$$

of (4.6.1, *UR1*). Note that both sides are continuous functions of x and X (or Λ standing for X) that vanish whenever x tends to points in X.

We now can give some hints to the results that follow in later sections. The Uncertainty Relation in the form (4.6.2, *UR2*) suggests to bound P^2 from above and σ from below, in order to have both upper bounds on the attainable error and on the numerical stability, measured by $1/\sigma$ due to (4.5.4, *Stab*). We shall see in 5.1 (*SecUBOPF)* that upper bounds for P^2 take the form (*FBound*)

$$P_{u^*}^2(x) \leq F(h_{X,\Omega}) \text{ for all } x \in \Omega \qquad (4.6.3)$$

where F is a monotonic function of the fill distance $h_{X,\Omega}$ defined in (2.1.2, *DDDef*) on page 18. On the other hand, the lower bounds for σ in 7.4 (*SecLBE)* will be of the form (*GBound*)

$$\sigma(A_X) \geq G(s_X) \text{ for all } X = \{x_1, \ldots, x_M\} \subset \Omega \qquad (4.6.4)$$

with the separation distance s_X defined in (2.1.1, *SDDef*). For gridded data on $\epsilon Z\!\!Z^d \cap \Omega$ we can roughly expect $h_{X,\Omega} = s_X\sqrt{d}$, and then the Uncertainty Relation necessarily implies (*UR3*)

$$F(t\sqrt{d}) \geq G(t) \qquad (4.6.5)$$

for all $t \geq 0$. This allows to check the quality of the bounds (4.6.3, *FBound*) and (4.6.4, *GBound*), since the lowest possible bounds F and the largest possible bounds G must necessarily satisfy (4.6.5, *UR3*) and are optimal, if they turn (4.6.5, *UR3*) into an equality. This opens the race for optimal bounds of the form (4.6.3, *FBound*) and (4.6.4, *GBound*), and this text will describe the current state-of-the-art.

4.6.2 Remarks

The Uncertainty Relation seems to occur first in [41](*schaback:95-1*).

5 Error Bounds

(*SecEB*)

5.1 Upper Bounds for the Optimal Power Function

(*SecUBOPF*) Here we proceed to prove upper bounds of the form (4.6.3, *FBound*) for the optimal power function of optimal recovery. This approach uses results from classical approximation theory and does not require Fourier transforms. Another proof technique, using transforms, will follow in section 6.5 (*SecEBTrans*).

5.1.1 Assumptions and First Results

We specialize here to the case of multivariate Lagrange interpolation by conditionally positive definite functions $\Phi \; : \; \Omega \times \Omega \to I\!R$ of order m on some domain Ω that can be embedded into $I\!R^d$. The data locations are supposed to form a $I\!P_m^d$-nondegenerate set $X = \{x_1, \ldots, x_M\} \subset \Omega$, and we use functions u_j on Ω with (4.3.6, *GenRec*) that reproduce $I\!P_m^d$.

The power function with respect to these data and the functions u_j takes the special form

(*DefPuxyLag*)

$$\begin{aligned} P_u(x)^2 \;\; := \;\; & \Phi(x,x) - 2\sum_{j=1}^{M} \Phi(x,x_j)u_j(x) \\ & + \sum_{j,k=1}^{M} \Phi(x_j,x_k)u_j(x)u_k(x) \end{aligned} \tag{5.1.1}$$

from (4.3.7, *DefPuxy*). Note that we allow quite arbitrary u_j here in view of Theorem 4.3.11 (*OPFT*). If optimal recovery leads to Lagrange basis functions u_j^*, $1 \leq j \leq M$, then

$$P_{u^*}(x) \leq P_u(x)$$

holds for all $x \in \Omega$, yielding a pointwise upper bound for the optimal power function.

To start with, we fix a polynomial order $\ell \geq m$ and a point $x \in \Omega$. Around x we shall approximate Φ by a polynomial φ in the following sense:

Assumption 5.1.2 *(FBAss1) For each $x \in \Omega$ and a specific choice of a polynomial order*

(EqEllgeqm)

$$\ell \geq m \tag{5.1.3}$$

there are positive constants ρ, h_0, and C_1 and a polynomial $\varphi : I\!R^d \times I\!R^d \to I\!R$ of order not exceeding ℓ in each d-variate variable, such that

(PhiApp)

$$\|\Phi(x+u, x+v) - \varphi(u,v)\|_\infty \leq C_1 h^\rho \tag{5.1.4}$$

for all $h \in [0, h_0]$ and all $u, v \in [0, h]^d$.

We shall vary x and ℓ later, and then all of the above quantities will be studied as functions of x and ℓ. Equation (5.1.4, *PhiApp*) may be viewed as resulting from a Taylor expansion around (x, x) or by an L_∞ approximation process. It is no drawback to assume symmetry of φ in the sense $\varphi(x, y) = \varphi(y, x)$, because the arithmetic mean of these two polynomials will do the job.

We now define a function Q_u^2 that serves as a polynomial approximation to P_u^2 near x, but which will turn out to be zero later:

(*DefQuxyLag*)

$$\begin{aligned} Q_u(x)^2 \quad := \quad & \varphi(0,0) - 2\sum_{j=1}^{M} \varphi(0, x_j - x) u_j(x) \\ & + \sum_{j,k=1}^{M} \varphi(x_j - x, x_k - x) u_j(x) u_k(x). \end{aligned} \tag{5.1.5}$$

Now it is time to specify our choice of u_j, $1 \leq j \leq M$ via local polynomial reproduction of order ℓ near x. Since the dependence on x and h is crucial here, we stick to an explicit notation:

Assumption 5.1.6 *(FBAss2) For some $x \in \Omega$ and some $h \in [0, h_0]$ there is a subset $J_x(h)$ of $\{1, \ldots, M\}$, positive constants $C_2(x,h), C_3(x,h)$, and a choice of M real numbers $u_1^h(x), \ldots, u_M^h(x)$ such that*

(uDefJx)

$$J_x(h) \subseteq \{\, j \;:\; 1 \le j \le M,\ \|x - x_j\|_\infty \le C_2(x,h)h\}, \tag{5.1.7}$$

(uDef1)

$$u_j^h(x) = 0 \textit{ for all } j \notin J_x(h), \tag{5.1.8}$$

(uDef2)

$$p(x) = \sum_{j \in J_x(h)} u_j^h(x) p(x_j) \textit{ for all } p \in I\!P_\ell^d, \tag{5.1.9}$$

(uDef3)

$$1 + \sum_{j \in J_x(h)} |u_j^h(x)| \le C_3(x,h). \tag{5.1.10}$$

The first three items specify the local polynomial reproduction, while the last defines C_3 to be the corresponding Lebesgue constant. We apply (5.1.9, *uDef2*) to $\varphi(0, y - x)$ as a function of y to get

$$\varphi(0, x - x) = \varphi(0,0) = \sum_{j \in J_x(h)} u_j^h(x) \varphi(0, x_j - x)$$

to prove that Q_u is identically zero:

$$\begin{aligned} Q_u(x)^2 &= \varphi(0,0) - 2\varphi(0,0) + \sum_{j=1}^{M} \varphi(x_j - x, 0) u_j(x) \\ &= 0. \end{aligned}$$

We now bound the optimal power function by

(*FundBound*)

$$\begin{aligned} P_{u^*}^2(x) &\le P_u^2(x) \\ &= P_u^2(x) - Q_u^2(x) \\ &= \Phi(x,x) - \varphi(0,0) \\ &\quad -2 \sum_{j \in J_x(h)} u_j^h(x) \left(\Phi(x, x_j) - \varphi(0, x_j - x)\right) \\ &\quad + \sum_{j,k \in J_x(h)} u_j^h(x) u_k^h(x) \left(\Phi(x_j, x_k) - \varphi(x_j - x, x_k - x)\right) \qquad (5.1.11) \\ &\le \left(1 + \sum_{j \in J_x(h)} |u_j^h(x)|\right)^2 C_1(x) (C_2(x,h)h)^\rho \\ &\le C_3(x,h)^2 C_1(x) C_2^\rho(x,h) h^\rho \end{aligned}$$

for all h with $C_2(x,h)h \leq h_0$, where we have to keep in mind that everything still depends on ℓ. Nevertheless (5.1.11, *FundBound*) is the fundamental error bound for optimal power functions, and it can be applied in a large number of cases. We summarize:

Theorem 5.1.12 *(FundBoundT) Under the assumptions 5.1.2 (FBAss1) and 5.1.6 (FBAss2) the optimal power function has a local bound of order $\rho/2$ in x with respect to $h \to 0$, if the constants $C_2(x,h)$, $C_3(x,h)$ are bounded for $h \to 0$.* □

The applications of Theorem 5.1.12 (*FundBoundT*) come in two variations:

- To prove a **fixed** error order ρ, one fixes an appropriate ℓ and uses compactness arguments to bound all relevant "constants" with respect to x and h.
- To prove very strong non-polynomial error bounds like e^{-c/h^2} for fixed-scale Gaussians, one has to let ℓ tend to ∞ and study the variation of the "constants" with ℓ. This is a much harder task.

The two assumptions 5.1.2 (*FBAss1*) and 5.1.6 (*FBAss2*) require two different kinds of results to be proven in the following sections:

- an error bound for local polynomial approximation of Φ,
- and bounds on the Lebesgue constant for local polynomial interpolation in Ω.

5.2 Approximation Error in the Radial Case

(*SecAERC*) Here we consider the special situation of d-variate radial functions $\Phi(x,y) = \phi(\|x-y\|_2)$, and we want to check Assumption 5.1.2 (*FBAss1*). The crucial term in (5.1.4, *PhiApp*) takes the form $\Phi(x+u, x+v) = \phi(\|u-v\|_2)$ and usually will not be nicely expandable into a polynomial in u and v. Fortunately, it is independent of x, since we are in a translation-invariant situation, and we only need an approximation to ϕ near zero. More precisely, we approximate $\phi(r)$ by a polynomial $p_n \in I\!P_n^1$ in r^2 on the domain $[0,h]$ for small $h > 0$ and define the error as

(*EDef1*)

$$\begin{aligned} E_n(\phi,h) &:= \inf_{p\in I\!P_n^1} \|\phi(r) - p(r^2)\|_{\infty,[0,h]} \\ &= \inf_{p\in I\!P_n^1} \|\phi(\sqrt{r}) - p(r)\|_{\infty,[0,h^2]}. \end{aligned} \tag{5.2.1}$$

This error can be bounded by univariate Jackson type theorems from classical approximation theory. Less sophisticated bounds simply take p as the Taylor expansion of $\phi(\sqrt{\cdot})$ in zero. With $\ell = 2n-1$ and h replaced by $2\sqrt{d}h$ one can put the result into Assumption 5.1.2 (*FBAss1)*.

Let us evaluate a few cases by standard techniques and cite the stronger Jackson results from the literature later.

Example 5.2.2 (AEPHS) *In the polyharmonic spline case $\phi(r) = r^\beta$ with $\beta \in I\!\!R_{>0}\backslash 2I\!\!N$ we can simply rescale the approximation problem to the interval $[0,1]$. That is,*

$$
\begin{aligned}
E_n(r^\beta, h) &= \inf_{p\in I\!\!P_n^1} \|r^{\beta/2} - p(r)\|_{\infty,[0,h^2]} \\
&= \inf_{p\in I\!\!P_n^1} \|(h^2 s)^{\beta/2} - p(h^2 s)\|_{\infty,[0,1]} \\
&= h^\beta \inf_{p\in I\!\!P_n^1} \|s^{\beta/2} - h^{-\beta} p(h^2 s)\|_{\infty,[0,1]} \\
&= h^\beta E_n(r^\beta, 1).
\end{aligned}
$$

*This yields the exact dependence on h and leaves the dependence on β to a classical Jackson result on $[0,1]$. We get $\rho = \beta$, and this is independent of $\ell = 2n-1$, provided that $\ell = 2n-1 \geq m \geq \lceil \beta/2 \rceil$ holds, since we have to exceed the order m of conditional positive definiteness. The most important cases $\beta = 1$ and $\beta = 3$ have the bounds $E_1(r,1) = 1/2$ and $E_2(r^3,1) = 2/27$, and these are available by direct analysis, using the Alternation Theorem of linear univariate L_∞ approximation. For this, see any textbook on Approximation Theory, e.g.: the classical books by Cheney [9](cheney:66-1) and Meinardus [25](*meinardus:67-1*).*

Example 5.2.3 (AETPS) *Now consider the classsical thin-plate spline $\phi(r) = r^\beta \log r$ with $\beta \in 2I\!\!N$ and order $m > \beta/2$ of conditional positive definiteness. We proceed along the same lines and need $\ell = 2n-1 \geq m > \beta/2$. This implies $\beta/2 \leq n-1$, which is useful to get rid of the* log *term in*

$$
\begin{aligned}
E_n(r^\beta \log r, h) &= \inf_{p\in I\!\!P_n^1} \|\frac{1}{2} r^{\beta/2} \log r - p(r)\|_{\infty,[0,h^2]} \\
&= \inf_{p\in I\!\!P_n^1} \|\frac{1}{2} (h^2 s)^{\beta/2} \log(h^2 s) - p(h^2 s)\|_{\infty,[0,1]} \\
&= \inf_{p\in I\!\!P_n^1} \|\frac{1}{2} (h^2 s)^{\beta/2} (\log(h^2) + \log s) - p(h^2 s)\|_{\infty,[0,1]} \\
&= \inf_{p\in I\!\!P_n^1} \|\frac{1}{2} (h^2 s)^{\beta/2} \log s - p(h^2 s)\|_{\infty,[0,1]} \\
&= \tfrac{1}{2} h^\beta \inf_{p\in I\!\!P_n^1} \|s^{\beta/2} \log s - h^{-\beta} p(h^2 s)\|_{\infty,[0,1]} \\
&= h^\beta E_n(r^\beta \log r, 1).
\end{aligned}
$$

The case $\beta = 2$ has $E_2(r^2 \log r, 1) = e^{-1}$.

Example 5.2.4 *(AEWF) Here we treat Wendland's [46](wendland:95-1) function $\phi(r) = (1-r)_+^4(1+4r)$ which is positive definite on $I\!R^d$ for $d \leq 3$ and in $C^2(I\!R^d)$ if $r = \|x\|_2$ for $x \in I\!R^d$. But our approach will be applicable to the whole class of piecewise polynomial functions of the form*

$$\phi(r) = \left\{ \begin{array}{rl} u(r^2) + r^{2n-1}v(r) & r \in [0,1] \\ 0 & else \end{array} \right\},$$

where we pick a maximal n such that u lies in $I\!P_n^1$ and v is an arbitrary univariate polynomial with $v(0) \neq 0$. This means that u covers the first terms of even degree, while r^{2n-1} is the first term of odd degree. This includes the full range of Wendland's functions from [46](wendland:95-1) as well as Wu's functions from [47](wu:95-2) for certain values of n. In case of $\phi(r) = (1-r)_+^4(1+4r)$ we have $\phi(r) = 1 - 10r^2 + r^3(20 - 15r + 4r^2)$ with $n = 2$. We now use u as an approximation to ϕ on small intervals. In particular,

$$\begin{array}{rcl} E_n(\phi, h) & = & \inf_{p \in I\!P_n^1} \|\phi(r) - p(r^2)\|_{\infty,[0,h]} \\ & \leq & \|r^{2n-1}v(r)\|_{\infty,[0,h]} \\ & \leq & C_5 h^{2n-1} \end{array}$$

for $h \in [0,1]$ with a suitable constant C_5 depending on v, e.g.: $C_5 := \|v\|_{\infty,[0,1]}$. Note that for the function $\phi(r) = (1-r)_+^4(1+4r)$ we get the same order as for the polyharmonic spline $\phi(r) = r^3$.

So far we did not use sophisticated theorems from approximation theory, since we were interested in the correct power of h, not in the optimal behaviour of the bounds with respect to ℓ or n.

In the previous cases it did not make much sense to let ℓ or n be too large, because the approximation order with respect to h is not improved, and because we see later that large values of ℓ lead to bad Lebesgue constants when heading for Assumption (5.1.6, *FBAss2*). But the next case will be different in nature:

Example 5.2.5 *(AEGEl) The Gaussian $\phi(r) = \exp(-\alpha r^2)$ allows arbitrary values of $\ell = 2n-1$ because it is unconditionally positive definite. A crude bound is provided by chopping the exponential series:*

$$\begin{array}{rcl} E_n(\exp(-\alpha r^2, h) & = & \inf_{p \in I\!P_n^1} \|\exp(-\alpha r) - p(r)\|_{\infty,[0,h^2]} \\ & = & \inf_{p \in I\!P_n^1} \|\exp(-s) - p(s/\alpha)\|_{\infty,[0,\alpha h^2]} \\ & \leq & \frac{(\alpha h^2)^n}{n!} \end{array}$$

for $\alpha h^2 \leq n+1$, which is not a serious restriction. By using the Taylor residual one can get rid of the restriction, and by Bernstein's theorem 5.3.4 (BT1) on approximation of analytic functions we can get a similar bound that decays exponentially with $n \to \infty$. Anyway, we see that the bound improves dramatically with increasing n or $\ell = 2n-1$.

5.3 Jackson-Bernstein Theorems and Applications

This section contains the results from Approximation Theory that make the previous results somewhat sharper. We stick to radial functions and use univariate techniques. These consist of Jackson type theorems for the best approximation of functions $f \in C^n[a,b]$ by univariate polynomials in $I\!P^1_\ell$ in the supremum or Chebyshev or L_∞ norm:

(*EDef2*)

$$E(\ell, f, [a,b]) := \inf_{p \in I\!P^1_\ell} \|f-p\|_{\infty,[a,b]} \tag{5.3.1}$$

We additionally need the notion of **Lipschitz continuity**:

Definition 5.3.2 *A function f is Lipschitz continuous on $[a,b]$ of order $\alpha \in [0,1]$ with Lipschitz constant L, if*

$$|f(x) - f(y)| \leq L|x-y|^\alpha$$

holds for all $x, y \in [a,b]$.

Theorem 5.3.3 *(JT1) For all functions $f \in C^n[a,b]$ and all $\ell \geq n$ we have*

$$E(\ell, f, [a,b]) \leq \left(\frac{\pi}{4}\right)^n \frac{(b-a)^n}{(\ell+1)\ell\ldots(\ell-n+2)} \|f^{(n)}\|_\infty.$$

If $f^{(n)}$ is Lipschitz continuous of order α with Lipschitz constant L, then

$$E(\ell, f, [a,b]) \leq \left(\frac{\pi}{4}\right)^{n+1} \frac{(n+1)^n}{n!} \left(\frac{b-a}{\ell}\right)^{n+\alpha} L.$$

These results of Jackson (see e.g. Cheney [9](*cheney:66-1*) or Meinardus [25](*meinardus:67-1*)) yield bounds in terms of fixed negative powers of ℓ that depend on the smoothness of f. They can be proven to be optimal. For analytic functions, however, the parameter ℓ moves into the exponent of some quantity that is smaller than one, and this yields a much better asymptotic behaviour for $\ell \to \infty$ due to Bernstein (this is, for instance, in Natanson [36](*natanson:55-1*)) :

Theorem 5.3.4 *(BT1) Let f be a function on $[a,b]$ which has a holomorphic continuation into an ellipse in $\mathbb{C}$ with foci a,b and half-axes of length $0 < r \le R$. Then there is a constant K depending only on f, r, and R, but not on ℓ, such that*

$$E(\ell, f, [a,b]) \le K \left(\frac{b-a}{2(r+R)}\right)^{\ell}.$$

We cannot give proofs here, but the following weaker and easily accessible result shows how the previous result is possible.

Theorem 5.3.5 *(RSJT) Let f be a function on $[-r,r]$ which has a holomorphic continuation into the circle C_R in $\mathbb{C}$ with radius $R > r$ such that the continuation still is bounded on the boundary ∂C_R of the circle. Then*

$$E(\ell, f, [-r,r]) \le \|f\|_{\infty,\partial C_R} \frac{R}{R-r}\left(\frac{r}{R}\right)^{\ell},$$

and the bound is already achieved by the Taylor expansion around zero.

Proof: Just consider the power series of f in zero and bound it using Cauchy's inequality

$$|a_n| \le \|f\|_{\infty,\partial C_R} R^{-n}$$

for the coefficients. This yields

$$\begin{aligned} |f(z) - \sum_{j=0}^{\ell-1} a_j z^j| &= |\sum_{j=\ell}^{\infty} a_j z^j| \\ &\le \sum_{j=\ell}^{\infty} |a_j| r^j \\ &\le \|f\|_{\infty,\partial C_R} \sum_{j=\ell}^{\infty} \left(\frac{r}{R}\right)^j \\ &\le \|f\|_{\infty,\partial C_R} \left(\frac{r}{R}\right)^{\ell} \frac{R}{R-r} \end{aligned}$$

for all $|z| \le r$. □

We now work our way through the examples.

Example 5.3.6 *(AEPHS2) Consider thin-plate splines $\phi(r) = r^\beta$. These are conditionally positive definite of order $m \ge m_\beta := \lceil \frac{\beta}{2} \rceil$. We have to approximate $r^{\beta/2}$ on $[0, h^2]$ and do this directly by application of Jackson's* *theorem 5.3.3 (JT1). The function* $r^{\beta/2}$ *has* $m_\beta - 1$ *continuous derivatives, and the final derivative is Lipschitz continuous of order*

$$\alpha_\beta := \frac{\beta}{2} - m_\beta + 1 = \frac{\beta}{2} - \lfloor \frac{\beta}{2} \rfloor \in (0,1)$$

with constant

$$L_\beta = \frac{\beta}{2}(\frac{\beta}{2} - 1)\ldots(\frac{\beta}{2} - m_\beta) = (1+\alpha_\beta)(2+\alpha_\beta)\ldots(m_\beta - 1 + \alpha_\beta) \le m_\beta!.$$

Then the two slightly different notions of (5.2.1, EDef1) and (5.3.1, EDef2), which are related by the transformation $r \mapsto \sqrt{r}$ *in the argument of the function, come out to be*

$$E_n(r_\beta, h) = E(n, r^{\beta/2}, [0, h^2]) \le \left(\frac{\pi}{4}\right)^{m_\beta} \frac{(m_\beta)^{m_\beta - 1}}{(m_\beta - 1)!} \left(\frac{h^2}{n}\right)^{\beta/2} L_\beta$$

for all $\ell = 2n - 1 \ge m \ge m_\beta$*. The result has the same power of* h *as before, but now we can quantify the dependence on* β *and* n*. Unfortunately, the gain for large* n *or* $\ell = 2n-1$ *is much too weak to cope with the dramatic increase of Lebesgue constants for increasing polynomial degrees.*

Example 5.3.7 *(AETPS2) We now continue with Example 5.2.3 (AETPS). The radial function* $\phi(r) = r^\beta \log r$ *with* $\beta \in 2\mathbb{N}$ *is conditionally positive definite of order* $m \ge m_\beta := \beta/2 + 1$*. We have to consider polynomial approximations to* $r^{\beta/2} \log r$ *for orders* n *satisfying* $\ell = 2n - 1 \ge m \ge m_\beta = \beta/2 + 1$*. The derivatives of* $r^{\beta/2} \log r$ *for* $\beta \in 2\mathbb{N}$ *produce lower-order polynomials of type* $r^{\beta/2-1}, r^{\beta/2-2}, \ldots$ *which are subsumed in* $\mathbb{P}_n^1$ *and do not change the approximation error. Thus we only have to consider the terms of type* $r^\alpha \log r$*, and we see that we can take* $\beta/2 - 1$ *continuous derivatives. The final derivative is* $(\beta/2)! r \log r$*, which is Lipschitz continuous of order* < 1*, but not of order 1. The direct application of the second version of Jackson's theorem 5.3.3 (JT1) would not give the full order with respect to* h *due to this fact, and therefore we first do the scaling of Example 5.2.3 (AETPS) to extract the factor* h^β *out of* $E_n(r^\beta \log r, h)$*. Then the first version of Jackson's theorem yields*

$$\begin{array}{rcl} E_n(r^\beta \log r, h) & = & h^\beta E_n(r^\beta \log r, 1) \\ & = & E(n, r^{\beta/2} \log r, [0,1]) \\ & \le & \left(\frac{\pi}{4}\right)^{\beta/2} \frac{(\beta/2)!}{(n+1)n(n-1)\ldots(n-\beta/2+2)} \|r \log r\|_{\infty,[0,1]} \\ & = & \left(\frac{\pi}{4}\right)^{\beta/2} \left(\begin{array}{c} n+1 \\ \beta/2 \end{array}\right)^{-1} e^{-1} \end{array}$$

for all $\ell = 2n - 1 \ge m \ge m_\beta = \beta/2 + 1$*. Again, we have some improvement for increasing* n*, but it will not be enough to cope with the Lebesgue constants.*

Example 5.3.8 *(AEMQ) We now consider multiquadrics $\phi(r) = (c^2 + r^2)^{\beta/2}$ for $\beta \notin 2I\!N$ and $c > 0$. In case of $\beta > 0$ they are conditionally positive definite of order $m \geq m_\beta := \lceil \beta/2 \rceil$, while they are positive definite for $\beta < 0$. In this case we define $m_\beta := 0$. Multiquadrics are analytic around $r = 0$ and their polynomial approximation can be treated by application of Bernstein's theorem 5.3.4 (BT1) or by Theorem 5.3.5 (RSJT). This means that we should study the complex function $f(z) = (c^2 + z)^{\beta/2}$ which has a singularity at $z = -c^2$. For $\beta > 0$ the function is bounded on the circle C_{c^2}, but for negative β (inverse multiquadrics) we have to use a smaller radius. To be safe, we use $R = c^2/2$ in both cases and get*

$$|f(z)| \leq (3c^2/2)^{\beta/2} \leq 2^{|\beta/2|} c^\beta$$

for $\beta > 0$ and $|z| = R$, while

$$|f(z)| \leq (c^2/2)^{\beta/2} = 2^{|\beta/2|} c^\beta$$

for $\beta < 0$ yields the same bound. We approximate on $[0, h^2]$ and thus have the constraint

$$h^2 < R = c^2/2$$

on what follows. Now Theorem 5.3.5 (RSJT) yields

(MQB1)

$$E_n(\phi, h) = E(n, f, [-h^2, h^2]) \leq 2^{|\beta/2|} c^\beta \frac{c^2}{c^2 - 2h^2} \left(\frac{2h^2}{c^2} \right)^n \qquad (5.3.9)$$

for all $\ell = 2n - 1 \geq m \geq m_\beta$.

Example 5.3.10 *(AES1) We consider Sobolew radial basis functions*

$$\phi(r) = r^\nu K_\nu(r)$$

for $\nu > 0$. These generate Sobolev spaces $W_2^m(I\!R^d)$ for $\nu = m - d/2$ and are unconditionally positive definite. A direct application of Jackson's theorems requires the derivatives of ϕ, which are not easy to calculate and bound from above. We postpone treatment of this case to section 6.5.7 (EBSob), where we apply Fourier transform techniques.

Theorem 5.3.11 *(LipConvTh) If $\Phi(x, y) = \phi(\|x - y\|_2)$ is a conditionally positive definite radial basis function on $\Omega \subset I\!R^d$ and if ϕ is Lipschitz continuous in a neighborhood of the origin, then Assumption 5.1.2 (FBAss1) is satisfied for some positive exponent ρ.*

Proof: Just apply Theorem 5.3.3 (*JT1*). □

5.4 Lebesgue Constants

(*SecLebCon*) We now face the verification of Assumption 5.1.6 (*FBAss2*), which is a very hard problem. Let us first discuss some easy cases.

5.4.1 Lines and Triangles

Assume that we want to prove a bound for the error in a point x that lies on a line between two distinct data points, say x_1 and x_2, and assume that the distance between these points is $2h$. We can define linear functions u_1, u_2 by

$$u_1(y) := \frac{(y - x_2)^T(x_1 - x_2)}{\|x_1 - x_2\|_2^2},\ u_2(y) := 1 - u_1(y)$$

and see that $u_j(x_k) = \delta_{jk},\ j,k = 1,2$. Any linear polynomial p restricted to the line through x_1 and x_2 is uniquely recovered by $p(x) = p(x_1)u_1(x) + p(x_2)u_2(x)$. Note that Assumption 5.1.6 (*FBAss2*) only requires the recovery in x, not everywhere. If x is way between x_1 and x_2, then clearly $C_3 = 2$ suffices, since both $u_1(x)$ and $u_2(x)$ are in $[0,1]$ and sum up to 1. Furthermore, we can set $C_2 = 1$ and are done for cases with $\ell \leq 2$. This argument works for every space dimension, but only on lines between two nearby data points.

We now go over to three points x_1, x_2, $x_3 \in I\!\!R^d$ forming a nondegenerate triangle T, and we consider points x inside such a triangle. If x lies on an edge, we are in the previous case. Since our argument is carried out in a two-dimensional affine subspace containing the triangle, we assume that we are in R^2 right away, and there are no problems going back to the embedded plane in $I\!\!R^d$. Nondegeneracy of the triangle, when written in bivariate coordinates, means that the system

$$\begin{pmatrix} x_1 & x_2 & x_3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} u_1(y) \\ u_2(y) \\ u_3(y) \end{pmatrix} = \begin{pmatrix} y \\ 1 \end{pmatrix}$$

has a nonsingular matrix and a unique solution. The components of the solution are called the **barycentric coordinates** of y with respect to the triangle spanned by x_1, x_2, x_3, and they satisfy

- $u_j(y)$ is linear in y,
- $u_j(x_k) = \delta_{jk},\ 1 \leq j,k \leq 3$,
- $p(y) = \sum_{j=1}^{3} u_j(y)p(x_j)$ for all $p \in I\!\!P_2^2$,

- $u_j(y) = 0$ iff y lies on the boundary line opposite to x_j,
- all $u_j(y) > 0,\ 1 \leq j \leq 3$ iff y lies inside the triangle,
- the $u_j(y)$ are nonnegative and sum up to 1 for y not outside the triangle.

The reader will have noticed that this is a very simple generalization from the two-point case. This can be carried further, but it never yields more than reproduction by linear polynomials. It always works for $d+1$ points that lie at least in $\mathbb{R}^d$ but not in a $d-1$-dimensional affine subspace.

It is now clear that in our three-point case we get $C_3 = 2$ independent of x and h, and if h is taken as the fill distance 2.1.2 (*DDDef*)

$$h := h_{\{x_1,x_2,x_3\},T} := \sup_{x\in T} \min_{1\leq j\leq 3} \|x - x_j\|_2,$$

of the triangle T, then we have $C_2 = 1$. This argument works on all small triangles that are formed by three data points that are not on a line.

We now assemble the two cases into a general strategy that works in $\mathbb{R}^2$ for polynomial reproduction of order $\ell \leq 2$. Assume that the set $X = \{x_1, \ldots, x_M\} \subset \mathbb{R}^2$ of scattered data is given, and let Ω be the convex hull of X, i.e.: the smallest convex set containing X. Then Ω is a compact convex polygon, and each point x of Ω either lies on a line between two points of X or in a nondegenerate triangle formed by three points of X. Assume that X fills Ω with a fill distance

$$h := h_{X,\Omega} := \sup_{x\in\Omega} \min_{1\leq j\leq M} \|x - x_j\|_2.$$

If the situation of one of the two above cases occurs, there will not necessarily be two points on a line with distance at most $2h$ or a triangle T with local fill distance h. We thus have to determine which distances as factors of h are possible in these cases. We form the Delaunay triangulation of the set $X = \{x_1, \ldots, x_M\}$ as described in section 11.1 (*SecVor*). This splits Ω into triangles with vertices at the points of X, and where there is an edge from x_k to x_j iff the midpoint between x_k and x_j has both x_k and x_j as points of X with minimal distance. Since this distance is at most h, the Delaunay triangles have edges of length at most $2h$. If we work on a line joining two vertices of the Delaunay triangulation, we thus have $C_2 = 1$. Inside of such triangles, the maximum distance from an interior point to the vertices is achieved in the isosceles case, and thus the fill distance within Delaunay triangles is at most $2h/\sqrt{3}$. We thus get away with $C_2 = 2/\sqrt{3}$ and $C_3 = 1$ in both cases.

5.4.2 Univariate Data

The situation for local polynomial interpolation of order exceeding two is much harder, even in one space dimension, where the solution still can be given using elementary techniques. Let us do a simple, but nonoptimal bound. Consider an odd number $\ell = 2k+1$ points ordered locally on the real line like

$$x_1 < x_2 < \ldots < x_\ell$$

and let the fill distance of $X = \{x_1, \ldots, x_\ell\}$ be $h/2$, such that we have $x_{i+1} - x_i \leq h$. The Lagrange basis functions for interpolation of order ℓ are

$$u_j(x) = \prod_{i \neq j} \frac{x - x_i}{x_j - x_i}, \ 1 \leq j \leq \ell$$

and they get large if there are points with $x_j - x_i$ extremely small. But those points can be neglected if our points are a local subsample of a much larger set. Let us thus assume a real number $\alpha \in (0,1]$ with $\alpha h \leq x_{i+1} - x_i \leq h$. Then the $M - 1 = 2k$ factors in the numerator can be bounded above by $(2k)!h^{2k}$, while the denominator can be bounded below by $(k!)^2\alpha^{2k}h^{2k}$. We have to sum $M = 2k+1$ of these quotients, and thus

$$C_3 \leq 1 + \frac{(2k+1)!}{(k!)^2\alpha^{2k}}, \ \ell = 2k+1.$$

This bound increases dramatically with ℓ, unfortunately, but it is independent of h. We can get an idea of the behaviour of C_3, if we apply **Stirling's inequality**

(*Stirling*)

$$1 \leq \frac{n!}{\sqrt{2\pi n}n^n e^{-n}} \leq \exp(-\frac{1}{12n}). \tag{5.4.1}$$

The result is

$$C_3 \leq \mathcal{O}\left(\sqrt{k}\left(\frac{2}{\alpha}\right)^{2k}\right)$$

for $k \to \infty$ or in simplified form and as a function of ℓ,

$$C_3 \leq \mathcal{O}(\gamma^\ell)$$

for $\ell \to \infty$ with some $\gamma > 1$.

Now let us apply this globally, and assume an ordered, but scattered set $X = \{x_1, \ldots, x_M\} \subset I\!R$ with fill distance h. For a uniform distribution of points of meshwidth $3h$ over $\Omega = [x_1, x_M]$ we associate a scattered point to each meshpoint, and then this selection of a subset of $N \leq M$ points has a fill distance of $2.5h$ and each adjacent pair of points is at least h and at most $5h$ apart. We then can apply the above bound with $\alpha = 1/5$ by local selection of ℓ points and an appropriate scaling. If we use a uniform distribution with spacing Kh, we end up with $\alpha = (K-2)/(K+2)$ which can be pushed towards 1 for K large. To check the value of C_2, we have to assume the worst case, in which some x lies at the boundary, while the next interpolation point is $2.5h$ away, and the interpolation points are at maximal distance $5h$. Then the maximal value of $|x - x_j|$ is $2.5h + (\ell - 1) \cdot 5 \cdot h < 5\ell h$ such that we have $C_2 = 5\ell$.

The above approach is unfeasible for multivariate cases, because we relied heavily on the ordering of the points. But it gives us two pieces of useful information: the good news is that we might get along with a quantity C_3 that does not depend on h, but the bad news is that C_3 will crucially depend on the order ℓ of local polynomial interpolation. We address the two topics one after another.

5.4.3 Independence of h

As we saw in the univariate case, one can expect that the scaling parameter h cancels out in the bounds for C_3. To generalize this statement, we repeat the technique that we already used before:

1. If a set $X = \{x_1, \ldots, x_M\}$ is given in Ω with fill distance h, we pick an integer $k \geq 3$ and lay a grid $G = kh\mathbb{Z}^d$ over Ω.

2. For each point from $G \cap \Omega$ we pick the nearest data point from X. This yields a subset Y of X of points that are only mildly scattered and are at least $(k-2)h$ apart from each other. We need this to avoid degeneration of the local polynomial interpolation that we want to construct. Since the diagonal in the unit cube in R^d has length $\sqrt{d}$, the fill distance of Y in Ω is at most $h(1 + k\sqrt{d})$.

3. If $x \in \Omega$ is given, we pick a selection of points from Y which are near to x, and the indices of these points define the set $J_x(h)$ occurring in Assumption 5.1.6 (*FBAss2*).

4. The main problem now is to prove that the selection guarantees solvability of polynomial interpolation of order ℓ.

5. We then evaluate the Lebesgue constants for this local interpolation.

If k is large, the set Y will consist of points that are relatively near to the grid $G = kh\mathbb{Z}^d$, since they can be only h away from gridpoints. Thus the local interpolation takes place on data that are slight perturbations of gridded data. We thus have to study polynomial interpolation on gridded data first, and then ask for admissible perturbations.

We write multivariate polynomials $p \in \mathbb{P}_\ell^d$ as

$$p(x) = \sum_{|\alpha|<\ell} p_\alpha x^\alpha \qquad (PolRep) \quad (5.4.2)$$

with the usual multiindex notation:

$$\alpha \in \mathbb{Z}_{\geq 0}^d, \ |\alpha| := \|\alpha\|_1, \ x^\alpha := \prod_{j=1}^d x_j^{\alpha_j}.$$

The number of data points should equal the number of basis functions, and thus we simply use the data set

$$X_\ell^d := \{ \beta \in \mathbb{Z}_{\geq 0}^d \ : \ |\beta| < \ell \}.$$

For $d = 2$ these are the points $(j,k) \in \mathbb{Z}^2$ with $0 \leq j,k \leq j+k < \ell$ forming a "triangle" in $\mathbb{Z}^2$.

Lemma 5.4.3 *(LemPIG) The set X_ℓ^d is a minimal nondegenerate set in $\mathbb{R}^d$ for polynomials in $\mathbb{P}_\ell^d$. Thus polynomial interpolation of order ℓ is uniquely possible.*

Proof: Since the dimension of $\mathbb{P}_\ell^d$ and the number of points in X_ℓ^d agree, it suffices to prove nondegeneracy. Let p be a polynomial of the form (5.4.2, *PolRep*) that vanishes on X_ℓ^d, and we want to show that p vanishes everywhere. We do this by induction on the space dimension d, and the case $d = 1$ is well-known. So we assume that for $k < \ell$ all polynomials from $\mathbb{P}_k^d$ that vanish on X_k^d must be identically zero. Now we extract the variable x_d from each of the terms in (5.4.2, *PolRep*), split x as $x = (\tilde{x}, x_d)$, and rearrange the sum. This yields

$$p(x) = p(\tilde{x}, x_d) = \sum_{j=0}^{\ell-1} p_j(\tilde{x}) x_d^j$$

with polynomials $p_j \in \mathbb{P}_{\ell-j}^{d-1}$. Setting $x = (0,k)$ for $0 \leq k < \ell$ we see that $(0,k) \in X_\ell^d$ and the univariate polynomial

$$p(0, x_d) = \sum_{j=0}^{\ell-1} p_j(0) x_d^j$$

in $I\!P_\ell^1$ vanishes in the ℓ distinct points $k,\ 0 \le k < \ell$. Thus it is zero as a polynomial in x_d, proving $p_j(0) = 0,\ 1 \le j < \ell$. Looking at $p_{\ell-1} \in I\!P_1^{d-1}$ we see that $p_{\ell-1}$ must be zero.

Now let us start an inner induction over an integer $j = 1, 2, \dots$ and assume that we already have proven that $p_{\ell-j}, \dots, p_{\ell-1}$ are identically zero, and that all of the p_i vanish on X_j^{d-1}. This is precisely what we have proven for $j = 1$ and $X_1^{d-1} = \{0\}$. Now fix an arbitrary $\beta \in X_{j+1}^{d-1} \setminus X_j^{d-1}$. Then $|\beta| = j$ and we can form the data points $x = (\beta, k) \in X_\ell^d$ for $0 \le k < \ell - j$. Considering these points, the univariate polynomial

$$p(\beta, x_d) = \sum_{i=0}^{\ell-j-1} p_i(\beta) x_d^i$$

must have vanishing coefficients, and thus all p_i vanish on X_{j+1}^{d-1}. For $p_{\ell-j-1} \in I\!P_{j+1}^{d-1}$ the hypothesis of the outer induction yields that this polynomial vanishes identically, and this finishes the inner and outer induction. $\square$

Now we know that (in a fixed enumeration of X_ℓ^d used for rows as well as columns) the matrix with elements α^β for $\alpha, \beta \in X_\ell^d$ is nonsingular. It is a continuous function of the data locations α, and thus it is still nonsingular when all the points vary in local balls of some positive radius $\rho \in (0, 1/2)$ around the integer points of X_ℓ^d. This radius is a function of both d and ℓ, and to give an explicit positive lower bound for it is a formidable task. We neglect this problem now and consider d and ℓ as fixed, leading to some mysterious, but clearly positive radius ρ for admissible perturbations.

Lemma 5.4.4 *(LemPIP) For each space dimension d and each polynomial order ℓ there is a positive quantity $\rho(d, \ell)$ such that interpolation by polynomials in $I\!P_\ell^d$ is uniquely possible in all data sets that pick a point in each of the balls*

$$B_\rho(\alpha) := \{\ y \in I\!R^d\ :\ \|y - \alpha\|_2 \le \rho\}$$

for all $\alpha \in X_\ell^d$. The maximum Lebesgue constant for all of these polynomial interpolation processes, measured on a fixed ball B_R of some large radius R containing the set X_ℓ^d is some finite positive quantity $C_3^*(d, \ell, R)$.

Proof: Each pick of points defines a nice interpolation problem that has Lagrange basis functions $\{u_\alpha\}_\alpha$ depending continuously on the locations of the points. Thus also the Lebesgue constant

$$1 + \sup_{x \in B_R} \sum_{\alpha \in X_\ell^d} |u_\alpha(x)|$$

varies (via the u_α) continuously with the data locations. Since these vary in a compact set, the Lebesgue constants, as defined above, attain a finite maximum under variation of the data locations. $\square$

Of course, one could replace the domain B_R of "measurement" in the Lebesgue constant by any compact set in $\mathbb{R}^d$, but note that the actual upper bound of the Lebesgue constants remains mysterious, and enlarging B_R will have a nasty blow-up effect.

The next step is the independence of the above situation under shifts and scaling:

Lemma 5.4.5 *(LemPIS) Let $Y \subset \mathbb{R}^d$ be a data set where interpolation by $\mathbb{P}_\ell^d$ is uniquely possible, and let $\{u_y(\cdot)\}_{y\in Y}$ be the associated Lagrange basis satisfying $u_y(x) = \delta_{xy}$ for $x, y \in Y$. If Y is translated by some $z \in \mathbb{R}^d$ and scaled by some $h > 0$ to go over into*

$$Z := h(Y - z) := \{\ y_h := h(y - z)\ :\ y \in Y\},$$

then interpolation in Z is equally possible, using the basis functions

$$u_{y_h}(\cdot) := u_y(z + \cdot/h)$$

and yielding the same Lebesgue constants, if the domain of measurement of those is translated and scaled accordingly, i.e.: the domain B is transformed into $h(B - y)$.

Proof: The definition of the new functions makes sure that they are polynomials in $\mathbb{P}_\ell^d$ and satisfy the Lagrange interpolation property. Looking at the definition of the Lebesgue constant proves the rest. $\square$

We now go back to our data set $X = \{x_1, \ldots, x_M\}$ that fills Ω with a fill distance h, and we pick points from X that are perturbations of points from a grid $kh\mathbb{Z}^d$ laid over Ω. The perturbations thus stay within h of the grid points, while these are kh apart along the axes. Scaling them down by division with kh will bring them to the unit grid Z^d, and the scaled perturbations will stay within a radius $1/k$. We thus have to make sure that $1/k \le \rho(d, \ell)$ holds and that we use a shifted, scaled, and perturbed version of X_ℓ^d for local interpolation. The point x must lie in the shifted and scaled domain of measurement of the Lebesgue constant. We then can use the bound $C_3^*(d, \ell, R)$ of the Lebesgue constant from Lemma 5.4.4 (*LemPIP)* for all h, as asserted.

Thus we are left to determine the constant C_2 of Assumption 5.1.6 (*FBAss2)* that bounds the maximal distance of x in terms of multiples of h to the points we use for interpolation. This is no big deal when x is in the interior of Ω and h is small enough, and we can then get away with something like $C_2 = k\ell\sqrt{d}$, the diameter of the cube $[0, k\ell]^d$. If x lies near the boundary of Ω, we must be more careful, because the boundary could have awkward outgoing cusps that take boundary points far away from places where we can find enough data from X that lie near the gridpoints of $kh\mathbb{Z}^d$ and allow full interpolation up to order ℓ. We make the following assumption:

Definition 5.4.6 *(DefICC) A closed compact domain $\Omega \in \mathbb{R}^d$ with nonempty interior satisfies an* **interior cone condition***, if there is a fixed positive angle γ and a fixed height δ such that for any boundary point x there is a cylindrical cone within Ω that has vertex x, angle γ at the vertex, and height δ.*

If Ω satisfies an interior cone condition, we can consider coverings of Ω by fine grids $\epsilon\mathbb{Z}^d$, and we see that there is a constant K_c such that for all ϵ that are small enough, i.e. $\epsilon \leq \epsilon_c$, any point of Ω is only $K_c\epsilon$ away from a grid cell of $\epsilon\mathbb{Z}^d$ that is completely contained in Ω. We apply this for $\epsilon := kh\ell \leq \epsilon_c$ and get that any x is at most $K_c\epsilon = K_ckh\ell$ away from a fully interior cell of sidelength $kh\ell$ in which we can do the local interpolation. Thus all interpolation points will be at most $(K_c + 1)kh\ell\sqrt{d}$ away from x and we can use $C_2 = (K_c + 1)k\ell\sqrt{d}$.

We have two restrictions up to now:

$$1/k \leq \rho(d, \ell) \text{ and } kh\ell \leq \epsilon_c.$$

This yields the conditions

$$h \leq h_c := \frac{\epsilon_c}{k\ell} \text{ and } k \geq \frac{1}{\rho(d, \ell)}$$

which are no problem for fixed values of d, ℓ, and the interior cone condition on Ω. We summarize:

Theorem 5.4.7 *(LPIT) For given values of d and ℓ and for a fixed cone condition there are positive constants h_c, C_2, C_3 such that Assumption 5.1.6 (FBAss2) is valid for all x in domains $\Omega \subset \mathbb{R}^d$ satisfying the cone condition, and for all $h \leq h_c$.* □

Theorem 5.4.7 (*LPIT)* is useful for all cases where the local approximation uses only a finite degree ℓ, and where the exact value of the constants does not matter much. We defer a more detailed analysis to the next lemmas, where we rely on [24](*madych-nelson:92-1*).

To treat the general case, we cite without proof a deep result from [24](*madych-nelson:92-1*):

Theorem 5.4.8 *(MNL) Let R be a cube in $I\!R^d$ which is divided into K^d identical subcubes for some large integer K, and define $\gamma_d := 2d(1+\gamma_{d-1})$ starting with $\gamma_1 := 2$. Consider arbitrary polynomials p from $I\!P_\ell^d$ and assume $K \geq \ell\gamma_d$. Then*

(MNLBound)

$$\|p\|_{\infty,R} \leq e^{2d\ell\gamma_d}\|p\|_{\infty,Y} \tag{5.4.9}$$

holds for any set $Y \subset R$ that picks a point from each of the subcubes. In particular, all these sets Y are $I\!P_\ell^d$-nondegenerate.

We now bring this into line with Lemma 5.4.4 (*LemPIP)* and assume $\ell \geq 2$ throughout. We want to let the little subcubes be centered around the points of X_ℓ^d. If their sidelength is 2ρ to make balls of radius ρ safely contained in the cubes, we have to take $(2\rho)^{-1} =: M \in I\!N$ and let the large cube be $R = [-\rho, \ell-1+\rho]^d$. Splitting it into K^d subcubes yields the equation

$$2\rho = \frac{\ell - 1 + 2\rho}{K}$$

which leads to

$$K = 1 + M(\ell-1),\ \rho = \frac{\ell-1}{2(K-1)},\ M = \lfloor \frac{\ell\gamma_d - 1}{\ell - 1} \rfloor.$$

This allows the application of Theorem 5.4.8 (*MNL)*. We first check the value of ρ as a function of d and ℓ. If we bound M crudely from above by $2\gamma_d$, we get $\rho \geq (4\gamma_d)^{-1}$, which is independent of ℓ.

The linear functional $\delta_x : p \mapsto p(x)$ can be written in the form

$$\sum_{\alpha \in X_\ell^d} p(z_\alpha)u_\alpha(x)$$

where we have picked z_α from the ball $B_\rho(\alpha)$, and where the u_α are the Lagrange interpolation polynomials. Then we fix $x \in R$ and interpolate data sgn $(u_\alpha(x))$ in z_α by some polynomial $\tilde{p} \in I\!P_\ell^d$ and get the bound

$$\sum_{\alpha \in X_\ell^d} |u_\alpha(x)| = \tilde{p}(x) \leq \|\tilde{p}\|_{\infty,R} \leq e^{2d\ell\gamma_d}.$$

Thus the Lebesgue constant in Lemma 5.4.4 (*LemPIP)* is bounded by

(*C3def*)

$$C_3^* \leq 1 + e^{2d\ell\gamma_d}. \tag{5.4.10}$$

We now go over to the situation in Theorem 5.4.7 (*LPIT)*. We fix a cone condition and a space dimension. This fixes the constants ϵ_c, h_c, and K_c from the cone condition. The integer k can be chosen as $k = 4\gamma_d$ to satisfy $k\rho \geq 1$, and we are left with the condition

(*hrestr*)

$$h \leq h_c = \frac{\epsilon_c}{4\ell\gamma_d} \tag{5.4.11}$$

under which we can use (5.4.10, *C3def*) and

(*C2def*)

$$C_2 = 4(K_c + 1)\ell\sqrt{d}\gamma_d. \tag{5.4.12}$$

Theorem 5.4.13 *(LPIT2) The assumption 5.1.6 (FBAss2) can be satisfied for each compact domain $\Omega \subset I\!R^d$ satisfying an interior cone condition 5.4.6 (DefICC) that defines positive constants ϵ_c, K_c. If γ_d is defined as in 5.4.8 (MNL), the constants C_2 and C_3 can be bounded by (5.4.12, C2def) and (5.4.10, C3def), respectively, while the polynomial order ℓ and the fill distance h must satisfy (5.4.11, hrestr).*

5.5 Error Bounds in Terms of Fill Distance

(*hrhodef*) We can now assemble the previous results into bounds of the form (4.6.3, *FBound*) for the power function from optimal recovery. Together with (4.1.4, *EqgSg1*) from page 81 this yields error bounds for the reconstruction of functions g from native spaces $\mathcal{G}$. Depending on the situation, we get quite explicit bounds for the power function in cases of small space dimensions and polynomial orders, while for fixed orders and arbitrary space dimension we use Theorem 5.4.7 (*LPIT)* to carry the order of the local bounds on the power functions over to the errors of optimal recoveries, the constants being mysterious. We list the orders (without the factors) of our L_∞ bounds on the power function in Table 6 (*TCPDEB)*, but delay the cases with exponential convergence somewhat. The additional data (parameters, domains, smoothness, dimension, order) should be looked up from tables 1 (*TCPDFct)* and 2 (*TPDFct)* on page 19. Note that the actual approximation orders of optimal recoveries may be better than the squares of these bounds.

$\phi(r)$	L_∞ Bound of Power Function
r^β	$h^{\beta/2}$
$r^\beta \log r$	$h^{\beta/2}$
$(r^2+\gamma^2)^{\beta/2}$	$\exp(-c/h),\ c>0$
$e^{-\beta r^2}$	$\exp(-c/h^2),\ c>0$
$r^\nu K_\nu(r)$	h^ν
$(1-r)_+^2(2+r)$	$h^{1/2}$
$(1-r)_+^4(1+4r)$	$h^{3/2}$

Table 6: L_∞ Bounds of Power Function Based on Lagrange Data (*TCPDEB*)

For radial functions $\Phi(x,y)=\phi(\|x-y\|_2)$ such that ϕ is Lipschitz continuous around the origin, we can apply Theorem 5.3.11 (*LipConvTh)* together with Theorem 5.4.7 (*LPIT)* to get convergence of the L_∞ norm of the power function to zero for $h\to 0$. We have used this fact in section 4.4.4 (*SecHSE)*.

Unfortunately, the factor $|g-S(g)|_\Phi$ in the actual error bound (4.1.4, *EqgSg1*) still is somewhat mysterious, if we start with a conditionally positive definite function Φ and construct the corresponding native space. If, on the other hand, we have started with $\mathcal{G}$, we are done. But note that these bounds can be improved, if g satisfies additional conditions. These improvements cannot come from better bounds on the power function, because we shall see that our techniques often provide optimal orders with respect to h. They rely on a deeper analysis of the term $|g-S(g)|_\Phi$, and this analysis will be done in 5.6 (*SecEBStage2)* and 5.7 (*SecEBStage3)*.

We now discuss the cases of multiquadrics and Gaussians, where we can push the polynomial order ℓ up to infinity. The overall bound for $P^2(x)$ is given by (5.1.11, *FundBound*), and we have to insert (5.4.12, *C2def*), (5.4.10, *C3def*) and the replacement of h by $2h\sqrt{d}$ from section 5.2 (*SecAERC)*. The values of ρ and C_1 depend on the special case chosen.

Let us first look at multiquadrics. The bound on $P^2(x)$ then is

$$P^2(x) \quad\le\quad 2^{|\beta/2|+1}c^\beta\left(1+e^{2d\ell\gamma_d}\right)^2\left(4(K_c+1)\ell\sqrt{d}\gamma_d\right)^{2n}\left(\tfrac{8h^2d}{c^2}\right)^n$$

under the restrictions $8h^2d<c^2/2$ and (5.4.11, *hrestr*). We now treat everything as fixed except h and $\ell=2n-1$. This turns the bound into something of the form

$$C_4\left(C_5nh\right)^{2n},$$

and we shall pick $n = (\ell + 1)/2$ as a function of h as large as possible, but such that the constraints

$$C_5 nh \le \gamma < 1,\ 4\ell h\gamma_d \le \epsilon_c$$

are satisfied. This works with $n = ch^{-1}/2$ and some positive constant c. Then the bound becomes

$$C_4\gamma^{c/h} = C_4 \exp(-|\log\gamma|c/h)$$

and proves exponential behaviour for $h \to \infty$.

The Gaussian case is quite similar and can easily be reduced to a bound like

$$C_4 \frac{(C_5 nh)^{2n}}{n!},$$

which allows the same treatment. But now we can use the additional $n!$ in the denominator to speed up the convergence. We first insert Stirling's formula (5.4.1, *Stirling*) into the denominator to cancel an n^n factor in the numerator, introducing some change in the constants C_4 and C_5. This yields

$$C_4 \left(C_5\sqrt{n}h\right)^{2n},$$

and we now pick $n = ch^{-2}/2$ to get

$$C_5\sqrt{n}h \le \gamma < 1.$$

The second restriction, induced by the interior cone condition, cannot be satisfied in this case. Furthermore, there are problems on bounded domains, because the data points needed for reconstruction at x spread out to distance $\mathcal{O}(h\ell) = \mathcal{O}(h^{-1})$. This is why our final result for Gaussians will only hold for $\Omega = I\!R^d$ and infinite data sets. We get the bound

$$C_4\gamma^{c/h^2} = C_4 \exp(-|\log\gamma|c/h^2)$$

with "Gaussian" exponential behaviour for $h \to \infty$.

Theorem 5.5.1 *(GMCEBT) The power functions of Lagrange interpolation by multiquadrics and Gaussians have L_∞ bounds of the form* $\exp(-c/h)$ *with $c > 0$ for compact domains $\Omega \subset I\!R^d$ satisfying an interior cone condition. The bound for the Gaussian can be improved to* $\exp(-c/h^2)$ *for $\Omega = I\!R^d$.*

5.5.1 Remarks

The proof of bounds on the power function via polynomial approximation goes back to Duchon [10](*duchon:76-1*) for thin-plate splines and was successfully generalized by Madych and Nelson [21](*madych-nelson:88-1*), [24](*madych-nelson:92-1*). The special cases of lines and triangles were done for thin-plate splines by Powell [39](*powell:93-1*).

5.6 Doubling the Approximation Order

(*SecEBStage2*) Here we show how the error bounds of the form (4.1.4, *EqgSg1*) can be improved by adding some assumptions on the function g that is reconstructed. A third enhancement, based on a localization argument, will follow in 5.7 (*SecEBStage3*).

We work in the setting of section 3.5.5 (*SecCLC*) and use local integration of the square of the error to get

$$\|r_0(g - S(g))\|_{L_2(\Omega_0)} \le \|P_\Lambda(\cdot)\|_{L_2(\Omega_0)}\|g - S(g)\|_{\Phi,\Omega}$$

for the optimal power function P_Λ and the optimal recovery $S(g)$ of $g \in \mathcal{G}_\Omega$.

For any $g \in \mathcal{G}_\Omega$ we can consider the function

$$C(g) := \Pi_{\mathcal{P}} g + c^0 r_0(g - \Pi_{\mathcal{P}} g) \in \mathcal{G}_\Omega.$$

Then we have

$$\begin{array}{rcl} (Cg, f)_{\Phi,\Omega} & = & (c^0 r_0(g - \Pi_{\mathcal{P}} g), f - \Pi_{\mathcal{P}} f)_{\Phi,\Omega} \\ & = & (r_0(g - \Pi_{\mathcal{P}} g), r_0(f - \Pi_{\mathcal{P}} f))_{L_2(\Omega_0)} \end{array}$$

for all $f, g \in \mathcal{G}_\Omega$. We now define the subspace

$$\mathcal{H}_\Omega := C(\mathcal{G}_\Omega) \subseteq \mathcal{G}_\Omega$$

of $\mathcal{G}_\Omega$ and consider optimal recovery of functions $g = C(f_g) \in \mathcal{H}_\Omega$ with $f_g \in \mathcal{G}$ by $g^* = S(g)$. The orthogonality (3.1.34, *EqOrtho*) then implies

$$\begin{array}{rcl} \|g - S(g)\|^2_{\Phi,\Omega} & = & (g - S(g), g - S(g))_{\Phi,\Omega} \\ & = & (g, g - S(g))_{\Phi,\Omega} \\ & = & (Cf_g, g - S(g))_{\Phi,\Omega} \\ & = & (r_0(f_g - \Pi_{\mathcal{P}} f_g), r_0(g - S(g) - \Pi_{\mathcal{P}} g + \Pi_{\mathcal{P}} S(g)))_{L_2(\Omega_0)} \\ & \le & \|r_0(f_g - \Pi_{\mathcal{P}} f_g)\|_{L_2(\Omega_0)} \|r_0(g - S(g))\|_{L_2(\Omega_0)} \\ & \le & \|r_0(f_g - \Pi_{\mathcal{P}} f_g)\|_{L_2(\Omega_0)} \|P_\Lambda(\cdot)\|_{L_2(\Omega_0)} \|g - S(g)\|_{\Phi,\Omega} \end{array}$$

and this allows to bound $\|g - S(g)\|_{\Phi,\Omega}$ nicely by

$$\|g - S(g)\|_{\Phi,\Omega} \leq \|P_\Lambda\|_{L_2(\Omega_0)} \|r_0(f_g - \Pi_{\mathcal{P}} f_g)\|_{L_2(\Omega_0)}$$

for all $g = Cf_g \in \mathcal{H}$. If we combine this with (4.1.4, *EqgSg1*), we get

Theorem 5.6.1 *(EBStage2T) For optimal reconstruction of functions $g \in \mathcal{H} \subseteq \mathcal{G}$ with $g = Cf_g$ by optimal recovery functions $S(g)$ we have the improved error bound*

$$|(g - S(g))(x)| \leq P(x)\|P\|_{L_2(\Omega_0)} \|r_0(f_g - \Pi_{\mathcal{P}} f_g)\|_{L_2(\Omega_0)}$$

for all $x \in \Omega_0$.

5.6.1 Remarks

The results of 5.6 (*SecEBStage2)* are from [40](*schaback:96-2*) and derived from the arguments used in classical spline theory [2](*ahlberg-et-al:86-1*) to improve the approximation order via the “second integral relation”.

5.7 Improvement by Localization

(*SecEBStage3*) Here we use a localization argument from [2](*Light-Wayne:96-1*) dating back to Duchon [10](*duchon:76-1*) to get some additional powers of h for error bounds of optimal recovery via Lagrange interpolation. We delay the formulation of these results.

6 Advanced Results on $I\!R^d$

(*SecARRd*) Here we apply Fourier transforms in $I\!R^d$, and derive a series of results that require related techniques. These include bounds on the stability and error bounds for the multilevel method. The reader should look into section 12.5 (*SecFTRd)* for backup material on Fourier transforms.

6.1 Transforms of Translation-Invariant Basis Functions

(*SecCNST*) In section 3.2.4 (*SecIP)* we have seen that on $I\!R^d$ we can restrict ourselves to cases where the recovery problem is translation-invariant or even invariant under Euclidean rigid-body transformations. In the first case, the conditionally positive definite functions $\Phi(x, y)$ take the form $\Phi(x - y)$, while in the second they are radial: $\Phi(x, y) = \phi(\|x - y\|_2)$. We start with the more general case, but we restrict ourselves to **un**conditionally positive definite functions first.

6.1.1 Unconditionally Positive Definite Functions

So let us now consider a function $\Phi : \mathbb{R}^d \to \mathbb{R}$ with $\Phi(-\cdot) = \Phi(\cdot)$ such that $\Psi(x, y) := \Phi(x - y)$ is a candidate for an unconditionally positive definite function on $\mathbb{R}^d$. We want to look at conditions that allow us to conclude that Φ actually is unconditionally positive definite on $\mathbb{R}^d$. Having the Gaussian in mind as a prominent example, we assume Φ to have a Fourier transform $\hat{\Phi}$ on $\mathbb{R}^d$ such that the Fourier inversion formula holds. We now want to construct the native space $\mathcal{G}$ by the techniques of section 3.3 (*SecNS*). Clearly, the representation (3.3.2, *DefBil*) of the bilinear form $(\cdot, \cdot)_\Phi$ can now be rewritten as

(*DefBil2*)

$$(\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta})_\Phi = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{\Phi}(\omega) \sum_{j=1}^{M} \sum_{k=1}^{N} \alpha_j \overline{\beta_k} e^{i(x_j - y_k)\cdot\omega} d\omega. \qquad (6.1.1)$$

The functions $F(\lambda_{X,M,\alpha})$ have Fourier transforms

$$\begin{aligned} \widehat{F(\lambda_{X,M,\alpha})}(\omega) &= \hat{\Phi}(\omega) \sum_{j=1}^{M} \alpha_j e^{-x_j\cdot\omega} \\ &= \hat{\Phi}(\omega) \hat{\lambda}_{X,M,\alpha} \end{aligned}$$

if we use the definition of Fourier transforms of functionals from Example 12.5.20 (*Exlxma*). This allows the shorthand representations

$$\begin{aligned} (\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta})_\Phi &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{\Phi}(\omega) \widehat{\lambda_{X,M,\alpha}}(\omega) \overline{\widehat{\lambda_{Y,N,\beta}}(\omega)} d\omega \\ (F\lambda_{X,M,\alpha}, F\lambda_{Y,N,\beta})_\Phi &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{1}{\hat{\Phi}(\omega)} \widehat{F\lambda_{X,M,\alpha}}(\omega) \overline{\widehat{F\lambda_{Y,N,\beta}}(\omega)} d\omega \\ &. \end{aligned}$$

We still have to add some arguments that convince us that the above representations define positive definite bilinear forms. Equation (6.1.1, *DefBil2*) yields

(*DefNorm2*)

$$\|\lambda_{X,M,\alpha}\|_\Phi^2 := (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{\Phi}(\omega) \left| \sum_{j=1}^{M} \alpha_j e^{ix_j\cdot\omega} \right|^2 d\omega, \qquad (6.1.2)$$

and we see that nonnegativity of this integral is closely related to nonnegativity of $\hat{\Phi}$. Starting from positive definiteness of Φ, it is hard (or even impossible) to deduce positivity almost everywhere of $\hat{\Phi}$ from this equation.

Conversely, in all known cases the Fourier transform $\hat{\Phi}$ can directly proven to be nonnegative almost everywhere anyway. Then the representation (6.1.1, *DefBil2*) yields a useful sufficient condition for positive definiteness:

Theorem 6.1.3 *(NCPDFTT) If $\Phi : I\!R^d \to I\!R$ is absolutely integrable with a real-valued Fourier transform $\hat{\Phi}$ that is positive almost everywhere on $I\!R^d$, then Φ is even, continuous, and unconditionally positive definite on $I\!R^d$.*

Proof: From Lemma 12.5.17 (*FTLoneLem*) we get that $\hat{\Phi}$ is continuous and even, and from (6.1.2, *DefNorm2*) we see that Φ is positive semidefinite. To prove definiteness, we have to prove $\alpha = 0$ if

$$\sum_{j=1}^{M} \alpha_j e^{ix_j \cdot \omega} = 0$$

holds almost everywhere on $I\!R^d$. But then the equation must hold on all of $I\!R^d$, and we can use the argument of Theorem 12.5.6 (*GaussPD*) to get $\alpha = 0$. $\square$

6.1.2 Conditionally Positive Definite Functions

We now go over to the treatment of general unconditionally positive definite functions. To do this, we shall introduce Fourier transforms in a somewhat more general way that will later save us quite some work. The direct attack is impossible, because some of the most important conditionally positive definite functions on $I\!R^d$ are radial functions $\Phi(\cdot) = \phi(\|\cdot\|_2)$ that grow towards infinity, e.g.: thin-plate splines $\phi(r) = r^2 \log r$ or multiquadrics $\phi(r) = \sqrt{r^2 + c^2}$. These do not have classical Fourier transforms, but since they grow at most polynomially, they are tempered functions in $\mathcal{K}$. Thus they have generalized Fourier transforms defined via the Fourier transforms of the functionals that they induce on $\mathcal{S}$. These generalized Fourier transforms are not straightforward to handle and require quite some machinery from distribution theory. Two different ways to do this are treated in the Ph.D. theses by A. Iske [18](*iske:94-1*) and M. Weinrich [45](*weinrich:94-1*).

We go a different way [42](*schaback:96-1*) by picking a very specific set of assumptions to start with, and then we can work our way without distributions. We do not even assume Φ to be a conditionally positive definite function; this will be a consequence of our assumptions and lead to an important technique to prove conditional positive definiteness for specific examples.

Assumption 6.1.4 *(FTAss1) Let $\Phi : I\!R^d \to I\!R$ be even and continuous. Furthermore, let there be a continuous nonnegative function*

$$\hat{\Phi} : I\!R^d \setminus \{0\} \to I\!R$$

which is positive almost everywhere. It may possibly have an algebraic singularity

(PhiSingCond)

$$\hat{\Phi}(\omega) = \mathcal{O}(\|\omega\|^{-d-\beta_0}) \tag{6.1.5}$$

with some real value β_0 for ω near zero, and it must have the behavior

(PhiInfCond)

$$\hat{\Phi} \in L_1 \text{ near infinity.} \tag{6.1.6}$$

Then define $m := \max(0, \lfloor \beta_0 \rfloor) \geq 0$ to get the restriction

(BetaZCond)

$$\beta_0 < 2m \tag{6.1.7}$$

that will often occur later. Finally, let the usual bilinear form on $\mathcal{P}_\Omega^- = (I\!P_m^d)_{I\!R^d}^-$ be representable by (6.1.1, DefBil2).

Theorem 6.1.8 *(NCCPDFTT) Under the above assumptions the function $\Phi(x-y)$ is conditionally positive definite of order $\geq m$ on $I\!R^d$.*

Proof: Note first that the functionals $\lambda_{X,M,\alpha} \in (I\!P_m^d)_{I\!R^d}^-$ have Fourier transforms with zeros of order at least m in the origin. Thus the integrand in (6.1.1, *DefBil2*) is of order $\mathcal{O}(\|\omega\|^{2m-d-\beta_0})$ near zero, and the integral is well-defined due to (6.1.7, *BetaZCond*) and (6.1.6, *PhiInfCond*). Nonnegativity of $\hat{\Phi}$ proves that the bilinear form is positive semidefinite. The rest is as in the proofs of Theorems 6.1.3 (*NCPDFTT)* and 12.5.6 (*GaussPD)*. □

The reader should be aware that we did not assume $\hat{\Phi}$ to be the usual Fourier transform. We thus cannot use equations (12.5.2, *FT*) or (12.5.9, *IFT*), but we have the general identity

$$\sum_{j=1}^{M}\sum_{k=1}^{N} \alpha_j \beta_k \Phi(x_j - y_k) = (2\pi)^{-d/2} \int_{I\!R^d} \hat{\Phi}(\omega) \sum_{j=1}^{M}\sum_{k=1}^{N} \alpha_j \overline{\beta_k} e^{i(x_j - y_k)\cdot\omega} d\omega.$$

that is identical to (6.1.1, *DefBil2*) and is valid for all functionals in $(I\!P_m^d)_{I\!R^d}^-$ due to Assumption 6.1.4 (*FTAss1)*. It will nicely serve as a substitute for (12.5.9, *IFT*) in the form (12.5.10, *IFT2*), but note that it does not allow single point-evaluation functionals in the left-hand side.

6.2 Connection to $L_2(I\!R^d)$

Assumption 6.1.4 (*FTAss1)* makes sure that the mapping

$$L \;:\; \lambda \mapsto \widehat{\lambda}\sqrt{\widehat{\Phi}},\;\; (I\!P_m^d)^-_{I\!R^d} \to L_2(I\!R^d)$$

is well-defined. Indeed, the function $L(\lambda)$ is in L_2 near infinity due to (6.1.6, *PhiInfCond*), and it is continuous around zero due to (6.1.7, *BetaZCond*), since $\widehat{\lambda}$ has a zero of order at least m at the origin.

Theorem 6.2.1 *Let Assumption 6.1.4 (FTAss1) be satisfied, and let m be minimal with respect to (6.1.7, BetaZCond). Then the map L extends by continuity to clos$((I\!P_m^d)^-_{I\!R^d})$, and it yields an isometry between clos$((I\!P_m^d)^-_{I\!R^d})$ and* all *of $L_2(I\!R^d)$.*

Proof: It is evident from (6.1.1, *DefBil2*) that L is isometric, and thus L extends to clos$((I\!P_m^d)^-_{I\!R^d})$ by continuity. But the density of $L($clos$((I\!P_m^d)^-_{I\!R^d}))$ in $L_2(I\!R^d)$ does not follow from abstract Hilbert space arguments. We thus need an additional analytic argument. We first prove the assertion for continuous $\widehat{\Phi}$ with $\widehat{\Phi} > 0$ on $I\!R^d \setminus \{0\}$.

Let some function $f \in L_2(I\!R^d)$ and some $\varepsilon > 0$ be given. Then there is a compactly supported C^∞ function $g \in L_2(I\!R^d)$ such that $\|f-g\|_2 \le \varepsilon$ due to Lemma 12.4.5 (*FTDC)*. Now define $\hat{u} := g/\sqrt{\widehat{\Phi}}$ on $I\!R^d$, where the (possible) singularity of $\widehat{\Phi}$ at zero does no harm. Clearly $\hat{u}$ is continuous and compactly supported, thus in $L_2(I\!R^d)$ and u is band–limited, of exponential type, and in $L_2(I\!R^d)$. We now invoke the multivariate sampling theorem 8.1.1 (*MST)* to recover u exactly from its function values on a grid in $I\!R^d$ with spacing h, where h is sufficiently small and related to the support of $\hat{u}$.

Thus we have

$$u(x) = \sum_{j\in \mathbb{Z}^d} u(jh)\,\mathrm{Sinc}_d\left(\frac{x-jh}{h}\right), \quad x \in I\!R^d$$

where

$$\mathrm{Sinc}_d(x_1,\ldots,x_d) = \prod_{j=1}^d \frac{\sin \pi x_j}{\pi x_j},$$

and

$$\hat{u}(\omega) = \sum_{j\in \mathbb{Z}_d} u(jh)e^{ihj\cdot\omega}, \qquad \omega \in I\!R^d$$

has the form $\hat{u} = \widehat{\lambda_u}$ for the functional

$$\lambda_u(v) = \sum_{j\in\mathbb{Z}^d} v(jh)u(jh).$$

We now have to make sure that $\lambda_u \in \operatorname{clos}\left((I\!P_m^d)^-_{I\!R^d}\right)$. If this is done, we are finished because of $L(\lambda_u) = g$ and

$$\|f - \sqrt{\hat{\Phi}}\widehat{\lambda u}\|_2 = \|f - g\|_2 \le \varepsilon.$$

For all $p \in I\!P_m^d$ we have to show that $\lambda_u(p) = 0$. By a standard argument in Fourier analysis this requires a zero of order at least m of $\hat{u}$ at zero. But our assumption (6.1.5, *PhiSingCond*) on $\hat{\Phi}$ and the minimality of m in (6.1.7, *BetaZCond*) imply that $\hat{u}$ has a zero of order at least

$$\frac{1}{2}(d+\beta_0) > \frac{1}{2}(d+2m-2) = m-1+\frac{d}{2},$$

thus of order $\ge m$.

We then evaluate the norm formally as

$$\|\lambda_u\|_\Phi^2 = \|\sqrt{\hat{\Phi}}\cdot\widehat{\lambda_u}\|_2^2 = \|\sqrt{\hat{\Phi}}\hat{u}\|_2^2 = \|g\|_2^2 < \infty.$$

Now we can proceed to prove that λ_u lies in $\operatorname{clos}\left((I\!P_m^d)^-_{I\!R^d}\right)$ by defining the function

$$f_{\lambda_u}(x) := (\lambda_u, \delta_{x,\Xi})_\Phi,\ x \in I\!R^d$$

via the explicit form of the inner product, and using the finiteness of the norm $\|\lambda_u\|_\Phi$ to show that the definition is valid. Then for all $\lambda_{Y,N,\beta} \in (I\!P_m^d)^-_{I\!R^d}$ we get

$$\lambda_{Y,N,\beta}(f_{\lambda_u}) = (\lambda_u, \lambda_{Y,N,\beta})_\Phi$$

and this proves that $f_{\lambda_u} \in \mathcal{F}$. Finally, we get $\lambda_u = F^{-1}(f_{\lambda_u})$ by checking

$$\begin{aligned}(\lambda_u, \lambda_{Y,N,\beta})_\Phi &= \lambda_{Y,N,\beta}(f_{\lambda_u})\\ &= (\lambda_{Y,N,\beta}F^{-1}f_{\lambda_u}))_\Phi\end{aligned}$$

for all $\lambda_{Y,N,\beta} \in (I\!P_m^d)^-_{I\!R^d}$, and this concludes the proof in case of $\hat{\Phi} > 0$.

Now let $\hat{\Phi}$ be positive up to a set of Lebesgue measure zero. We cover the set of zeros by intervals I_k, where k varies over some index set K and the total area $\sum_k |I_k|$ is less than some given δ. Now let $\hat{\Phi}_\delta(\omega) \ge \hat{\Phi}(\omega)$ be a strictly positive continuous function that differs from $\widehat{\Phi}$ only on the I_k. Then $\widehat{\Phi}_\delta$ will also satisfy our assumptions, and we can use (6.1.1, *DefBil2*) in the form

$$(\mu, \lambda)_{\Phi_\delta} := (2\pi)^{-d/2} \int_{I\!R^d} \widehat{\Phi}_\delta(\omega)\widehat{\lambda}(\omega)\overline{\widehat{\mu}(\omega)}d\omega$$

as a definition of an inner product, but we do not need Φ_δ explicitly.

Now we approximate a given $f \in L_2(I\!R^d)$ by some $\sqrt{\widehat{\Phi}_\delta} \cdot \widehat{\lambda}$ up to $\varepsilon/2$ in the L_2 norm, picking a suitable λ for each δ and ε. Then

$$\|f - \sqrt{\widehat{\Phi}}\widehat{\lambda}\|_2 \le \|f - \widehat{\lambda}\sqrt{\widehat{\Phi}_\delta}\|_2 + \|\widehat{\lambda}(\sqrt{\widehat{\Phi}_\delta} - \sqrt{\widehat{\Phi}})\|_2$$

and

$$\begin{aligned} \|\widehat{\lambda}(\sqrt{\widehat{\Phi}_\delta} - \sqrt{\widehat{\Phi}})\|_2^2 &= \|\widehat{\lambda} \cdot \sqrt{\widehat{\Phi}_\delta}(1 - \sqrt{\widehat{\Phi}/\widehat{\Phi}_\delta})\|_2^2 \\ &\le \sum_k \int_{I_k} |\widehat{\lambda}(\omega)|^2 \widehat{\Phi}_\delta(\omega) d\omega. \end{aligned}$$

The full integral

$$\int_{I\!R^d} |\widehat{\lambda}(\omega)|^2 \widehat{\Phi}_\delta(\omega) d\omega = \|\widehat{\lambda} \cdot \sqrt{\widehat{\Phi}_\delta}\|_2^2$$

can be bounded independent of δ, because it approximates $\|f\|_2^2$. Thus we are able to pick δ small enough to guarantee

$$\sum_k \int_{I_k} |\widehat{\lambda}(\omega)|^2 \widehat{\Phi}_\delta(\omega) d\omega \le \varepsilon/2$$

yielding an overall bound $\|f - \sqrt{\widehat{\Phi}}\widehat{\lambda}\|_2 \le \varepsilon$. □

6.3 Characterization of Native Spaces

We now use Theorem 6.2.1 (*LCTh)* to characterize the native space $\mathcal{G}_{I\!R^d}$ for Φ via $L_2(I\!R^d)$. Starting with an arbitrary $h \in L_2(I\!R^d)$ and a fixed $I\!P_m^d$-unisolvent set $\Xi \subset I\!R^d$, we mimic the technique of (3.3.15, *lf3*) to define a function

(*fhdef*)

$$f_h(x) := (\widehat{h}, L\delta_{x,\Xi})_{L_2(I\!R^d)}. \tag{6.3.1}$$

It is in $\mathcal{F}_{I\!R^d}$, because

$$\lambda f_h = (\widehat{h}, L\lambda)_{L_2(I\!R^d)}$$

follows easily from (6.3.1, *fhdef*) for all $\lambda \in (I\!P^d_m)^-_{I\!R^d}$. We can transform this equation further into

$$\begin{array}{rcl}\lambda f_h & = & (\widehat{h}, L\lambda)_{L_2(I\!R^d)}\\ & = & (L^{-1}\widehat{h}, \lambda)_\Phi\\ & = & (F^{-1}f_h, \lambda)_\Phi\end{array}$$

to see that

(*FTDef*)

$$\widehat{f_h} = \sqrt{\widehat{\Phi}}\widehat{h} = \sqrt{\widehat{\Phi}}LF^{-1}f_h \tag{6.3.2}$$

is another way to define $\widehat{f_h}$. We can rewrite (6.3.1, *fhdef*) as

$$\begin{array}{rcl}f_h(x) & = & (2\pi)^{-d/2}\displaystyle\int_{I\!R^d} \widehat{h}(\omega)\sqrt{\widehat{\Phi}(\omega)}\left(e^{ix\cdot\omega} - \sum_{j=1}^{Q} p_j(x)e^{i\xi_j\cdot\omega}\right)d\omega\\ & = & (2\pi)^{-d/2}\displaystyle\int_{I\!R^d} \widehat{f_h}(\omega)\left(e^{ix\cdot\omega} - \sum_{j=1}^{Q} p_j(x)e^{i\xi_j\cdot\omega}\right)d\omega\end{array}$$

where we define

$$\widehat{f_h} := \widehat{h}\cdot\sqrt{\widehat{\Phi}},$$

which is fully consistent with the usual notation for Fourier transforms in case of $m = 0$. We then get

Theorem 6.3.3 *The native space $\mathcal{G}_{I\!R^d}$ for a conditionally positive definite function of order m on $I\!R^d$ satisfying Assumption 6.1.4 (FTAss1) coincides with the space of all functions f on $I\!R^d$ that can be written as*

(FTfdef)

$$f(x) = (2\pi)^{-d/2}\int_{I\!R^d} \widehat{f}(\omega)\left(e^{ix\cdot\omega} - \sum_{j=1}^{Q} p_j(x)e^{i\xi_j\cdot\omega}\right)d\omega \tag{6.3.4}$$

plus polynomials from $I\!P^d_m$, and where $\widehat{f}$ is a function that can be defined via (6.3.2, FTDef) and satisfies

$$\widehat{f}/\sqrt{\widehat{\Phi}} \in L_2(I\!R^d).$$

The inner product on $\mathcal{G}_{I\!R^d}$ can be rewritten on the subspace $\mathcal{F}_{I\!R^d}$ as

$$(f,g)_\Phi = (2\pi)^{-d/2}\int_{I\!R^d} \frac{\widehat{f}(\omega)\cdot\overline{\widehat{g}(\omega)}}{\widehat{\Phi}(\omega)}d\omega.$$

□

Note that $\hat{f}$ is only defined for functions in $\mathcal{F}_{I\!R^d}$. In many cases, even with Φ increasing polynomially towards infinity, the functions $f = g - \Pi_{I\!P^d_m} g \in \mathcal{F}_{I\!R^d}$ for arbitrary $g \in \mathcal{G}_{I\!R^d}$ will decay sufficiently fast to have classical Fourier transforms, and then (6.3.4, *FTfdef*) coincides with the Fourier inversion formula (12.5.10, *IFT2*). The expression in brackets makes sure that the left-hand side automatically is $(g - \Pi_{I\!P^d_m} g)(x)$ if we insert $f = g - \Pi_{I\!P^d_m} g$ into the right-hand side. Thus (6.3.4, *FTfdef*) is nothing else than the Fourier inversion formula modulo polynomials.

If we define the mapping

$$M \; : \; \mathcal{F}_{I\!R^d} \to L_2(I\!R^d) \;\; f \mapsto \hat{f}/\sqrt{\hat{\Phi}},$$

it is easy to see that

$$\lambda(f) = (Mf, L\lambda)_{L_2(I\!R^d)} = (\hat{f}/\sqrt{\hat{\Phi}}, \hat{\lambda}\sqrt{\hat{\Phi}})_{L_2(I\!R^d)} = \int_{I\!R^d} \hat{f}\overline{\hat{\lambda}}$$

holds for all $\lambda \in \operatorname{clos}(I\!P^d_m)^{-}_{I\!R^d}$ and all $f \in \mathcal{F}_{I\!R^d}$. Along the same lines or directly from (6.3.2, *FTDef*) we get

$$L = M \circ F$$

as expected. Note that (so far) we only have *inverse* explicit formulae that allow to calculate f or $\lambda(f)$ from the transforms. The opposite direction is not explicitly given on its full domain, but rather on the special functionals and functions of the form $\lambda_{X,M\alpha} \in (I\!P^d_m)^{-}_{I\!R^d}$ and $F\lambda_{X,M\alpha}$. In particular, we have

$$\begin{aligned} \widehat{(F\lambda_{X,M\alpha})}(\omega) &= \sqrt{\hat{\Phi}}(\omega) MF\lambda_{X,M\alpha}(\omega) \\ &= \sqrt{\hat{\Phi}}(\omega) L\lambda_{X,M\alpha}(\omega) \\ &= \hat{\Phi}(\omega)\hat{\lambda}_{X,M\alpha}(\omega) \\ &= \hat{\Phi}(\omega)\sum_{j=1}^{M} \alpha_j e^{-ix_j\cdot\omega}, \end{aligned}$$

which nicely agrees with the formula

$$\widehat{\left(\sum_{j=1}^{M} \alpha_j \Phi(x_j - \cdot)\right)}(\omega) = \hat{\Phi}(\omega)\sum_{j=1}^{M} \alpha_j e^{-ix_j\cdot\omega}$$

we would expect from classical Fourier transforms. But note that the latter cannot be obtained or defined termwise, because the single terms do not have classical Fourier transforms. Furthermore, both sides still may be singular at the origin.

6.4 Condition Numbers

(*SecCNTrans*) This section uses Fourier transform techniques to prove results concerning the condition of the matrices that occur in the basic equations (1.7.3, *BDef*) for optimal recovery. This requires upper bounds for the largest, and lower bounds for the smallest eigenvalue. We start with the latter and restrict ourselves to the Lagrange case. The bounds should (if possible) should neither depend on the specific data locations $X = \{x_1, \ldots, x_M\}$, nor on the number M of data points, but rather on certain real-valued quantities like the separation distance (2.1.1, *SDDef*).

6.4.1 Stability Bounds

(*SecSB*) We go back to the setting in section 4.5 (*SecCondition)* and want to calculate lower bounds for the smallest eigenvalue σ defined in (4.5.3, *Defsigma*) and providing the stability bound (4.5.4, *Stab*).

6.4.2 Narcowich-Ward Technique

(*SecNWT*) We generalize the technique of Narcowich and Ward [32](*narcowich-ward:91-1*) [33](*narcowich-ward:91-2*) [34](*narcowich-ward:92-2*) for calculating stability bounds, but we introduce Fourier transforms right from the start, which makes it much easier to treat large values of m, the order of conditional positive definiteness of Φ.

The starting point is that any conditionally positive definite function Φ of order m satisfying Assumption 6.1.4 (*FTAss1)* allows the formula

(*EqNFT*)

$$\sum_{j=1}^{M}\sum_{k=1}^{M}\alpha_j\alpha_k\Phi(x_j - x_k) = (2\pi)^{-d/2}\int_{I\!R^d}\widehat{\Phi}(\omega)\left|\sum_{j=1}^{M}\alpha_j e^{ix_j\cdot\omega}\right|^2 d\omega \qquad (6.4.1)$$

for all $I\!P_m^d$–nondegenerate sets $X = \{x_1, \ldots, x_M\}$ and all vectors $\alpha \in I\!R^M$ such that $\lambda_{X,M,\alpha}$ is a functional that annihilates $I\!P_m^d$. This is just another way of writing (6.1.1, *DefBil2*).

The left–hand side of (6.4.1, *EqNFT*) is the quantity $\alpha^T A_{X,\Phi}\alpha$ that we want to bound from below, and we can do this by any minorant $\widehat{\Psi}$ on $I\!R^d \setminus \{0\}$ of $\widehat{\Phi}$ that satisfies

(*EqPhiPsi*)

$$\widehat{\Phi}(\omega) \geq \widehat{\Psi}(\omega) \qquad \text{on } I\!R^d \setminus \{0\} \qquad (6.4.2)$$

and that itself leads to a similar quadratic form

(*EqPsiQF*)

$$\sum_{j=1}^M \sum_{k=1}^M \alpha_j \alpha_k \widehat{\Psi}(x_j - x_k) = (2\pi)^{-d/2} \int_{I\!R^d} \widehat{\Psi}(\omega) \left| \sum_{j=1}^M \alpha_j e^{ix_j \cdot \omega} \right|^2 d\omega \tag{6.4.3}$$

for another basis function $\widehat{\Psi}$ and a weaker constraint on $\alpha \in I\!R^M$ (or none at all). Furthermore, there should be an easy lower bound

$$\alpha^T A_{X,\Psi} \alpha \geq \sigma \|\alpha\|_2^2$$

for the left–hand side $\alpha^T A_{X,\Psi} \alpha$ of (6.4.3, *EqPsiQF*). Then clearly for all $\alpha \in I\!R^M$ that are admissible,

$$\alpha^T A_{X,\Phi} \alpha \geq \alpha^T A_{X,\Psi} \alpha \geq \sigma \|\alpha\|_2^2,$$

as required. The basic trick of Narcowich and Ward now is to make $A_{X,\Psi}$ diagonally dominant, while $\widehat{\Psi}$ is obtained by chopping off $\widehat{\Phi}$ appropriately near infinity.

Before we proceed any further, here is the main result:

Theorem 6.4.4 *(ThNWLB) Let Φ be a conditionally positive definite function on $I\!R^d$ that satisfies Assumption 6.1.4 (FTAss1). Furthermore, let $X = \{x_1, \ldots, x_M\} \subset I\!R^d$ be any set of Lagrange data locations having separation distance*

$$q := \min_{1 \leq i \neq j \leq M} \|x_i - x_j\|_2.$$

With the function

(EqDefPhi0)

$$\phi_0(r) := \inf_{\|\omega\|_\infty \leq 2r} \widehat{\Phi}(\omega), \tag{6.4.5}$$

the smallest eigenvalue σ of the quadratic form associated to the matrix

$$A_{X,\Phi} = \left(\Phi(x_j - x_k)\right)_{1 \leq j,k \leq M},$$

restricted as usual to the subspace of $I\!R^M$ that contains the coefficient vectors α of functionals $\lambda_{X,M,\alpha} \in \mathcal{P}_\Omega^-$ has the lower bound

(EqCLB)

$$\sigma \geq \frac{1}{2} \frac{\phi_0(K)}{?\ (d/2+1)} \left(\frac{K}{\sqrt{2}}\right)^d \tag{6.4.6}$$

for any $K > 0$ satisfying

(EqKBound)

$$K \geq \frac{4}{q}\left(2\pi ?^2 (d/2+1)\right)^{\frac{1}{d+1}} \tag{6.4.7}$$

or, a fortiori,

(EqKBound2)

$$K \geq \frac{9.005\ d}{q}. \tag{6.4.8}$$

Proof: We start with any $K > 0$ and the characteristic function

$$\chi_K(x) = \left\{ \begin{array}{ll} 1 & \|x\|_2 \leq K \\ 0 & \text{else} \end{array} \right\}$$

of the L_2 ball $B_K(0)$ in $I\!R^d$ with radius K. Then we define

$$\widehat{\Psi}(\omega) := \widehat{\Psi}_K(\omega) := \frac{\phi_0(K)?\ (d/2+1)}{K^d\ \ \pi^{d/2}}(\chi_K * \chi_K)(\omega)$$

and immediately see that the support is

$$\text{supp}\,(\widehat{\Psi}_K) = \left\{x \in I\!R^d : \|x\|_2 \leq 2K\right\} =: B_{2K}(0).$$

We now use the formula (12.3.3, *EqVolBall*) for the volume of the unit ball to get the L_∞ bound

$$\|\chi_K * \chi_K\|_\infty \leq vol(B_K(0)) = K^d \frac{\pi^{d/2}}{?\ (d/2+1)}$$

via the usual convolution integral. We adjusted the factors in the definition of $\widehat{\Psi}$ to guarantee (6.4.2, *EqPhiPsi*) on all of $I\!R^d$.

This is part of what we wanted, but we still have to evaluate Ψ itself or at least to show diagonal dominance of $A_{X,\Psi}$. The radial basis function Ψ_K corresponding to $\widehat{\Psi}_K$ is obtained via the inverse Fourier transform as

$$\begin{aligned} \check{\chi}_K(x) &= \check{\chi}_1(\cdot/K)(x) \\ &= K^d \check{\chi}_1(Kx) \\ &= K^d (K\|x\|)^{-d/2} J_{d/2}(K \cdot \|x\|_2) \\ &= \left(\frac{K}{\|x\|}\right)^{d/2} J_{d/2}(K \cdot \|x\|_2) \end{aligned}$$

using (12.5.5, *EqFTScale*) and (9.2.3, *EqFTCharF*). Then we apply (12.5.3, *EqFTC*) to the convolution to get

$$\begin{aligned}\Psi_K(x) &= \phi_0(K)?\ (d/2+1)\, K^{-d}\pi^{-d/2}(\chi_K * \chi_K)^\vee(x)\\ &= \phi_0(K)?\ \left(\frac{d}{2}+1\right) 2^{d/2}\|x\|^{-d}J^2_{d/2}(K\cdot\|x\|).\end{aligned}$$

Equation (12.3.16, *EqJsqInfty*) yields

$$\Psi_K(0) = \frac{\phi_0(K)}{?\ (d/2+1)}\left(\frac{K}{\sqrt{2}}\right)^d$$

and we assert diagonal dominance of the quadratic form in (6.4.3, *EqPsiQF*) by a suitable choice of K. We have

$$\alpha^T A_{X,\Psi}\alpha \geq \|\alpha\|_2^2\left(\Psi_K(0) - \max_{1\leq j\leq M}\sum_{\substack{k=1\\ k\neq j}}^{M}\Psi_K(x_j-x_k)\right)$$

by Gerschgorin's theorem, and the final bound will be of the form

$$\sigma \geq \frac{1}{2}\Psi_K(0) = \frac{\phi_0(K)}{2?\ (d/2+1)}\left(\frac{K}{\sqrt{2}}\right)^d,$$

because we shall choose K such that

(*EqCBGer*)

$$\max_{1\leq j\leq M}\sum_{\substack{k=1\\ k\neq j}}\Psi_K(x_j-x_k) \leq \frac{1}{2}\Psi_K(0). \tag{6.4.9}$$

This is done by a tricky summation argument of Narcowich and Ward [35](*narcowich-ward:92-1*) using (12.3.15, *EqJsqBound*) which proves (6.4.9, *EqCBGer*) for K satisfying (6.4.7, *EqKBound*). Since the technique is nice and instructive, we repeat it here in full detail.

To proceed towards diagonal dominance of the matrix, we should fix a point $x_j \in X = \{x_1,\ldots,x_M\}$ and exploit the observation that many of the distances $x_j - x_k$ to the remaining points should be large, if the separation distance $q > 0$ does not let two points to be too near to each other. But the number of far-away points will strongly depend on the space dimension d, and we need a precise argument to put the above reasoning on a solid basis.

To this end, define the sets

$$E_n := \{ \ x_k \in X \ : \ nq \leq \|x_j - x_k\|_2 < (n+1)q \ \}$$

for all $n \in I\!N$ and observe that E_1 is empty due to the definition of the separation distance q, which implies

$$\|x_j - x_k\|_2 \geq 2q \text{ for all } 1 \leq j \neq k \leq M.$$

Now we can put a little ball $B_q(x_k)$ of radius q around each of the $x_k \in E_n$. Any two of these balls cannot overlap due to the definition of q. Since none of the x_k is farther away from x_j than $(n+1)q$, the balls are all contained in the ball $B_{(n+2)q}(x_j)$ of radius $(n+2)q$ around x_j. But all of the x_k are at least nq away from x_j, such that their surrounding balls cannot intersect the smaller ball $B_{(n-1)q}(x_j)$ around x_j of radius $(n-1)q$. Adding their volumes using (12.3.3, *EqVolBall*) we get the bound

$$\begin{aligned}
|E_n|\frac{q^d\pi^{d/2}}{?\,(1+d/2)} &\leq \frac{(q(n+2))^d\pi^{d/2}}{?\,(1+d/2)} - \frac{(q(n-1))^d\pi^{d/2}}{?\,(1+d/2)} \\
|E_n| &\leq (n+2)^d - (n-1)^d.
\end{aligned}$$

for the number $|E_n|$ of elements of E_n. If both terms on the right-hand side are expanded with the binomial formula, the leading positive term is $3n^{d-1}$, and all the terms must combine into powers of n with nonnegative factors. Thus we arrive at

$$|E_n| \leq 3n^{d-1}.$$

For points $x_k \in E_n$ we can bound the values of Ψ via (12.3.15, *EqJsqBound*) as follows:

$$\begin{aligned}
\Psi_K(x_j - x_k) &= \phi_0(K)?\left(\frac{d}{2}+1\right)2^{d/2}\|x_j - x_k\|^{-d}J^2_{d/2}(K\cdot\|x_j - x_k\|) \\
&= \phi_0(K)?\left(\frac{d}{2}+1\right)2^{d/2}K^{-1}\|x_j - x_k\|^{-d-1} \\
&\quad\cdot(K\cdot\|x_j - x_k\|_2)J^2_{d/2}(K\cdot\|x_j - x_k\|) \\
&\leq \phi_0(K)?\left(\frac{d}{2}+1\right)2^{d/2}K^{-1}((n-1)q)^{-d-1}\frac{2^{d+2}}{\pi} \\
&= \Psi_K(0)\left(\frac{4}{K(n-1)q}\right)^{d+1}\pi^{-1}?^{\,2}\left(\frac{d}{2}+1\right).
\end{aligned}$$

Now it is time to do the summation over all $k \neq j$, and this summation can be done by summing the points in the sets E_n. This yields

$$\begin{aligned}
\sum_{k\neq j} \Psi_K(x_j - x_k) &= \sum_{n=2}^{\infty} \sum_{x_k \in E_n} \Psi(x_j - x_k) \\
&\leq \Psi_K(0) \left(\frac{4}{Kq}\right)^{d+1} \pi^{-1} ?^2 \left(\tfrac{d}{2}+1\right) \sum_{n=2}^{\infty} 3n^{d-1}(n-1)^{-d-1} \\
&\leq \Psi_K(0) \left(\frac{4}{Kq}\right)^{d+1} \pi^{-1} ?^2 \left(\tfrac{d}{2}+1\right) 6\sum_{n=2}^{\infty} (n-1)^{-2} \\
&\leq \Psi_K(0) \left(\frac{4}{Kq}\right)^{d+1} \pi^{-1} ?^2 \left(\tfrac{d}{2}+1\right) \pi^2 \\
&= \Psi_K(0) \left(\frac{4}{Kq}\right)^{d+1} \pi ?^2 \left(\tfrac{d}{2}+1\right) \\
&\leq \tfrac{1}{2}\Psi_K(0)
\end{aligned}$$

if we choose K according to (6.4.9, *EqCBGer*).

It remains to show that (6.4.8, *EqKBound2*) implies (6.4.7, *EqKBound*). We use a variation of Stirling's formula in the form

$$?(1+x) \leq \sqrt{2\pi x}\, x^x e^{-x} e^{1/12x}, \qquad x > 0$$

to get

$$\begin{aligned}
2\pi ?^2 (d/2+1) &\leq 2\pi^2 d^{d+1} (2e)^{-d} e^{1/3d}, \\
\left(2\pi ?^2 (d/2+1)\right)^{\frac{1}{d+1}} &\leq \frac{d}{2e} \left(4e\pi^2\right)^{\frac{1}{d+1}} e^{\frac{1}{3d(d+1)}} \\
&\leq d\frac{\pi}{\sqrt{e}} \cdot e^{1/6} \leq d \cdot 2.2511
\end{aligned}$$

such that

$$K \geq \frac{9.005}{qd}$$

is satisfactory for all cases. □

We now want to look at the specific cases for applications. From (6.4.6, *EqCLB*) and (6.4.7, *EqKBound*) we see that

$$\sigma = \sigma(q) =\geq \mathcal{O}\left(q^{-d}\phi_0(cd/q)\right)$$

with some positive constant c. Thus we only need to look at the decay of the Fourier transforms to get the asymptotics of σ with respect to $q \to 0$, keeping the space dimension d fixed. Comparison with Table 10 (*TFT*) then yields the results of Table 7 (*TCPDC*).

$\phi(r)$	Lower Bound in $\mathcal{O}$ form for $q \to 0$
r^β	q^β
$r^\beta \log r$	q^β
$(r^2+\gamma^2)^{\beta/2}$	$q^{-d}\exp(-c/q),\ c>0$
$e^{-\beta r^2}$	$q^{-d}\exp(-c/q^2),\ c>0$
$r^\nu K_\nu(r)$	$q^{2\nu}$
$(1-r)_+^2(2+r)$	q
$(1-r)_+^4(1+4r)$	q^3

Table 7: Lower Bounds of Smallest Eigenvalue Based on Lagrange Data with Separation Distance q (*TCPDC*)

6.5 Error Bounds

(*SecEBTrans*) We now want to apply Fourier transform techniques to get error bounds. The starting points are the representations (6.1.1, *DefBil2*) of the bilinear form and (5.1.1, *DefPuxyLag*) of the square of the power function for Lagrange data on $X = \{x_1, \ldots, x_M\} \subset I\!R^d$. These combine into

(*EqPFFT*)

$$\begin{aligned} P_u^2(x) &= (2\pi)^{-d/2}\int_{I\!R^d} \widehat{\Phi}(\omega)\left|e^{-ix\cdot\omega} - \sum_{j=1}^M u_j(x)e^{-ix_j\cdot\omega}\right|^2 d\omega \\ &= (2\pi)^{-d/2}\int_{I\!R^d} \widehat{\Phi}(\omega)\left|1 - \sum_{j=1}^M u_j(x)e^{i(x-x_j)\cdot\omega}\right|^2 d\omega \end{aligned} \qquad (6.5.1)$$

and we insert the real numbers $u_j^h(x)$ into this representation as postulated in Assumption 5.1.6 (*FBAss2)*. Thus we assume (5.1.3, *EqEllgeqm*), (5.1.7, *uDefJx*) to (5.1.10, *uDef3*) to be satisfied.

There are basically two choices to bound the square factor in the integrand of equation (6.5.1, *EqPFFT*), and these two are done after splitting the integration domain into two parts, say I_1 and I_2. Bounding $e^{i(x-x_j)\cdot\omega}$ by one, we can write

$$\int_{I_1} \widehat{\Phi}(\omega) \left| 1 - \sum_{j=1}^{M} u_j^h(x) e^{i(x-x_j)\cdot\omega} \right|^2 d\omega \leq C_3(x,h)^2 \int_{I_1} \widehat{\Phi}(\omega) d\omega$$

if the latter integral exists. This usually is no problem, if I_1 excludes the origin in case of conditionally positive definite functions of positive order. If I_1 is chosen as a function of h to represent a shrinking neighbourhood of infinity, say as the complement of the ball $B_{1/h}(0)$, this integral will nicely decay for $h \to 0$. But since it does not depend on the u_j and their polynomial reproduction properties in $I\!P_\ell^d$, it does not furnish the relevant part of the bound of the power function.

This will be obtained by the second technique. We take the complex-valued Taylor expansion $p_\ell \in I\!P_\ell^1$ of e^{it} at zero, and denote the residual by $r_\ell(t) = e^{it} - p_\ell(t)$. Then the polynomial reproduction property (5.1.9, *uDef2*), when applied to the function $p_\ell((\cdot - x)\cdot\omega)$ for fixed values of x and ω, yields

$$1 = p_\ell(0) = p_\ell((x-x)\cdot\omega) = \sum_{j\in J_x(h)} u_j^h(x) p_\ell((x_j - x)\cdot\omega)$$

and we use Example 12.1.4 (*ExaExpIma*) to bound part of (6.5.1, *EqPFFT*) by

(*EqPFFT2*)

$$(2\pi)^{-d/2} \int_{I_2} \widehat{\Phi}(\omega) \left| \sum_{j=1}^{M} r_\ell((x_j - x)\cdot\omega) \right|^2 d\omega \quad \leq \quad (2\pi)^{-d/2} \frac{(C_2(x,h)h)^\ell}{\ell!} \int_{I_2} \widehat{\Phi}(\omega) \|\omega\|^{2\ell} d\omega. \tag{6.5.2}$$

As explained before, the domain I_2 is either all of $I\!R^d$ or a large neighbourhood of zero that grows towards infinity when $h \to 0$. Thus the bound in (6.5.2, *EqPFFT2*) roughly depends on the smoothness of Φ. It exists on all of $I\!R^d$, if Φ has Fourier transformable derivatives of order up to 2ℓ. Note that (5.1.3, *EqEllgeqm*) and (6.1.7, *BetaZCond*) combine into

$$\beta_0 < 2m \leq 2\ell,$$

making the integral well-defined around zero.

Before we look at single examples, let us check the basic two situations, the first of which is easily obtained by picking $I_2 = I\!R^d$.

Theorem 6.5.3 *(TheEBFT1) If for a specific $\ell \geq m$ the integral of $\widehat{\Phi(\cdot)}\|\cdot\|^{2\ell}$ on $I\!R^d$ is finite, and if Assumptions 5.1.6 (FBAss2) and 6.1.4 (FTAss1) are satisfied, then there is a bound of order ℓ for the power function.* □

If Theorem 6.5.3 (*TheEBFT1*) cannot be applied, there typically is a rather slow algebraic decay of $\widehat{\Phi}$ at infinity, e.g.:

(*EqPhiDecay*)

$$\widehat{\Phi(\omega)} \leq C_4\|\omega\|_2^{-d-\beta_\infty} \tag{6.5.4}$$

with a real number β_∞ that cannot be pushed up to be larger than 2ℓ. We now set $I_2 = B_{1/h}(0)$ and look at the result for sufficiently small h. The first integral will then be of order β_∞, while the second consists of the "inner" part of order 2ℓ and the "outer" part, where the asymptotics of $\widehat{\Phi}$ yield the order $2\ell + (\beta_\infty - 2\ell) = \beta_\infty$. Due to $2\ell \geq \beta_\infty$ we are left with an overall order β_∞. We summarize:

Theorem 6.5.5 *(TheEBFT2) Let $\widehat{\Phi}$ satisfy Assumption 6.1.4 (FTAss1) and have a decay like (6.5.4, EqPhiDecay) with $\beta_\infty \leq 2\ell$ near infinity. If the data satisfy Assumption 5.1.6 (FBAss2) near x, then the power function has a bound of order $\beta_\infty/2$ at x.* □

Example 6.5.6 *(ExaTPSFT1) The typical case for Theorem 6.5.5 (TheEBFT2) is furnished by thin-plate or polyharmonic functions $\Phi(x) = \phi(\|x\|_2) = \|x\|_2^\beta \log \|x\|_2$ or $\|x\|_2^\beta$, dependent on β being an even integer or not. In both cases the Fourier transform in $I\!R^d$ is $\|\omega\|_2^{-d-\beta}$ up to multiplicative constants, and we have to set $\beta = \beta_0 = \beta_\infty \leq 2\ell$. Then we get a bound of order $\beta/2$ for the power function, as in section 5.5 (*hrhodef*).*

Most other cases are applications of Theorem 6.5.3 (*TheEBFT1*), because one can take $2\ell < \beta_\infty \geq \beta_0$ to get order ℓ for the pointwise bound of the power function. The only drawback is the case $2\ell = \beta_i nfty$, which needs a similar split as the thin-plate-spline case.

Example 6.5.7 *(EBSob) Sobolev radial basis functions have the native space $W_2^k(I\!R^d)$ which can be characterized as the space of functions with classical Fourier transforms in a weighted L_2 space with weight $(1 + \|\cdot\|_2^2)^k$. Thus the reciprocal of this function must be (up to a factor) the Fourier transform of the radial function generating the space as a reproducing kernel. We thus can set $\widehat{\Phi} = (1 + \|\cdot\|_2^2)^{-k}$ and use Theorem 6.5.3 (TheEBFT1) for any $\ell < k - d/2 = \beta_\infty/2$. The case $\ell = k - d/2$ is handled again by splitting the integral. This yields terms of type $\mathcal{O}(h^{2k-d}) + \mathcal{O}(h^{2\ell})(\mathcal{O}(1) + \mathcal{O}(h^{2k-2\ell-d}))$, and we get a pointwise bound of order $k - d/2$, as expected.*

6.6 Error Bounds and Scaling

(*SecError*) Here we study the effect of scaling a translation-invariant basis function $\Phi : I\!R^d \to I\!R^d$ by some factor δ in the sense

$$\Phi_\delta(x, y) := \Phi((x - y)/\delta) \text{ for all } x, y \in I\!R^d.$$

We assume that $\Phi_1(x, y) = \Phi(x - y)$ is conditionally positive definite with respect to some finite-dimensional space $\mathcal{P}$, and the latter should be invariant under scaling of arguments. From (3.3.2, *DefBil*) we see that

(*DefBildelta*)

$$\begin{array}{rcl} (\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta})_{\Phi_\delta} & = & \sum_{j=1}^{M}\sum_{k=1}^{M} \alpha_j\beta_k\Phi_\delta(x_j, y_k) \\ & = & \sum_{j=1}^{M}\sum_{k=1}^{M} \alpha_j\beta_k\Phi((x_j - y_k)/\delta) \\ & = & (\lambda_{X/\delta,M,\alpha}, \lambda_{Y/\delta,N,\beta})_{\Phi_1}. \end{array} \quad (6.6.1)$$

The condition $\lambda_{X,M,\alpha} \in \mathcal{P}_\Omega^-$ implies $\lambda_{X/\delta,M,\alpha} \in \mathcal{P}_\Omega^-$ due to the invariance of $\mathcal{P}$ under scaling. Thus we have

.....

7 Special Theory

(*SecST*) Here we introduce general transforms and generalize the results that we had on $I\!R^d$. We start with generalizing the notion of a transform in order to cover some other cases we consider in some detail later:

1. Fourier series on $[0, 2\pi]^d$,
2. General expansions in orthogonal series,
3. Harmonic analysis on locally compact topological groups.

It will turn out that certain results can be formulated for general transforms, while others take advantage of the special structure of the underlying space.

7.1 Results for General Transforms

(*SecT*) This section covers the necessary results about general transforms. The applications except for R^d will follow later. We start from the general setting and add the specific details later.

7.1.1 General Transforms

(*SecGTR*) Here we formulate the general notions that apply to all kinds of transforms that we consider later. The setting is general enough to allow **generalized** transforms in addition to the classical ones. This turns out to be absolutely necessary even in the simple case of Fourier transforms on $\mathbb{R}^d$. For this reason we do not rely on other sources on transforms.

Assume that our basic space $\mathcal{G}$ with positive definite bilinear form $(\cdot,\cdot)_\Phi$ and nullspace $\mathcal{P}$ is a space of real-valued functions on some domain Ω. Forget about $\Phi, \mathcal{G}$, and $\mathcal{P}$ for a moment, and concentrate on Ω.

Assumption 7.1.1 *(TAss1) For a specific space $\mathcal{S}$ of real-valued* **test functions** *on some domain Ω there is a linear and injective* **transform mapping**

$$g \mapsto \hat{g} \; : \; \mathcal{S} \to L_{2,\sigma}(D)$$

whose values are complex-valued functions on some domain D that carries a measure σ such that the space

$$L_{2,\sigma}(D) := \left\{ f \; : \; D \to \mathbb{C} \; : \; \int_D |f|^2 d\sigma \; < \infty \right\}$$

is well-defined and a Hilbert space over $\mathbb{C}$ with inner product

$$(u,v)_{L_{2,\sigma}(D)} := \int_D u\overline{v} d\sigma \text{ for all } u,v \in L_{2,\sigma}(D).$$

In particular, the measure σ can be Lebesgue measure on $D = \mathbb{R}^d$ for the classical multivariate Fourier transform, or Haar measure on a locally compact topological group, or plain summation for series transforms, e.g.: $D = \mathbb{Z}^d$ for Fourier series on $\Omega = [0, 2\pi]^d$. Note that the transform domain D and its measure σ are independent of the functions Φ that we are going to consider, but they will crucially depend on Ω. We shall often write $(\cdot,\cdot)_2$ as shorthand for the above inner product, and we use the phrase **almost everywhere** to stand for "on D except for a set of σ-measure zero". Assumption 7.1.1 (*TAss1*) is usually satisfied by proper definition of D, σ, and the transform mapping. Injectivity of the latter is often proved by an inverse transform.

Assumption 7.1.2 *(TAss2) The space $L_{2,\sigma}(D)$ contains the image of the space $\mathcal{S}$ of test functions under the transform mapping as a dense subspace and coincides with its closure under the inner product $(\cdot,\cdot)_{L_{2,\sigma}(D)}$.*

This makes sure that the test function space $\mathcal{S}$ is rich enough to generate all of $L_{2,\sigma}(D)$ by continuity arguments acting on transforms.

Assumption 7.1.3 *(TAss3) There is a 1-1 correspondence between L_2 spaces on Ω and D in the sense that there is a measure ω on Ω such that the spaces $L_{2,\sigma}(D)$ and $L_{2,\omega}(\Omega)$ are isometrically isomorphic under the transform mapping:*

(Planch)

$$(f,g)_{L_{2,\omega}\Omega} := \int_\Omega f\overline{g}d\omega = (\widehat{f},\widehat{\overline{g}})_{L_{2,\sigma}(D)}. \tag{7.1.4}$$

Identities like (7.1.4, *Planch*) are usually called **Plancherel's equation**. Of course, one could use the structure on $L_{2,\sigma}(D)$ to define an inner product for functions on Ω by using (7.1.4, *Planch*) without the representation via integrals as a definition. Thus the actual meaning of Assumption 7.1.3 (*TAss3)* is that this abstract inner product can be respresented as a standard L_2 inner product.

7.1.2 Spaces Induced by Basis Functions

We restrict ourselves to basis functions Φ that satisfy

Assumption 7.1.5 *(PFTAss1) The conditionally positive definite function $\Phi : \Omega\times\Omega \to I\!R$ has an associated real-valued nonnegative function $\widehat{\Phi}$ which is defined and positive almost everywhere on the transform domain D.*

For reasons to become apparent later, we do not require $\widehat{\Phi}$ to be the image of Φ under the transform mapping, since we shall encounter cases where Φ is not in the domain of the transform. One should rather consider $(\widehat{\Phi})^{-1}$ as a weight function on D. But there will also be cases where actually $\widehat{\Phi}$ is the transform of Φ, thus the notation. The relation between Φ and $\widehat{\Phi}$ will be clarified after introducing some additional notation.

We use $(\widehat{\Phi})^{-1}$ as a weight function to define the operator

$$L_\Phi \ : \ g \mapsto \frac{\widehat{g}}{\sqrt{\widehat{\Phi}}}.$$

To turn it into a continuous map with image in $L_{2,\sigma}(D)$, we restrict its domain to the subspace

$$\mathcal{S}_\Phi := \{u \in \mathcal{S} \ : \ L_\Phi u \in L_{2,\sigma}(D)\}$$

of the space $\mathcal{S}$ of test functions on D. We now can define an inner product

(*PTIP*)

$$(f,g)_\Phi = \int_D \widehat{f}(\widehat{\Phi})^{-1}\widehat{\overline{g}}d\sigma \tag{7.1.6}$$

on all $f, g \in \mathcal{S}_\Phi$.

We are now ready to link $\widehat{\Phi}$ to Φ and its native space $\mathcal{G} = \mathcal{P} + \mathcal{F}$ by the requirement

Assumption 7.1.7 *(PTAss2) The closure of $\mathcal{S}_\Phi$ under the inner product (7.1.6, PTIP) coincides with the Hilbert space $\mathcal{F}$.*

Then the mapping L_Φ can be identified with its continuous extension to all of $\mathcal{F}$, and it can be further extended to $\mathcal{G}$ by defining it as being zero on $\mathcal{P}$. The image of $\mathcal{F}$ under L_Φ is a closed Hilbert subspace of $L_{2,\sigma}(D)$, and we shall require some additional work in special cases to prove

Assumption 7.1.8 *(PTAss3) The mapping*

(LSurj)

$$L_\Phi \ : \ \mathcal{G} \to L_{2,\sigma}(D) \tag{7.1.9}$$

as the canonical extension of

$$L_\Phi(g) := \frac{\widehat{g}}{\sqrt{\widehat{\Phi}}}$$

for $g \in \mathcal{S}_\Phi$ is surjective.

The extension allows to define a **generalized transform** on the space $\mathcal{G}$ via

$$\widehat{g} := \sqrt{\widehat{\Phi}} L_\Phi(g),$$

and these are by definition in the weighted L_2 space

$$L_{2,\sigma,1/\widehat{\Phi}}(D) := \left\{ u \ : \ \int_D \frac{|u(\omega)|^2}{\widehat{\Phi}(\omega)} d\sigma(\omega) < \infty \right\}.$$

7.2 Theory on the Torus using Fourier Series

7.3 Theory on Spheres using Expansions

7.4 Lower Bounds for Eigenvalues

(*SecLBE*) Here we proceed to prove lower bounds of the form (4.6.4, *GBound*) for the smallest eigenvalue of the matrix occurring in optimal recovery problems with Lagrange data. We had to postpone them until now, because they require transforms.

7.5 Generalizations of Results Using Transforms

8 Theory on Grids

Using Fourier transforms, we treat the case of gridded data $h\mathbb{Z}^d$ here.

8.1 Sampling Theory

Theorem 8.1.1 *(MST) Classical Sampling Theorem......*

8.2 Strang-Fix Theory

8.3 Application to Radial Basis Functions

8.4 Shift Invariant Spaces

9 Construction of Positive Definite Functions

(*SecCCPD*) This section is intended to give the proofs of conditional positive definiteness of the classical radial basis functions. We include a toolbox of operators on radial functions that allow the construction of compactly supported positive definite functions. Except for the first subsection, we shall rely on properties of Fourier transforms as compiled in section 12.5 (*SecFTRd).*

9.1 General Construction Techniques

(*SecGCT*) This section is planned to give an overview of methods for the construction of new conditionally positive definite functions from existing ones. For the time being, we restrict ourselves to translation-invariant cases in $\mathbb{R}^d$.

9.1.1 Simple Cases of Positive Semidefinite Functions

(*SecSCPSDF*) Let us start with the function

$$\Phi_\zeta(x, y) := \cos((x - y) \cdot \zeta),$$

where $\zeta \in I\!R^d$ is fixed and $x \cdot \zeta$ stands for the inner product on $I\!R^d$. Using the terminology of (3.3.2, *DefBil*) on page 55, we get

$$\begin{aligned}
(\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta})_{\Phi_\zeta} &= \sum_{j=1}^{M}\sum_{k=1}^{N} \alpha_j\beta_k \cos((x_j - y_k)\cdot\zeta) \\
&= \sum_{j=1}^{M}\sum_{k=1}^{N} \alpha_j\beta_k \cos(x_j\cdot\zeta)\cos(y_k\cdot\zeta) + \sin(x_j\cdot\zeta)\sin(y_k\cdot\zeta) \\
&= \left(\sum_{j=1}^{M}\alpha_j\cos(x_j\cdot\zeta)\right)\left(\sum_{k=1}^{N}\beta_k\cos(y_k\cdot\zeta)\right) + \\
&\quad + \left(\sum_{j=1}^{M}\alpha_j\sin(x_j\cdot\zeta)\right)\left(\sum_{k=1}^{N}\beta_k\sin(y_k\cdot\zeta)\right)
\end{aligned}$$

and this is a well-defined bilinear form which is positive semidefinite because of

$$(\lambda_{X,M,\alpha}, \lambda_{X,M,\alpha})_{\Phi_\zeta} = \left(\sum_{j=1}^{M}\alpha_j\cos(x_j\cdot\zeta)\right)^2 + \left(\sum_{j=1}^{M}\alpha_j\sin(x_j\cdot\zeta)\right)^2.$$

If we allow complex-valued functions temporarily, we can generalize the above case to

$$\Phi_\zeta(x,y) = e^{i(x-y)\cdot\zeta},$$

which yields a positive semidefinite sesquilinear form with

$$(\lambda_{X,M,\alpha}, \lambda_{X,M,\alpha})_{\Phi_\zeta} = \left|\sum_{j=1}^{M}\alpha_j e^{ix_j\cdot\zeta}\right|^2.$$

These quadratic forms are not positive definite, if ζ happens to be a zero of the analytic function

$$\lambda_{X,M\alpha}\Phi_\zeta(\cdot, z) = \sum_{j=1}^{M}\alpha_j e^{ix_j\cdot z},$$

but we shall overcome this drawback by integration over ζ.

9.1.2 Elementary Operations

(*SecCTElOps*) It is very easy to see that (conditionally) positive (semi-) definite functions on Ω form a **cone** in the space of all functions on $\Omega \times \Omega$. In particular, if Φ and Ψ are (conditionally) positive (semi-) definite, so are

$\alpha\Phi + \beta\Psi$ for $\alpha, \beta > 0$. Furthermore, if a family Φ_ζ of (conditionally) positive (semi-) definite functions can be integrated against a positive function $w(\zeta)$, the result

$$\Phi(x, y) := \int w(\zeta)\Phi_\zeta(x, y)d\zeta$$

will again be (conditionally) positive (semi-) definite.

9.1.3 Autocorrelation Method

(*SecCTAut*) If we cannot start with a (conditionally) positive (semi-) definite function but have an arbitrary function $\Psi \in L_2(\mathbb{R}^d)$, we can form the **autocorrelation function**

$$\Phi(x, y) := \int_{\mathbb{R}^d} \Psi(x - z)\Psi(y - z)dz.$$

This always yields a symmetric positive semidefinite function which even is positive definite, if all translates $\Phi(x_j - \cdot)$ for different points x_j are linearly independent in $L_2(\mathbb{R}^d)$.

9.1.4 Integration Method

(*SecCTInt*) The previous method easily generalizes for any Ω. For any function Ψ on $\Omega \times \Pi$ one can formally consider

$$\Phi(x, y) := \int_\Pi \Psi(x, \zeta)\Psi(y, \zeta)w(\zeta)d\zeta$$

with a positive weight function w on Π. If the integral is well-defined, the result will be a symmetric positive semidefinite function on Ω.

9.2 Construction of Positive Definite Radial Functions on $\mathbb{R}^d$

(*SecCTPDRF*) This subsection contains tools from [47](*wu:95-2*) as generalized in [43](*schaback-wu:95-1*) for the construction of positive definite radial functions on $\mathbb{R}^d$. We start with the standard reduction of d-variate Fourier transforms of radial functions to Hankel transforms of univariate functions. Introducing $t = r^2/2$ as a new variable, two such transforms for different space dimensions are related to each other by a simple univariate differential or integral operator that preserves compact supports. This fundamental trick of Z. Wu then opens up the way for the easy derivation of various compactly supported radial basis functions.

9.2.1 Hankel Transforms

(*SecHT*) We assume a radial function $\Phi(\cdot) = \phi(\|\cdot\|_2)$ to be given such that $\phi : I\!R_{>0} \to I\!R$ has some decay towards infinity that we are going to quantify later. Let us formally look at the Fourier transform formula (12.5.2, *FT*) and simplify it, using radiality, and introducing polar coordinates for x:

$$\begin{aligned}
\hat{\Phi}(\omega) &= (2\pi)^{-d/2}\int_{I\!R^d}\Phi(x)e^{-ix\cdot\omega}dx \\
&= (2\pi)^{-d/2}\int_{I\!R^d}\phi(\|x\|_2)e^{-ix\cdot\omega}dx \\
&= (2\pi)^{-d/2}\int_0^\infty \phi(r)r^{d-1}\int_{\|y\|_2=1} e^{-ir\|\omega\|_2 y\cdot\frac{\omega}{\|\omega\|_2}}dydr.
\end{aligned}$$

This contains the function $F(r\|\omega\|_2, d)$ defined in (12.3.5, *EqDefFtd*) by the integral

$$F(t,d) := \int_{\|y\|_2=1} e^{-ity\cdot z}dy$$

for $t \geq 0$ and some $\|z\|_2 = 1,\ z \in I\!R^d$. Using its representation (12.3.7, *EqFtdRep*) via a Bessel function, we get the very important equation

(*EqFTR*)

$$\begin{aligned}
\hat{\Phi}(\omega) &= (2\pi)^{-d/2}\sigma_{d-2}\int_0^\infty \phi(r)r^{d-1}\frac{?\,(\frac{d-1}{2})?\,(\frac{1}{2})}{(r\|\omega\|_2/2)^{(d-2)/2}}J_{(d-2)/2}(r\|\omega\|_2)dr \\
&= \|\omega\|_2^{-(d-2)/2}\int_0^\infty \phi(r)r^{d/2}J_{(d-2)/2}(r\|\omega\|_2)dr.
\end{aligned} \tag{9.2.1}$$

that allows the Fourier transform of a radial function to be written as a univariate **Hankel transform**. Equation (9.2.1, *EqFTR*) implies that the Fourier transform of a radial function Φ is again a radial function. It holds also for $d = 1$, as can be proven by direct calculation and

(*JBh*)

$$\sqrt{\frac{\pi}{2z}}J_{-1/2}(z) = \frac{\cos z}{z}. \tag{9.2.2}$$

This equation is not directly compatible with (12.3.6, *JBF*), because the latter does not exist for $\nu = -1/2$. But we can use the usual power series representation (12.3.8, *JBFP*) of Bessel functions to get (9.2.2, *JBh*) from (12.3.10, *JBh2*).

We conclude this section by evaluating the Fourier transform of the characteristic function χ_1 of the unit ball in $I\!R^d$. This is needed in the proof of

Theorem 6.4.4 (*ThNWLB*). In particular, we apply (12.3.14, *CSPnuFT*) and get

(*EqFTCharF*)

$$\begin{aligned}\widehat{\chi_1}(\omega) &= \|\omega\|_2^{-(d-2)/2}\int_0^1 r^{d/2}J_{(d-2)/2}(r\|\omega\|_2)dr \\ &= \|\omega\|_2^{-d/2}J_{d/2}(\|\omega\|_2).\end{aligned} \tag{9.2.3}$$

Considered as a univariate radial function, this is an entire analytic function of exponential type that we shall meet again in the next section.

9.2.2 Change of Variables

(*SecCTHTCV*) We now introduce $t = r^2/2$ as a new variable, writing a radial basis function Φ as

(*EqCHV*)

$$\Phi(\cdot) = \phi(\|\cdot\|_2) = f(\|\cdot\|_2^2/2), \tag{9.2.4}$$

and we shall use Latin characters $f, g, \ldots$ to distinguish the transformed functions from the original ones ϕ, ψ, etc. Note that going over from Φ to ϕ and further to f loses the information on the dimension of the space that we want to work on. But we can take advantage of this loss and write dimension-dependent operations like Fourier transforms as univariate operations with a scalar parameter d.

We keep the dimension d in mind and rewrite the d-variate Fourier transform equation (9.2.1, *EqFTR*) in terms of the transformed function f to get

$$\begin{aligned}\widehat{\Phi}(\omega) &= \|\omega\|_2^{-\frac{d\perp 2}{2}}\int_0^\infty f(s^2/2)s^{d/2}J_{\frac{d\perp 2}{2}}(s\cdot\|\omega\|_2)ds \\ &= \int_0^\infty f\left(\frac{s^2}{2}\right)\left(\frac{s^2}{2}\right)^{\frac{d\perp 2}{2}}\left(\frac{s\cdot\|\omega\|_2}{2}\right)^{-\frac{d\perp 2}{2}}J_{\frac{d\perp 2}{2}}(s\cdot\|\omega\|_2)s\,ds \\ &= \int_0^\infty f\left(\frac{s^2}{2}\right)\left(\frac{s^2}{2}\right)^{\frac{d\perp 2}{2}}H_{\frac{d\perp 2}{2}}\left(\frac{s^2}{2}\cdot\frac{\|\omega\|_2^2}{2}\right)s\,ds\end{aligned}$$

with the functions J_ν and H_ν defined by

$$\left(\frac{z}{2}\right)^{-\nu}J_\nu(z) = H_\nu(z^2/4) = \sum_{k=0}^{\infty}\frac{(-z^2/4)^k}{k!?\,(k+\nu+1)}$$

for $\nu \in \mathbb{C}$ as in (12.3.9, *EqHnuDef*). If we substitute $t = s^2/2$, we find

(*EqHnuProp*)

$$\begin{array}{rcl}\hat{\Phi}(\omega) & = & \displaystyle\int_0^\infty f(t)t^{\frac{d-2}{2}} H_{\frac{d-2}{2}}\left(t\cdot\frac{\|\omega\|^2}{2}\right)dt \\ & =: & \left(F_{\frac{d-2}{2}}f\right)(\|\omega\|^2/2)\end{array} \tag{9.2.5}$$

with the general operator

(*EqFnuGen*)

$$(F_\nu f)(r) := \int_0^\infty f(t)t^\nu H_\nu(tr)dt. \tag{9.2.6}$$

This operator is formally defined for all $\nu > -1$ and sufficiently nice functions f, but we can extend it to all $\nu \in I\!R$, if we omit terms in the series of H_ν that have a singularity of the Gamma function in their denominator. However, we want to check its domain of definition with respect to functions f on $I\!R_{>0}$ for $\nu > -1$. Near zero, the function $f(t)t^\nu$ should be absolutely integrable, because the analyticity of H_ν causes no problems at zero. For large ν this allows a moderate singularity of f at zero. Near infinity we have to check the decay of H_ν. But since the Bessel functions J_ν have a $\mathcal{O}(t^{-1/2})$ behaviour for $t \to \infty$ due to (12.3.18, *EqBFBound2*), we see that $H_\nu(t)$ decays like $t^{-\nu/2-1/4}$. Thus we require integrability of $f(t)t^{\nu/2-1/4}$ at infinity for $\nu > -1$. Since we do not need the weakest conditions, we can simply assume

(*Eqbb*)

$$f(t)t^\nu \in L_1(I\!R_{>0}). \tag{9.2.7}$$

Note that both F_ν and H_ν generalize to arbitrary $\nu \in I\!R$, provided that certain restrictions on f like (9.2.7, *Eqbb*) hold. Furthermore, by symmetry of radial functions and our definition of Fourier transforms we have

$$F^{-1}_{\frac{d-2}{2}} = F_{\frac{d-2}{2}} \qquad \text{for } d \in I\!N$$

on sufficiently smooth functions with sufficient decay. We shall see later that this generalizes to $F_\nu^{-1} = F_\nu$ for all $\nu \in I\!R$, wherever both operators are defined. Please keep in mind that the parameter ν is related to the space dimension d via $\nu = (d-2)/2$. We shall work with ν instead of d for notational simplification. Furthermore, we consider a space $\mathcal{S}_{rad}$ of **tempered radial functions**. It could be defined as a subspace of the space $\mathcal{S}$ of d-variate tempered test functions, comprising all radial test functions after introducing $\|x\|_2^2/2$ as a new variable. However, we prefer to define it as the space of real-valued functions on $[0, \infty)$ that are infinitely differentiable such that all derivatives vanish faster than any polynomial at infinity. Taking derivatives of (9.2.4, *EqCHV*), one can easily see that this yields a subspace of radial test functions on $I\!R^d$ for all space dimensions d. Conversely, any radial test function Φ in the form (9.2.4, *EqCHV*) yields a function f that is in $\mathcal{S}_{rad}$. To see this one can proceed inductively, using

$$\frac{\partial^m}{\partial \omega_j^m}\Phi(\omega) = f^{(m)}(\|\omega\|_2^2/2)\omega_j^m + \text{ lower derivatives with polynomial factors.}$$

Thus the two notions of $\mathcal{S}$ coincide, and each radial function which yields a test function for a specific space dimension will provide a test function for any dimension. Thus $\mathcal{S}_{rad}$ is the proper space to define the operators F_ν on, and it clearly contains e^{-r}, which can easily proven to be a fixed point of any F_ν, using the definitions (12.3.9, *EqHnuDef*) of H_ν and (12.3.1, *GammaDef*) of the Gamma function.

9.2.3 Calculus on the Halfline

(*SecCoHL*) In the space $\mathcal{S}_{rad}$ we can introduce a quite useful generalization of the classical calculus operations. We start with the family of operators

(*EqIaDef*)

$$I_\alpha(f)(r) := \int_0^\infty f(s)\frac{(s-r)_+^{\alpha-1}}{?\,(\alpha)}ds \tag{9.2.8}$$

on $\mathcal{S}_{rad}$ for all $\alpha > 0$. The simplest special case is

$$I_1(f)(r) := \int_r^\infty f(s)ds$$

with the inverse

$$I_{-1}(f)(r) := -f'(r).$$

Note that this operation implies that the resulting function vanishes at infinity, and thus there is no additive constant in the integration. Furthermore, the identity

$$Id = I_1^n \circ I_{-1}^n$$

is Taylor's formula at infinity, as follows from (9.2.8, *EqIaDef*). The identity (12.3.2, *EqGxy*) allows a direct proof of the property

(*EqIab*)

$$I_\alpha \circ I_\beta = I_{\alpha+\beta} \tag{9.2.9}$$

for all $\alpha,\ \beta > 0$ by application of Fubini's theorem. Differentiation and integration by parts imply

$$\begin{array}{rcll} I_{-1}^n \circ I_\alpha & = & I_{\alpha-n} & 0 < \alpha < n \\ I_{n+\alpha} \circ I_{-1}^n & = & I_\alpha & \alpha > 0, n > 0. \end{array}$$

By $I_\alpha = I_\alpha \circ I_n \circ I_{-1}^n = I_n \circ I_\alpha \circ I_{-1}^n$ we get

$$I_{-1}^n \circ I_\alpha = I_\alpha \circ I_{-1}^n,$$

and this suffices to prove that (9.2.9, *EqIab*) holds for all $\alpha, \beta \in I\!R$ if we define

$$\begin{array}{rcl} I_0 & := & Id \\ I_{-n} & := & I_{-1}^n,\ n > 0 \\ I_\alpha & := & I_{\alpha-\lfloor\alpha\rfloor} \circ I_{\lfloor\alpha\rfloor} \end{array}$$

for the remaining cases of α. Altogether, we have

Theorem 9.2.10 *(TheIab) The operators I_α on $\mathcal{S}_{rad}$ form an abelian group under composition which is isomorphic to $I\!R$ under "+" via $\alpha \mapsto I_\alpha$.*

Proof: The remaining things are easy to prove using the above rules. □

Let us do some simple examples of differentiation and integration of fractional order. The independent variable will be denoted by t, and we indicate the domain of validity of the different cases, because we do not restrict ourselves to tempered radial functions.

$$\begin{array}{rcll} I_\alpha(f(t+x))(r) & = & I_\alpha(f(t))(r+x) & \alpha \in I\!R,\ x \geq 0 \\ I_\alpha(f(tx))(r) & = & x^{-\alpha} I_\alpha(f(t))(rx) & \alpha \in I\!R,\ x \geq 0 \\ I_\alpha(e^{-st})(r) & = & s^{-\alpha} e^{-sr} & \alpha \in I\!R,\ s > 0 \\ I_\alpha(t^{-\beta} ?\,(\beta))(r) & = & r^{-(\beta-\alpha)} ?\,(\beta - \alpha) & \beta > 0,\ \alpha < \beta \\ I_\alpha((x+t)^{-\beta} ?\,(\beta))(r) & = & (x+r)^{-(\beta-\alpha)} ?\,(\beta-\alpha) & \beta > 0,\ \alpha < \beta,\ x > 0 \\ I_\alpha\left(\dfrac{(s-t)_+^{\beta-1}}{?\,(\beta)}\right)(r) & = & \dfrac{(s-r)_+^{\alpha+\beta-1}}{?\,(\alpha+\beta)} & \beta > 0,\ \alpha+\beta > 0 \end{array}$$

We shall make specific use of the "semi-integration" operator and its inverse, the "semi-differentiation", as given by

(*EqDefI12*)

$$\begin{array}{rcl} I_{1/2}(f)(r) & = & \displaystyle\int_r^\infty \frac{f(s)}{\sqrt{\pi(s-r)}} ds \\ I_{-1/2}(f)(r) & = & \displaystyle -\int_r^\infty \frac{f'(s)}{\sqrt{\pi(s-r)}} ds \\ & = & I_{1/2} \circ I_{-1}(f)(r), \end{array} \qquad (9.2.11)$$

that are inverses of each other.

A very simple representation of the operators I_α is possible via the **Laplace transform**

(*EqLapDef*)

$$L(\varphi)(r) := \int_0^\infty \varphi(s)e^{-rs}ds \tag{9.2.12}$$

which exists classically for any continuous function φ on $[0,\infty)$ that grows at most polynomially towards infinity. For the time being, we ignore the more general definitions of Laplace transforms and observe that the action of I_α can be written down as

$$I_\alpha(L(\varphi)(\cdot)) := L(\varphi(\cdot)(\cdot)^{-\alpha}),$$

where all real α are formally possible (provided that φ behaves nicely enough).

9.2.4 Basic Transitions

(*SecBT*) The main advantage of $\mathcal{S}_{rad}$ and the definition (9.2.6, *EqFnuGen*) of the radial Fourier transform using (9.2.5, *EqHnuProp*) is that we can compare Fourier transforms for various dimensions, while working on a simple space of univariate functions. But the most surprising fact, as discovered by Wu, shows up when we simply take the derivative of $F_\nu(f)(r)$. We use (12.3.11, *EqHnuDer*) to get

(*EqDFF*)

$$\begin{aligned}
-\frac{d}{dr}F_\nu(f)(r) &= (I_{-1}\circ F_\nu)(f)(r)\\
&= -\frac{d}{dr}\int_0^\infty f(t)t^\nu H_\nu(rt)dt\\
&= -\int_0^\infty f(t)t^\nu \frac{d}{dr}H_\nu(rt)dt\\
&= \int_0^\infty f(t)t^{\nu+1}H_{\nu+1}(rt)dt\\
&= F_{\nu+1}(f)(r).
\end{aligned} \tag{9.2.13}$$

Going back to $\nu=(d-2)/2$, we see that the $(d+2)$-variate Fourier transform of a radial function after the substitution (9.2.4, *EqCHV*) is nothing else than the negative univariate derivative of the d-variate Fourier transform after

(9.2.4, *EqCHV*). We shall generalize the above identity later to $I_\alpha \circ F_\nu = F_{\nu-\alpha}$ on $\mathcal{R}$, but we already know that $I_1 \circ F_\nu = F_{\nu+1}$ allows to proceed from $(d+2)$-variate radial Fourier transforms to d-variate Fourier transforms by univariate integration.

Let us apply (12.3.12, *EqHnuDerNu*) to get another identity on tempered functions:

(*EqFIF*)

$$\begin{aligned} F_\nu(-f')(r) &= \int_0^\infty -f'(s)s^\nu H_\nu(sr)ds \\ &= \int_0^\infty f(s)s^{\nu-1}H_{\nu-1}(sr)dsdt \qquad (9.2.14) \\ &= = F_{\nu-1}(f)(r). \end{aligned}$$

This will generalize to $F_\nu \circ I_\alpha = F_{\nu+\alpha}$ and is a trivial consequence of $I_\alpha \circ F_{\nu+\alpha} = F_\nu$ and $F_\mu^2 = Id$, if the latter holds in general.

Note that in both cases we have operators that preserve compact supports. The integral operator even preserves nonegativity (it is a **monotone operator**). The explicit construction of compactly supported radial functions relies heavily on these features. But we also want to proceed from d-variate Fourier transforms to $(d+1)$- or $(d-1)$-variate Fourier transforms. This will be achieved by the operator $I_{1/2}$ and its inverse from (9.2.11, *EqDefI12*). We shall treat this problem in general, comparing two arbitrary instances F_ν and F_μ.

9.2.5 Identities for Transforms, First Version

(*SecIfT1*) We can easily evaluate the action of the Fourier operator on the Laplace transform as

$$\begin{aligned} F_\nu(L(\varphi))(r) &= \int_0^\infty s^\nu H_\nu(sr)\int_0^\infty \varphi(t)e^{-st}dtds \\ &= \int_0^\infty \varphi(t)\int_0^\infty s^\nu H_\nu(sr)e^{-st}dsdt \\ &= \int_0^\infty \varphi(t)t^{-\nu-1}\int_0^\infty x^\nu H_\nu(xr/t)e^{-x}dxdt \\ &= \int_0^\infty \varphi(t)t^{-\nu-1}e^{-r/t}dt \\ &= \int_0^\infty \varphi(1/s)s^{\nu-1}e^{-sr}ds \\ &= L\left(\varphi(1/\cdot)(\cdot)^{\nu-1}\right). \end{aligned}$$

Then, again as formal operations,

$$\begin{aligned} F_\nu(L(\varphi(\cdot))) &= L\left(\varphi(1/\cdot)(\cdot)^{\nu-1}\right) \\ &= I_{\mu-\nu}L\left(\varphi(1/\cdot)(\cdot)^{\mu-1}\right) \\ &= I_{\mu-\nu}F_\mu(L(\varphi(\cdot))), \end{aligned}$$

$$\begin{aligned} F_\nu(F_\mu(L(\varphi(\cdot)))) &= F_\nu\left(L\left(\varphi(1/\cdot)(\cdot)^{\mu-1}\right)\right) \\ &= L\left(\varphi(\cdot)(\cdot)^{-\mu+1}(\cdot)^{\nu-1}\right) \\ &= I_{\mu-\nu}(L(\varphi(\cdot))), \end{aligned}$$

as expected. Note that this implies $F_\nu^2 = Id$ for all ν. All of these identities are valid at least on Laplace transforms of functions φ that vanish faster than any polynomial at zero and at infinity, but continuity arguments can be used to enlarge the scopes.

9.2.6 Identities for Transforms, Second Version

(*SecIfT2*) The previous section showed that the identity

$$F_\nu \circ F_\mu = I_{\mu-\nu}$$

holds for all $\mu, \nu \in I\!R$ on a small space of functions, and where I_α is an operator that roughly does α-fold integration for $\alpha \in I\!R$. We now want to make this more precise and explicit. In particular, we assert $F_\nu^2 = Id$ for all ν, which we only know for $\nu \in \frac{1}{2}Z\!\!Z_{>-2}$. Furthermore, we want to use our explicit representations for the operators I_α.

To proceed towards inversion of the operator F_ν, let us start calculating the Fourier transform of the simplest compactly supported function, i.e.: a truncated power. The outcome is somewhat surprising, because we run into the function H_ν again:

Lemma 9.2.15 *(Leml2) For $\nu > \mu > -1$ and all $s, r \geq 0$ we have*

$$F_\mu \left(\frac{s^{-\nu}(s-\cdot)_+^{\nu-\mu-1}}{?\,(\nu-\mu)} \right)(r) \quad = \quad H_\nu(rs).$$

Proof: We directly calculate the assertion and use (12.3.13, *JBI*) from page 190. In detail,

$$\begin{aligned}
& F_\mu \left(\frac{s^{-\nu}(s-\cdot)_+^{\nu-\mu-1}}{?\,(\nu-\mu)} \right)(r) \\
= & \int_0^\infty t^\mu \frac{s^{-\nu}(s-t)_+^{\nu-\mu-1}}{?\,(\nu-\mu)} H_\mu(tr)dt \\
= & \frac{s^{-\nu}}{?\,(\nu-\mu)} \int_0^s t^\mu (s-t)^{\nu-\mu-1} H_\mu(tr)dt \\
= & \frac{s^{-\nu}}{?\,(\nu-\mu)} \int_0^s t^\mu (s-t)^{\nu-\mu-1} J_\mu(2\sqrt{rt})(rt)^{-\mu/2}dt,
\end{aligned}$$

and by substitution $t = su^2$, we get

$$\begin{aligned}
= & \frac{s^{-\nu}}{?\,(\nu-\mu)} \int_0^1 s^\mu u^{2\mu} s^{\nu-\mu-1}(1-u^2)^{\nu-\mu-1} J_\mu(2\sqrt{rs}u)(rsu^2)^{-\mu/2} 2sudu \\
= & \frac{2(rs)^{-\mu/2}}{?\,(\nu-\mu)} \int_0^1 u^{\mu+1}(1-u^2)^{\nu-\mu-1} J_\mu(2\sqrt{rs}u)du \\
= & \frac{2(rs)^{-\mu/2}}{?\,(\nu-\mu)} \frac{2^{\nu-\mu-1}?\,(\nu-\mu)}{(2\sqrt{rs})^{-\nu-\mu}} J_\nu(2\sqrt{rs}) \\
= & (\sqrt{rs})^{-\nu} J_\nu(2\sqrt{rs}) \\
= & H_\nu(rs).
\end{aligned}$$

□

We would like to invert the Fourier transform in the above assertion, but the decay of H_ν is not sufficient to see directly that F_μ is applicable at all. However, we can resort to specific tools from Special Functions to get

Lemma 9.2.16 *(LemFTInvTP) For $\nu > \mu > -1$ and all $r, s > 0$ we have*

$$(F_\mu H_\nu(s\cdot))(r) = \frac{s^{-\nu}(s-r)_+^{\nu-\mu-1}}{?\,(\nu-\mu)} .$$

Proof: The assertion is a consequence of the **Weber–Schafheitlin integral** (see (12.3.21, *EqWeSchaf*) or [1](*abramowitz-stegun:70-1*) p. 487, 11.4.41) after substitutions of the type $t = s^2/2$. In detail, we have

$$\begin{aligned}
& \left(F_\mu H_\nu\left(\frac{u^2}{2}\cdot\right)\right)\left(\frac{r^2}{2}\right) \\
= \quad & \int_0^\infty t^\mu H_\mu\left(\frac{r^2}{2}\,t\right) H_\nu\left(\frac{u^2}{2}\,t\right) dt \\
= \quad & \int_0^\infty \left(\frac{s^2}{2}\right)^\mu \cdot s \cdot H_\mu\left(\frac{r^2}{2}\cdot\frac{s^2}{2}\right) H_\nu\left(\frac{u^2}{2}\cdot\frac{s^2}{2}\right) ds \\
= \quad & \int_0^\infty 2^{-\mu} s^{2\mu+1}\left(\frac{rs}{2}\right)^{-\mu}\left(\frac{us}{2}\right)^{-\nu} J_\nu(us) ds \\
= \quad & 2^\nu r^{-\mu} r^{-\nu} \int_0^\infty s^{\mu-\nu+1} J_\mu(rs) J_\nu(us) ds \\
= \quad & \frac{2^\nu r^{-\mu} u^{-\nu} 2^{\mu-\nu+1} r^\mu (u^2-r^2)_+^{\nu-\mu-1}}{u^\nu ?\,(\nu-\mu)} \\
= \quad & \frac{1}{?\,(\nu-\mu)}\left(\frac{u^2}{2}\right)^{-\nu}\left(\frac{u^2}{2}-\frac{r^2}{2}\right)_+^{\nu-\mu-1}.
\end{aligned}$$

□

We now know that $F_\nu \circ F_\nu = Id$ holds on Laplace transforms, on truncated powers, and on functions of the form $H_\mu(s\cdot)$. But before we generalize this to a larger class of functions, we generalize it to other F_μ operators:

Theorem 9.2.17 *(TheFFI) Let $\nu > \mu > -1$. Then for all tempered radial test functions $f \in \mathcal{S}_{rad}$ we have*

(Eqdd)

$$F_\mu \circ F_\nu = I_{\nu-\mu} \tag{9.2.18}$$

where the integral operator I_α is given by

$$(I_\alpha f)(r) = \int_0^\infty f(s)\frac{(s-r)_+^{\alpha-1}}{?\,(\alpha)}\,ds, \qquad r > 0,\ \alpha > 0.$$

Proof: For any tempered radial test function $f \in \mathcal{S}_{rad}$ we evaluate $(F_\mu \cdot$

$F_\nu)f(r)$ by means of Lemma 9.2.16 (*LemFTInvTP)* to obtain

$$\begin{aligned}
& \int_0^\infty H_\mu(tr)t^\mu \int_0^\infty H_\nu(st)s^\nu f(s)dsdt \\
= & \int_0^\infty s^\nu f(s) \int_0^\infty t^\mu H_\mu(tr)H_\nu(ts)dt\; ds \\
= & \int_0^\infty s^\nu f(s) \cdot F_\mu(H_\nu(s\cdot))(r)ds \\
= & \int_0^\infty f(s)\frac{(s-r)_+^{\nu-\mu-1}}{?\,(\nu-\mu)}\, ds = (I_{\nu-\mu}f)(r).
\end{aligned}$$

□

By the above theorems it is easy to see that

$$I_\alpha H_\nu = H_{\nu-\alpha}$$

for all $\alpha < \nu + 1$, generalizing (12.3.11, *EqHnuDer*).

9.2.7 Wendland's Functions

(*SecWF*) Due to a result of Askey [3](*askey:73-1*) the radial truncated power function

$$A_\mu(\cdot) := (1 - \|\cdot\|_2)_+^\mu$$

is positive definite on $I\!R^d$ for $\mu \geq \lfloor d/2 \rfloor + 1$, because it has a strictly positive radial Fourier transform in this case. Its radial form after substitution is $(1-\sqrt{2r})_+^\mu$, and due to its finite support we can apply any F_ν operator for $\nu > -1$. We use the identity $F_{\nu+\alpha} = F_\nu \circ I_\alpha$ from (9.2.14, *EqFIF*) for this function and get

$$F_{\nu+k}A_\mu = F_\nu(I_k(A_\mu)),\; k \in I\!N,$$

where the left-hand side is strictly positive whenever

(*EqWeCond*)

$$\mu \geq \lfloor d/2 \rfloor + 1 + k. \tag{9.2.19}$$

Thus the function $I_k(A_\mu)$ is positive definite on $I\!R^d$ for the same range of parameters. Since the I_k operators preserve compact supports, the resulting functions

$$\psi_{\mu,k}(r) := I_k(A_\mu(r^2/2))$$

lead to compactly supported positive definite functions

$$\Psi_{\mu,k}(\cdot) = \psi_{\mu,k}(\|\cdot\|_2) = I_k(A_\mu(\|\cdot\|_2^2/2))$$

on $\mathbb{R}^d$ under the condition (9.2.19, *EqWeCond*). Let us do a straightforward evaluation. This yields

(*EqWeGen*)

$$\begin{aligned} I_kA_\mu(r) &= \int_0^\infty (1-\sqrt{2s})_+^\mu \frac{(s-r)_+^{k-1}}{(k-1)!} \\ &= \int_{\sqrt{2r}}^1 t(1-t)^\mu \frac{(t^2/2-r)_+^{k-1}}{(k-1)!} \\ &= \int_x^1 t(1-t)^\mu \frac{(t^2-x^2)_+^{k-1}}{(k-1)!2^{k-1}} \end{aligned} \tag{9.2.20}$$

for $0 \leq r \leq 1/2$ or $0 \leq x = \sqrt{2r} \leq 1$. If μ is an integer, the resulting function is a single polynomial of degree $\mu+2k$ in the variable $x = \|\cdot\|_2$ on its support. The case $k = 1$ is particularly simple. We get the explicit representation

$$\begin{aligned} I_1A_\mu(x^2/2) &= \int_x^1 t(1-t)^\mu dt \\ &= \frac{x(1-x)^{\mu+1}}{\mu+1} + \frac{(1-x)^{\mu+2}}{(\mu+1)(\mu+2)} \\ &= \frac{(1-x)_+^{\mu+1}}{(\mu+1)(\mu+2)}(1+(\mu+1)x)\,. \end{aligned}$$

The smallest possible integer μ for $d \leq 3$ and $k = 1$ is $\mu = 3$, whence

$$I_1A_3(x^2/2) = \frac{1}{20}(1-x)_+^4(1+4x).$$

In addition to $A_{k,\mu} := I_kA_\mu$ let us define

$$B_{k,\mu} := \int_x^1 (1-t)^\mu \frac{(t^2-x^2)_+^{k-1}}{(k-1)!2^{k-1}}$$

and split the integral defining $A_{k,\mu}$ via $t = (t-1)+1$ into

$$A_{k,\mu} = -B_{k,\mu+1} + B_{k,\mu}.$$

Then do integration by parts for $B_{k,\mu}$ and $k > 1$ to get

$$B_{k,\mu} = \frac{1}{\mu+1}A_{k-1,\mu+1}.$$

Thus we have the recurrence relation

$$A_{k,\mu} = -\frac{1}{\mu+2}A_{k-1,\mu+2} + \frac{1}{\mu+1}A_{k-1,\mu+1}.$$

Looking at our result for $k = 1$ we see that we can assume

$$A_{k,\mu}(x^2/2) = (1-x)^{\mu+k}C_{k,\mu}(x)$$

with the recursion

$$C_{k,\mu}(x) = \frac{(x-1)}{\mu+2}C_{k-1,\mu+2}(x) + \frac{1}{\mu+1}C_{k-1,\mu+1}(x),$$

for $k \geq 1$, starting with

$$C_{0,\mu}(x) = 1.$$

Thus the polynomials $C_{k,\mu}$ have degree k with a positive leading coefficient. The number of continuous derivatives of $A_{k,\mu}(x^2/2)$ at $x = 1$ thus is $\mu+k-1 \geq 2k + \lfloor d/2 \rfloor \geq 2k$. To get the number of derivatives at zero we apply the binomial theorem to the last factor in the integrand. Then

$$\begin{aligned}
A_{k,\mu}(x^2/2) &= \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{(-1)^j x^{2j}}{(k-1)!} \int_x^1 t(1-t)^\mu t^{2k-2-2j}\,dt \\
q_{\mu,k-j}(x) &:= \int_x^1 t(1-t)^\mu t^{2k-2-2j}\,dt \\
&= q_{\mu,k-j}(1) - \int_0^x t(1-t)^\mu t^{2k-2-2j}\,dt \\
&= q_{\mu,k-j}(1) - \frac{x^{2k-2j}}{2k-2j} + \text{ higher-order terms}
\end{aligned}$$

shows that the first odd monomial occurring in $A_{k,\mu}(x^2/2)$ cannot have an exponent smaller than $2k+1$. Thus the function has $2k$ continuous derivatives at zero, and we get $2n - 1 = 2k + 1$ in the context of Example 5.2.4 (*AEWF*). In terms of continuity requirements, we get overall C^{2k} continuity at a minimal degree $\mu + 2k = \lfloor d/2 \rfloor + 3k + 1$, and Wendland proves in [46](*wendland:95-1*) that this degree is minimal, if we ask for a single polynomial piece on $[0, 1]$ that induces a positive definite radial function which is C^{2k} and positive definite on $I\!R^d$. Note that the order of smoothness at the boundary of the support is $\lfloor d/2 \rfloor$ larger than the smoothness at zero, which has a positive effect on the visual appearance of the reproduced functions.

We end this by giving the C^4 case for all dimensions d, where $\mu = \lfloor d/2 \rfloor + 3$:

$$\begin{aligned} & A_{2,\mu}(x^2/2) \\ = & \frac{(1-x)_+^{\mu+2}}{(\mu+1)(\mu+2)(\mu+3)(\mu+4)}(x^2(\mu+1)(\mu+3)+3x(\mu+2)+3) \end{aligned}$$

and the most frequent case for $d \leq 3$ is

$$A_{2,4}(x^2/2) = \frac{(1-x)_+^6}{1680}(35x^2+18x+3).$$

The Fourier transforms are

$$F_\nu I_k A_\mu = F_{\nu+k} A_\mu$$

and thus for $r = x^2/2$ of the form

$$\begin{aligned} F_{\nu+k}A_\mu(r) &= \int_0^{1/2}(1-\sqrt{2s})^\mu s^{\nu+k} H_{\nu+k}(rs)ds \\ &= \frac{x^{-\nu-k}}{2^{\nu+k}}\int_0^1 (1-t)^\mu t^{\nu+k+1} J_{\nu+k}(xt)dt \\ &= \frac{x^{-\mu-2\nu-2k-2}}{2^{\nu+k}}\int_0^x (x-u)^\mu u^{\nu+k+1} J_{\nu+k}(u)du. \end{aligned}$$

Due to a result of Gasper [**?**](*gasper:75-1*), the above integral can be written as a positive sum of squares of Bessel functions, at least in the odd-dimensional case $d = 2n-1$ with $\mu = n+k+1$, which leads to $\nu = m - 1/2$ and $\mu = m+1$ for $m = n+k \geq n$. Results of Wendland [46](*wendland:95-1*) then imply the asymptotic behaviour

$$F_\nu I_k A_\mu(r^2/2) = F_{\nu+k}A_\mu(r^2/2) \geq cr^{-d-2k-1}$$

for the necessary values of μ from (9.2.19, *EqWeCond*).

9.2.8 Fourier Transforms of Conditionally Positive Definite Functions

(*SecFTCPD*) We now work towards a proof of conditional positive definiteness of the functions

$$\Phi(x) = \varphi(\|x\|_2) = \|x\|_2^\beta$$

for $\beta \in \mathbb{R}_{>0} \setminus 2\mathbb{N}$. Let us first informally explain how the argument works in general. If we have a radial function

(*EqRadg*)

$$\Phi(x) = \varphi(\|x\|_2) = g(\|x\|_2^2/2) \tag{9.2.21}$$

such that g is recoverable via (9.2.6, *EqFnuGen*) from a radial Fourier transform via

(*EqFTRad1*)

$$g(r) = \int_0^\infty f(t)t^\nu H_\nu(tr)dt, \tag{9.2.22}$$

then the action of functionals on Φ, as in (3.3.2, *DefBil*), is representable by

(*DefBil3*)

$$\begin{array}{rcl} (\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta})_\Phi & = & \displaystyle\sum_{j=1}^{M}\sum_{k=1}^{N} \alpha_j\beta_k g(\|x_j - y_k\|_2^2/2) \\ & = & \displaystyle\sum_{j=1}^{M}\sum_{k=1}^{N} \alpha_j\beta_k \int_0^\infty f(t)t^\nu H_\nu(t\|x_j - y_k\|_2^2/2)dt \\ & = & \displaystyle\int_0^\infty f(t)t^\nu \sum_{j=1}^{M}\sum_{k=1}^{N} \alpha_j\beta_k H_\nu(t\|x_j - y_k\|_2^2/2)dt. \end{array} \tag{9.2.23}$$

This is okay if all functions involved actually behave nicely, but in case of conditionally positive definite functions we have to account for a singularity of f at the origin. This must be cancelled by a zero of the double sum, and we thus assert that the application of functionals from $(I\!P_m^d)^-_{I\!R^d}$ kills off the first m terms of the power series of H_ν. In general:

Lemma 9.2.24 *(LemPDPow) For all polynomials $p \in I\!P_m^d$ and all functionals $\lambda_{X,M,\alpha}$, $\lambda_{Y,N,\beta} \in (I\!P_m^d)^-_{I\!R^d}$ we have*

$$\lambda^u_{X,M,\alpha}\lambda^v_{Y,N,\beta} p(\|u - v\|_2^2) = 0.$$

Proof: We evaluate the left-hand side for a monomial

$$\begin{array}{rcl} \|u - v\|_2^{2n} & = & (\|u\|_2^2 - 2(u,v)_2 + \|v\|_2^2)^n \\ & = & \displaystyle\sum_{\ell \in I\!N^3_{\geq 0},\ |\ell|=n} \frac{n!}{\ell_1!\ell_2!\ell_3!}\|u\|_2^{2\ell_1}(-2(u,v)_2)^{\ell_2}\|v\|_2^{2\ell_3} \end{array}$$

with $0 \leq n < m$. Since both functionals annihilate polynomials of order up to m, there can only be nonzero terms for

$$2\ell_1 + \ell_2 \geq m, \text{ and } \ell_2 + 2\ell_3 \geq m.$$

This implies $n = \ell_1 + \ell_2 + \ell_3 \geq 2m - n$ and $n \geq m$, which is impossible. □

This leads us to look at the Taylor remainder

$$H_{\nu,m}(t) := \sum_{k=m}^{\infty} \frac{(-t)^k}{k!?\,(\nu + k + 1)}$$

of H_ν after chopping the first m terms. If Φ is only a conditionally positive definite function of some order m, we should replace (9.2.22, *EqFTRad1*) by the assumption

(*EqFTRad2*)

$$g(r) = \int_0^\infty f(t) t^\nu H_{\nu,m}(tr) dt, \tag{9.2.25}$$

and let the functionals act like in (9.2.23, *DefBil3*), but on (9.2.25, *EqFTRad2*). Using Lemma 9.2.24 (*LemPDPow*), we can then conclude that the first and third lines of (9.2.23, *DefBil3*) are valid and equal, while the second is invalid. But then we can transform the integral back to non-radial form, letting the sums safely stay inside the integration. This yields precisely (6.1.1, *DefBil2*) for

(*EqwhPhiDef*)

$$\widehat{\Phi}(\cdot) = f(\| \cdot \|_2^2/2) \tag{9.2.26}$$

and is the crucial step to prove the validity of Assumption 6.1.4 (*FTAss1*) for conditionally positive definite functions.

Theorem 9.2.27 *(TheFTCPD) If a function Φ on $I\!R^d$ is radial in the sense of (9.2.21, EqRadg) such that its radial form satisfies (9.2.25, EqFTRad2) for a certain $m \geq 0$, then (6.1.1, DefBil2) holds for the function defined in (9.2.26, EqwhPhiDef) and all functionals from $(I\!P_m^d)^{\perp}_{I\!R^d}$. If f is positive almost everywhere on $(0, \infty)$, then Φ is conditionally positive definite of order m.*

Proof: The first assertion is already proven. The second is a shortcut to Theorem 6.1.8 (*NCCPDFTT*) in the radial case, and it just requires the same arguments as the proofs of Theorems 12.5.6 (*GaussPD*) and 6.1.3 (*NCPDFTT*). □

9.2.9 Application to Transforms of Powers

(*SecFTP*) Let us apply Theorem 9.2.27 (*TheFTCPD)* in case of $\phi(r) = r^\beta$. We can generalize the rule (12.3.11, *EqHnuDer*) for derivatives of H_ν to get (*EqHnuDer2*)

$$-\frac{d}{dt}H_{\nu,m} = H_{\nu+1,m-1} \tag{9.2.28}$$

for the Taylor remainders.

We next assert that the radial Fourier transform of $\phi(r) = r^\beta$ is of the form $r^{-d-\beta}$. This is motivated by the following illegal argument:

The Fourier transform of $\phi(r) = r^\beta$ in $I\!R^d$ will after substitution of $t = r^2/2$ be proportional to

$$\begin{aligned} & \int_0^\infty s^\nu s^{\beta/2} H_\nu(st)ds \\ = \;& t^{-1-\nu-\beta/2}\int_0^\infty u^{\nu+\beta/2}H_\nu(u)du \\ = \;& c(d,\beta)t^{(-d-\beta)/2} \end{aligned}$$

for $\nu = (d-2)/2$. The integral will not exist except for a small and uninteresting range of β and d, but there is a deeper argument by analytic continuation that can be used to turn this Euler-style calculation into a proof.

This makes it reasonable to restrict attention to integrals of monomials against Taylor residuals. Using (9.2.28, *EqHnuDer2*) and integration by parts, they can be reduced to the moment integrals (12.3.20, *EqHnuMom*). In fact, (*EqHnum*)

$$\begin{aligned} \int_0^\infty \frac{t^\rho}{?\,(\rho+1)}H_{\nu,m}(t)dt &= \int_0^\infty \frac{t^{\rho+m}}{?\,(\rho+1+m)}H_{\nu+m,0}(t)dt \\ &= \frac{1}{?\,(\nu-\rho)}. \end{aligned} \tag{9.2.29}$$

for a proper and hopefully wide enough range of ρ, ν, and m. We deliberately omitted the usual factor in the argument of the H_ν function, because we can proceed along the same lines as above to get rid of it by substitution. Let us now check the validity of the above procedure. The final equation will be valid for

(*EqRhoMB*)

$$\rho + m > -1 \text{ and } \nu > 2\rho + m + \frac{1}{2}, \tag{9.2.30}$$

as follows from (12.3.20, *EqHnuMom*). The way back to the first integral requires to check the boundary terms of the form $t^{\rho+j+1}H_{\nu+j,m-j}(t)$ for $0 \leq j < m$ arising from integration by parts. At zero, we have the behaviour $\mathcal{O}(t^{\rho+m+1}) = o(1)$ due to the first restriction in (9.2.30, *EqRhoMB*) and due to the leading powers of the Taylor residual. At infinity, we use the decay properties of $H_{\nu+j}$ as given in section 9.2.2 (*SecCTHTCV)* and get a $\mathcal{O}(t^\alpha)$ behaviour with

$$\alpha = \rho + j + 1 - \nu/2 - j/2 - 1/4 \leq \rho + m - \nu/2 - m/2 + 1/4 < 0$$

due to the second restriction in(9.2.30, *EqRhoMB*).

We now can head for the real thing, i.e.:

(*EqFTPow*)

$$\begin{aligned}
& \int_0^\infty t^{(d-2)/2}t^{-(d+\beta)/2}H_{(d-2)/2,m}(rt)dt \\
= & \int_0^\infty t^{-1-\beta/2}H_{(d-2)/2,m}(rt)dt \\
= & \; r^{\beta/2}\int_0^\infty s^{-1-\beta/2}H_{(d-2)/2,m}(s)ds \\
= & \; r^{\beta/2}\frac{?\,(-\beta/2)}{?\,((d+\beta)/2)}.
\end{aligned} \tag{9.2.31}$$

Here, we applied (9.2.29, *EqHnum*) and have to check (9.2.30, *EqRhoMB*) for the special choice

$$\nu = (d-2)/2, \quad \rho = -1 - \beta/2, \quad m = \lceil \beta/2 \rceil.$$

But this works fine, since $\beta \notin 2I\!N$ and

$$\begin{array}{rcl}
\rho + m & = & -1 - \beta/2 + \lceil \beta/2 \rceil > -1 \\
\nu = \frac{d-2}{2} \geq -\frac{1}{2} & > & 2\rho + m + \frac{1}{2} = -\frac{3}{2} - \beta + \lceil \beta/2 \rceil.
\end{array}$$

The worst cases for the final inequality are very small values of β with $\lceil \beta/2 \rceil = 1$, but they still are safe. We summarize:

Theorem 9.2.32 *(TheCPDPow) The radial basis function* $\Phi(\cdot) = (-1)^{\lceil\beta/2\rceil}\|\cdot\|_2^\beta$ *for* $\beta \in I\!R_{>0} \setminus 2I\!N$ *is conditionally positive definite of order* $m \geq \lceil\beta/2\rceil$*, and it satisfies Assumption 6.1.4 (FTAss1) with* $\beta_0 = \beta$*. The function* $\widehat{\Phi}$ *is*
(EqFTPow2)

$$\widehat{\Phi}(\cdot) = \frac{2^{\beta+d/2}?\,((d+\beta)/2)}{(-1)^{\lceil\beta/2\rceil}?\,(-\beta/2)}\|\cdot\|_2^{-d-\beta}. \tag{9.2.33}$$

□

Note how the sign $(-1)^{\lceil\beta/2\rceil}$ arises when accounting for the sign of $?\,(-\beta/2)$ to make the Fourier transform positive.

9.2.10 Sobolew Splines

(*SecSobSpl*) Before we treat multiquadrics, let us consider the simpler case of Sobolew splines, which are the Fourier transforms of inverse multiquadrics. We want to construct Φ such that the native space is $W_2^k(I\!R^d)$. But since Sobolew space $W_2^k(I\!R^d)$ consists of functions f such that the Fourier transforms satisfy

$$\hat{f}(\omega)(1+\|\omega\|_2)^k \in L_2(I\!R^d),$$

we have to take the Fourier transform of

$$\widehat{\Phi}(\cdot) = (1+\|\cdot\|_2^2)^k.$$

After the usual substitution and the replacement of 1 by an arbitrary positive value y we see that we should make use of (12.3.25, *EqKJ*) to transform it to meet our needs. In particular,

$$\begin{aligned}
&\int_0^\infty \frac{t^{\nu+1}J_\nu(at)}{(t^2+z^2)^{\mu+1}}dt\\
=\ &\int_0^\infty \frac{t^{\nu+1}(at/2)^\nu H_\nu(a^2t^2/4)}{(t^2+z^2)^{\mu+1}}dt\\
=\ &\frac{a^\nu}{2^\nu}\int_0^\infty \frac{t^{2\nu+1}H_\nu(a^2t^2/4)}{(t^2+z^2)^{\mu+1}}dt\\
=\ &r^{\nu/2}2^{\nu/2-\mu}\int_0^\infty \frac{s^\nu H_\nu(rs)}{(s+y)^{\mu+1}}ds\\
=\ &\frac{2^{\nu/2-\mu}r^{\mu/2}y^{(\nu-\mu)/2}}{?\,(\mu+1)}K_{\nu-\mu}(2\sqrt{ry})
\end{aligned}$$

after substitutions $s = t^2/2$, $y = z^2/2$, $r = a^2/2$. In a slightly more convenient form this means

(*EqSobSplRad*)

$$?\,(\mu+1)\int_0^\infty s^\nu(s+y)^{-(\mu+1)}H_\nu(rs)ds \;=\; \left(\frac{y}{r}\right)^{(\nu-\mu)/2}K_{\nu-\mu}(2\sqrt{ry}) \tag{9.2.34}$$

for the above range of μ and ν. If we use our parameters, we have

$$\nu = (d-2)/2,\quad -\mu-1=-k,$$

and then the usual Sobolew inequality $2k > d$ implies that we are safe with the condition $-1 < \nu < 2\mu + 3/2$, as required for (12.3.25, *EqKJ*). We could even work under the weaker condition $d < 4k-3$ without any loss, thus doing a continuous recovery of functions from spaces that contain discontinuous functions.

It remains to perform the substitutions properly: for $\hat{\Phi}(\cdot) = (1+\|\cdot\|_2)^{-k}$ we get $f(t) = 2^{-k}(1/2+t)^{-k}$ and

(*EqSobSpl*)

$$\Phi(\cdot) = 2^{-k}?\,(k)\|\cdot\|_2^{k-d/2}K_{k-d/2}(\|\cdot\|_2) \tag{9.2.35}$$

9.2.11 Multiquadrics

(*SecFTMQ*) Let us now turn the Sobolew spline case upside down. We want to take the d-variate Fourier transform of (9.2.34, *EqSobSplRad*) and come back to $\hat{\Phi}(\cdot) = (1+\|\cdot\|_2^2)^{-k}$. To this end, we can use that the K_ν functions decay expoentially towards infinity due toe (12.3.24, *KnuAsyInf*). At zero, we have to compensate the singularity of K_ν, as given by (12.3.23, *KnuAsyZero*), by introducing the function

(*EqLnuDef*)

$$L_\nu(s^2/4) := K_\nu(s)\left(\frac{s}{2}\right)^\nu. \tag{9.2.36}$$

This definition makes sense due to $K_\nu = K_{-\nu}$ and (12.3.22, *KnuDef*), allowing to write the right-hand side as a function of s^2. The function to be transformed by F_ν then is

$$\left(\frac{y}{s}\right)^{(\nu-\mu)/2}K_{\nu-\mu}(2\sqrt{sy}) = s^{\mu-\nu}L_{\mu-\nu}(sy),$$

and it leads to a function of ω in L_1 after substitution, if $2\mu-2\nu+d>0$ or $\mu > -1$ or $\beta < 0$ for $-\mu-1=\beta/2$. Thus at least for $\nu=(d-2)/2$ and $d < 2(1+\mu) = -\beta$ we can safely invert (9.2.34, *EqSobSplRad*) to get

(*EqMQFTRad*)

$$\int_0^\infty s^\nu \left(\frac{y}{s}\right)^{(\nu-\mu)/2} K_{\nu-\mu}(2\sqrt{sy})H_\nu(rs)ds = (r+y)^{-(\mu+1)}?\,(\mu+1) \quad (9.2.37)$$

without any further calculation. This covers a special range of **inverse multiquadrics**, namely those which can be handled by classical Fourier transforms. Our goal is to proceed towards positive values of $-1-\mu = \beta/2 \notin I\!N$ by a sequence of tricks.

The first step proceeds towards $-d \leq \beta = -2-2\mu < 0$, where both sides of (9.2.37, *EqMQFTRad*) are well-defined, the integrand being in L_1. The only thing that prevents us to use (9.2.37, *EqMQFTRad*) for this range is that we proved it via the Fourier inversion theorem, but the right-hand side fails to have a classical Fourier transform in the new range.

But we can make use of our calculus on the half-line, integrating both sides by application of I_α with respect to the variable y for $\alpha > 0$. This works fine on the right-hand side, but we have to check the action on the left-hand side, rewriting the equation as

(*EqMQFTRad2*)

$$\int_0^\infty s^\mu L_{\nu-\mu}(sy)H_\nu(rs)ds = (r+y)^{-(\mu+1)}?\,(\mu+1). \quad (9.2.38)$$

We use the differentiation rule (12.3.26, *EqKnuDif*) for the K_ν functions to get

$$\begin{aligned} \frac{d}{dz}L_\nu(z^2/4) &= \frac{z}{2}L'_\nu(z^2/4) \\ &= K'_\nu(z)\left(\frac{z}{2}\right)^\nu + \frac{\nu}{2}K_\nu(z)\left(\frac{z}{2}\right)^{\nu-1} \\ &= \left(\frac{z}{2}\right)^{\nu-1} K_{\nu-1}(z)\frac{z}{2} \end{aligned}$$

and thus $L'_\nu = L_{\nu-1}$ for all values of ν. This allows to apply the integral operator I_k for any integer k with $0 \leq k < \mu+1$ to (9.2.38, *EqMQFTRad2*). The result is

(*EqMQFTRad3*)

$$\int_0^\infty s^{\mu-k} L_{\nu-\mu+k}(sy)H_\nu(rs)ds = (r+y)^{-(\mu+1-k)}?\,(\mu-k+1), \quad (9.2.39)$$

and we cannot integrate any further because both sides would cause trouble. However, this settles the case of inverse multiquadrics for all negative exponents $\beta/2$. In fact, starting with some negative $\beta/2$, pick some $k > 0$ such that $2k > d+\beta$ and define $\mu = -1+k-\beta/2$. Then we can use the classical case due to

$$2(\mu+1) = 2k-\beta > d$$

and integrate k times to arrive at (9.2.39, *EqMQFTRad3*) with exponent $-\mu-1+k = \beta/2$ in the right-hand side. The final result is

(*EqMQFTRad3a*)

$$\int_0^\infty s^{-1-\beta/2} L_{(d+\beta)/2}(sy) H_{(d-2)/2}(rs)ds = (r+y)^{\beta/2} ?\,(-\beta/2) \qquad (9.2.40)$$

for all $\beta < 0$.

To proceed towards positive values of β, we have to avoid the singularities of the right-hand side by sticking to non-integer exponents. Furthermore, we have to apply functionals from $(I\!P_m^d)^-_{I\!R^d}$ to both sides in order to avoid a singularity of the integrand at zero and to make sure that the right-hand side vanishes at infinity.

Let us pick functionals $\lambda_{X,M,\alpha}, \lambda_{Y,N,\beta}$ from $(I\!P_m^d)^-_{I\!R^d}$ and apply them with respect to $r = \|x_j - y_k\|_2^2$ to (9.2.39, *EqMQFTRad3*). The result is

(*EqMQFTRad4*)

$$\begin{aligned}
& \int_0^\infty s^{-1-\beta/2} L_{(d+\beta)/2}(sy) \sum_{j=1}^M \sum_{\ell=1}^N \alpha_j \beta_\ell H_{(d-2)/2}(s\|x_j - y_\ell\|_2^2/2)ds \\
= & \int_0^\infty s^{-1-\beta/2} L_{(d+\beta)/2}(sy) \sum_{j=1}^M \sum_{\ell=1}^N \alpha_j \beta_\ell H_{(d-2)/2,m}(s\|x_j - y_\ell\|_2^2/2)ds \\
= & ?\,(-\beta/2) \sum_{j=1}^M \sum_{\ell=1}^N \alpha_j \beta_\ell \left(y + \|x_j - y_\ell\|_2^2/2\right)^{\beta/2}
\end{aligned} \qquad (9.2.41)$$

while still $\beta < 0$. We now can integrate the left-hand side m times with respect to y without running into difficulties whenever $\beta \notin 2\mathbb{Z}$. This yields

(*EqMQFTRad4a*)

$$\begin{aligned}
& \int_0^\infty s^{-1-\beta/2-m} L_{(d+\beta)/2+m}(sy) \sum_{j=1}^M \sum_{\ell=1}^N \alpha_j \beta_\ell H_{(d-2)/2,m}(s\|x_j - y_\ell\|_2^2/2)ds \\
= & ?\,(-\beta/2-m)(-1)^m \sum_{j=1}^M \sum_{\ell=1}^N \alpha_j \beta_\ell \left(y + \|x_j - y_\ell\|_2^2/2\right)^{\beta/2+m}.
\end{aligned} \qquad (9.2.42)$$

Looking at the Taylor expansion of the right-hand side assures us that it still vanishes at infinity, as required. We thus define $\gamma := \beta + 2m < 2m$ and rewrite the equation as

(*EqMQFTRad4a*)

$$\begin{aligned}
&\int_0^\infty s^{-1-\gamma/2} L_{(d+\gamma)/2}(sy) \sum_{j=1}^M \sum_{\ell=1}^N \alpha_j \beta_\ell H_{(d-2)/2,m}(s\|x_j - y_\ell\|_2^2/2) ds \\
= \; & ?\,(-\gamma/2)(-1)^m \sum_{j=1}^M \sum_{\ell=1}^N \alpha_j \beta_\ell \left(y + \|x_j - y_\ell\|_2^2/2\right)^{\gamma/2}.
\end{aligned} \tag{9.2.43}$$

This proves that multiquadrics $(c^2 + \|x\|_2^2)_{\gamma/2}$ for $\gamma \in I\!R_{>0} \setminus 2I\!N$ are conditionally positive definite of order $m = \lceil \gamma/2 \rceil$ with Fourier transform

$$\frac{2^{\gamma/2}(-1)^m}{?\,(-\gamma/2)} \left(\frac{\|omega\|_2}{c}\right)^{-(\gamma+d)/2} K_{(\gamma+d)/2}(c\|\omega\|_2)$$

after backsubstitution.

9.2.12 Nonexistence of CS Functions for All Dimensions

(*NECSAlld*)

9.3 Positive Definite Functions on Topological Groups

9.4 Positive Definite Zonal Functions on Spheres

10 Special Algorithms

(*SecSA*) This section contains some additional techniques that may be useful for the numerical solution of multivariate recovery problems.

10.1 Reduction of Enlarged System, Method 1

(*Red1*) Consider the enlarged system (1.7.3, *BDef*) and perform a partial Gaussian elimination algorithm on the matrix P with row interchanges. The result can be written in the form

$$LP\Pi = \begin{pmatrix} U \\ 0 \end{pmatrix}, \quad L = \begin{pmatrix} L_{11} & 0 \\ L_{21} & E \end{pmatrix}$$

with nonsingular lower triangular matrices L and L_{11} of size $M \times M$ and $q \times q$, respectively, with an $M \times M$ permutation matrix Π and a nonsingular $q \times q$ upper triangular matrix U, while L_{21} is some $M \times q$ matrix and E is the identity matrix. Now write α as a vector

(*split1*)

$$\alpha = L^T\beta = \begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & E \end{pmatrix} \begin{pmatrix} \beta^{(1)} \\ \beta^{(2)} \end{pmatrix} = \begin{pmatrix} L_{11}^T\beta^{(1)} + L_{21}^T\beta^{(2)} \\ \beta^{(2)} \end{pmatrix}, \tag{10.1.1}$$

again using a split of an M-vector into a q-vector followed by an $(M-q)$-vector. Ignoring the details of such obvious splits from now on, we evaluate

$$0 = \Pi^T P^T \alpha = \Pi^T P^T L^T \beta = U^T \beta^{(1)} + 0$$

and get $\beta^{(1)} = 0$. Now we split the system $LAL^T\beta + LP\gamma = Lf$ in the same way to get

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} 0 \\ \beta^{(2)} \end{pmatrix} + \begin{pmatrix} U \\ 0 \end{pmatrix} \delta = \begin{pmatrix} g^{(1)} \\ g^{(2)} \end{pmatrix}$$

introducing the vector δ via $\gamma = \Pi\delta$. This decomposes into two systems

$$A_{22}\beta^{(2)} = g^{(2)}, \quad A_{12}\beta^{(2)} + U\delta = g^{(1)}$$

that can be solved for $\beta^{(2)}$ and δ, respectively. From these it is easy to calculate α and γ.

To see the positive definiteness of the matrix A_{22}, observe that

$$(\beta^{(2)})^T A_{22} \beta^{(2)} = \begin{pmatrix} 0 \\ \beta^{(2)} \end{pmatrix}^T LAL^T \begin{pmatrix} 0 \\ \beta^{(2)} \end{pmatrix}$$

holds for all $\beta^{(2)} \in I\!R^{M-q}$, and all α with (1.6.3, *CPDef*) have a unique split in the form (10.1.1, *split1*). Thus A_{22} defines a positive definite quadratic form on R^{M-q}, and it must be a positive definite matrix.

To calculate the numerical effort, we now explicitly write down the algorithm:

1. Perform q Gaussian transformations on rows of P with pivoting. This requires $\mathcal{O}(Mq^2)$ operations and generates the matrices Π, U, L_{11}, and L_{21}. The latter three can be stored over P, and Π requires an integer array of length q for keeping track of row permutations.

2. Calculate the submatrices A_{ik} of LAL^T by applying the Gaussian transformations stored in L to A row- and columnwise. These are q transformations of M-vectors each, and the overall effort will be $\mathcal{O}(M^2q)$. Note that this operation will cause fill-in, if the original matrix was sparse.

3. Calculate Lf and split it into $g^{(1)}$ and $g^{(2)}$. Using the special form of L again, this amounts to $\mathcal{O}(Mq)$ operations.

4. Solve the positive definite $(M-q)\times(M-q)$ system $A_{22}\beta^{(2)} = g^{(2)}$ by your favourite method. We shall comment on such problems for unconditionally positive definite functions later in section 2.3 (*CompEffort*). Its computational complexity does not enter into the complexity of the transformation we consider here.

5. Now solve $A_{12}\beta^{(2)} + U\delta = g^{(1)}$ for δ. Using the upper triangular structure of U, the computational effort is $\mathcal{O}(Mq+q^2)$ for forming the system and solving it.

6. Backpermutation of elements of δ yields γ at $\mathcal{O}(q)$ cost.

7. Finally, α is an extension of $\beta^{(2)}$ by the q components of the vector $L_{21}^T\beta^{(2)}$, and these can be calculated by $\mathcal{O}(Mq^2)$ operations.

Since we started with a conditionally positive definite function of *positive* order m, the increase of Φ towards infinity leads to a matrix A that shows a more or less strong increase of elements when moving away from the main diagonal. After the reduction the resulting matrix behaves like one generated by a positive definite function (this is actually provable for the reduction method of the next section). Thus it usually shows off-diagonal decay, and numerical experiments indicate some improvement of the condition. Thus there is some hope that variations of these reduction methods can possibly be turned into efficient preconditioning techniques.

10.2 Reduction of Enlarged System, Method 2

(*Red2*) Again, we consider the enlarged system (1.7.3, *BDef*), but now we perform q Householder transformations on P^T with column pivoting. This means a reordering of the points in $X=\{x_1,\ldots,x_M\}$ and transition to a new basis in $\mathbb{P}_m^d$. In linear algebra terms we end up with a decomposition

(*Dec2*)

$$U^{-1}QP^T\Pi = (E,S) \tag{10.2.1}$$

with a nonsingular upper triangular $q\times q$ matrix U, an orthogonal $q\times q$ matrix Q, an $M\times M$ permutation matrix Π and a plain $q\times(M-q)$ matrix S. Note that the Householder transformations first produce $QP^T\Pi=(U,*)$, but we left-multiply this with U^{-1} to get (10.2.1, *Dec2*).

Now we permute and split α by

(*split3*)

$$\alpha = \Pi\beta, \quad \beta = \begin{pmatrix} \beta^{(1)} \\ \beta^{(2)} \end{pmatrix} \tag{10.2.2}$$

into a q-vector followed by an $(M-q)$-vector. Then we evaluate

$$0 = U^{-1}QP^T\alpha = U^{-1}QP^T\Pi\Pi^{-1}\alpha = (E,S)\begin{pmatrix} \beta^{(1)} \\ \beta^{(2)} \end{pmatrix}$$

and get $\beta^{(1)} = -S\beta^{(2)}$. Now we split the system $\Pi^T A\Pi\beta + \Pi^T P\gamma = \Pi^T f$ to get

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} \beta^{(1)} \\ \beta^{(2)} \end{pmatrix} + \begin{pmatrix} E \\ S^T \end{pmatrix}\delta = \begin{pmatrix} g^{(1)} \\ g^{(2)} \end{pmatrix}$$

introducing the vector $\delta = U^TQ\gamma$. This decomposes into two systems

$$\begin{array}{ccccccc} A_{11}\beta^{(1)} & + & A_{12}\beta^{(2)} & + & \delta & = & g^{(1)} \\ A_{21}\beta^{(1)} & + & A_{22}\beta^{(2)} & + & S^T\delta & = & g^{(2)} \end{array}$$

and we solve the first for δ by

(*DeltaSys*)

$$\delta = g^{(1)} - A_{11}\beta^{(1)} - A_{12}\beta^{(2)} = g^{(1)} + (A_{11}S - A_{12})\beta^{(2)}. \tag{10.2.3}$$

Putting this with $\beta^{(1)} = -S\beta^{(2)}$ into the second yields a symmetric $(M-q)\times(M-q)$ system

(*RedSys3*)

$$\left(A_{22} + S^TA_{11}S - S^TA_{12} - A_{21}S\right)\beta^{(2)} = g^{(2)} - S^Tg^{(1)} \tag{10.2.4}$$

that can be solved for $\beta^{(2)}$. To see the positive definiteness of the matrix (10.2.4, *RedSys3*), observe that

$$(\beta^{(2)})^T\left(A_{22} + S^TA_{11}S - S^TA_{12} - A_{21}S\right)\beta^{(2)} = \beta^T\Pi^TA\Pi\beta = \alpha^TA\alpha$$

holds for all $\beta^{(2)} \in I\!R^{M-q}$, and all α with (1.6.3, *CPDef*) have a unique split in the form (10.2.2, *split3*) with $\beta^{(1)} = -S\beta^{(2)}$. In 3.3.2 (*PhiNormalization*) we shall see that this matrix can be written in the form $A_{Y,\Psi}$ for a set Y of $M-q$ points and a function Ψ that is *unconditionally* positive definite on $I\!R^d \setminus (X \setminus Y)$..

Let us now explicitly write down the algorithm:

1. Perform q Householder transformations on P^T with pivoting by column permutation of P^T. This requires $\mathcal{O}(Mq^2)$ operations and generates the matrices Π, U, and US. The latter two can be stored over P^T, and Π requires an integer array of length M for keeping track of point permutations.

2. Solve for S by backward substitution, using U. This again requires $\mathcal{O}(Mq^2)$ operations and generates S, which can be stored over part of P.

3. Generate the submatrices A_{ik} of $\Pi^T A \Pi$ by applying the permutations defined by Π to A row- and columnwise. This requires $2M$ swaps of M-vectors, and the overall effort will be $\mathcal{O}(M^2)$. Note that this operation can be avoided by using indirect indexing in later steps, but be aware of the fact that indirect indexing spoils the positive effect of cache memory.

4. Permute the right-hand side of the system and split it into $g^{(1)}$ and $g^{(2)}$. This amounts to $\mathcal{O}(M)$ operations, but is unnecessary if indirect indexing is implemented.

5. The bulk of work in this reduction method lies in forming the positive definite matrix
$$A_{22} + S^T A_{11} S - S^T A_{12} - A_{21} S,$$
and it is of order $\mathcal{O}(M^2 q)$.

6. Now solve the positive definite $(M - q) \times (M - q)$ system (10.2.4, *RedSys3*) for $\beta^{(2)}$ by your favourite method. We considered such problems for unconditionally positive definite functions in section 2.3 (*CompEffort)*. Its computational complexity does not enter into the complexity of the transformation we describe here.

7. Now form $\beta^{(1)} = -S\beta^{(2)}$ with $\mathcal{O}(M^2 q)$ operations and

8. use (10.2.3, *DeltaSys*) to calculate δ with $\mathcal{O}(Mq^2)$ operations. The solution vector α just is a permuted version of β, but the calculation of γ requires solution of the system $U^T Q\gamma = \delta$ in two steps:

9. Calculate $Q\gamma$ from δ by backward substitution with $\mathcal{O}(q^2)$ operations, and

10. form $\gamma = Q^T(Q\gamma)$ by premultiplication of $Q\gamma$ with Q^T with $\mathcal{O}(q^3)$ operations. Since $M \geq q$ follows from $I\!\!P_m^d$-nongegeneracy of $X = \{x_1, \ldots, x_M\}$, this is at most an $\mathcal{O}(Mq^2)$ effort.

11 Computational Geometry Techniques

(*SecCGT*) This section contains algorithms from Computational Geometry that are useful for solving scattered data problems in the the large. The main topic will be the **k-nearest neighbor problem** and related **query problems**.

11.1 Voronoi Diagrams

(*SecVor*)

12 Analytic Background

(*SecAB*) This section collects the required facts from Functional and Real Analysis that the core of this text requires as basic knowledge. It is useful for teaching purposes, because it makes the text self-contained. Researchers and advanced students will not need to look into this, but beginners should brush up their background by checking it against the contents of this section. And in case of doubt or lack of memory, any reader should get an easy possibility to access the backing material without consulting too many different texts.

12.1 Calculus Facts

(*SecCalcFacts*) We start with some basics from calculus that we need for notational reference.

12.1.1 Taylor's Formula and Truncated Powers

(*Sec*) Using the **truncated power function**

(*EqTrPoFu*)

$$(x)_+^k := \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{12.1.1}$$

we write **Taylor's Formula** as

(*EqTayFor*)

$$\begin{aligned} f(x) &= \sum_{j=0}^{\ell-1} \frac{f^{(j)}(a)}{j!}(x-a)^j + \int_a^b \frac{(x-t)_+^{\ell-1}}{(\ell-1)!} f^{(\ell)}(x)dx \\ f(x) &=: p_{f,\ell}(x) + r_{f,\ell}(x) \end{aligned} \tag{12.1.2}$$

for all functions f with absolutely integrable ℓ-th derivative on $[a,b] \subset I\!R$ and all arguments $x \in [a,b]$. The polynomial part $p_{f,\ell}$ is in $I\!P^1_\ell$, while the residual $r_{f,\ell}$ has the crude bounds

$$\begin{aligned} |r_{f,\ell}(x)| &\leq \frac{(b-a)^\ell}{\ell!}\|f^{(\ell)}\|_{L_\infty[a,b]} \\ |r_{f,\ell}(x)| &\leq \frac{(b-a)^{\ell-1/2}}{\sqrt{(2\ell-1)!}}\|f^{(\ell)}\|_{L_2[a,b]} \end{aligned}$$

depending on how $f(\ell)$ is extracted from the integral. The second case uses the Cauchy-Schwarz inequality, and it is a first case where half-integers enter naturally into an approximation order. Here are some special instances that are needed in the text:

Example 12.1.3 *(ExaExpNeg) For $f(x) = e^{-x}$ on $[0,h]$ we have*

$$|f(x) - p_{f,\ell}(x)| = |r_{f,\ell}(x)| \leq \frac{h^\ell}{\ell!}.$$

Example 12.1.4 *(ExaExpIma) For the real and imaginary parts of $f(x) = e^{ix}$, i.e. the sine and cosine function, we know that all derivatives are bounded by one. Thus we get*

$$|f(x) - p_{f,\ell}(x)| = |r_{f,\ell}(x)| \leq \frac{h^\ell}{\ell!}.$$

on $[-h,h]$ for $h > 0$. We avoid a factor of 2 in the bound of the residual by using Taylor's formula on both $[0,h]$ and $[-h,0]$.

12.2 Hilbert Space Basics

(*SecHSB*) This is intended as a short tutorial on Hilbert spaces as required in Section 3 (*SecHST)*. We only require fundamentals on linear spaces, bilinear forms, and norms. If a reader has problems with any of the stated facts below, it is time to go back to an introductory course on Calculus and Numerical Analysis.

Definition 12.2.1 *(DefPHS) A set $\mathcal{H}$ and a mapping $(\cdot,\cdot)_\mathcal{H} : \mathcal{H}\times\mathcal{H} \to I\!R$ form a* **pre-Hilbert space** *over $I\!R$, if the following holds:*

1. *$\mathcal{H}$ is a vector space over $I\!R$.*
2. *$(\cdot,\cdot)_\mathcal{H}$ is a symmetric positive definite bilinear form.*

A symmetric positive bilinear form as $(\cdot,\cdot)_{\mathcal{H}} : \mathcal{H}\times\mathcal{H}\to I\!R$ is often called an **inner product** on $\mathcal{H}$. Then

(*NormDef*)

$$\|x\|^2_{\mathcal{H}} := (x,x)_{\mathcal{H}},\ x\in\mathcal{H} \tag{12.2.2}$$

defines a norm on $\mathcal{H}$, and we assume all readers to be familiar with this notion. Sometimes, we shall weaken the assumptions on $(\cdot,\cdot)_{\mathcal{H}}$ and only ask for symmetry and positive semidefiniteness. Even in this more general situation, we have the **Cauchy-Schwarz inequality**

$$|(u,v)_{\mathcal{H}}| \le |u|_{\mathcal{H}}|v|_{\mathcal{H}}$$

for all $u,v\in\mathcal{H}$, where we use the notation $|x|^2_{\mathcal{H}} := (x,x)_{\mathcal{H}}$ to denote a **seminorm** instead of a norm as in (12.2.2, *NormDef*). To prove the Cauchy-Schwarz inequality as a warm-up, just consider the quadratic function

$$\varphi(t) := |u+tv|^2_{\mathcal{H}} = |u|^2_{\mathcal{H}} + 2t(u,v)_{\mathcal{H}} + t^2|v|^2_{\mathcal{H}}.$$

It must be nonnegative, and thus it has none or a double real zero. This property is satisfied for a general function $\varphi(t) = at^2+2bt+c$, iff $b^2\le ac$ holds. But this is the square of the Cauchy-Schwarz inequality.

For completeness, we recall some basics from normed linear spaces:

1. A sequence $\{u_n\}_{n\in I\!N}\subset\mathcal{N}$ of a normed linear space $\mathcal{N}$ with norm $\|\cdot\|_{\mathcal{N}}$ is a **zero sequence** in $\mathcal{N}$, if the sequence $\{\|u_n\|_{\mathcal{N}}\}_{n\in I\!N}$ converges to zero in $I\!R$.

2. A sequence $\{u_n\}_{n\in I\!N}\subset\mathcal{N}$ is a **convergent sequence** in $\mathcal{N}$ with limit u, if the sequence $\{u_n-u\}_n$ is a zero sequence.

3. A subspace $\mathcal{M}$ of $\mathcal{N}$ is a **closed subspace**, if for every convergent sequence $\{u_n\}_{n\in I\!N}\subset\mathcal{M}\subset\mathcal{N}$ with limit u one can conclude that the limit u also belongs to $\mathcal{M}$.

4. The normed linear space $\mathcal{N}$ is **complete** or a **Banach space**, if every sequence which is a Cauchy sequence in the norm $\|\cdot\|_{\mathcal{V}}$ is necessarily convergent in $\mathcal{V}$.

5. A complete normed linear space is closed, since each convergent sequence is a Cauchy sequence.

6. A subset $\mathcal{M}$ of a normed linear space $\mathcal{N}$ is **dense**, if each element of $\mathcal{N}$ can be written as a limit of a convergent sequence from $\mathcal{M}$.

Now we add some simple facts about pre-Hilbert spaces:

1. A mapping (or operator) $A : \mathcal{H} \to \mathcal{N}$ with values in a normed linear space $\mathcal{N}$ with norm $\|\cdot\|_{\mathcal{N}}$ is a **continuous mapping** or a **bounded mapping**, if there is a constant C such that
$$\|Ax\|_{\mathcal{N}} \le C\|x\|_{\mathcal{H}}$$
holds for all $x \in \mathcal{H}$.

2. The mapping A then has an **operator norm**
$$\|A\|_{\mathcal{H},\mathcal{N}} := \sup_{x\in\mathcal{H}\setminus\{0\}} \frac{\|Ax\|_{\mathcal{N}}}{\|x\|_{\mathcal{H}}} \le C$$
and the bound
$$\|Ax\|_{\mathcal{N}} \le \|A\|_{\mathcal{H},\mathcal{N}}\|x\|_{\mathcal{H}}$$
is best possible.

3. Two subspaces $\mathcal{U}$, $\mathcal{V}$ of a pre-Hilbert space are **orthogonal**, if all vectors $u \in \mathcal{U}$, $v \in \mathcal{V}$ are orthogonal, i.e.: $(u, v)_{\mathcal{H}} = 0$.

Definition 12.2.3 *An element u^* of a subspace $\mathcal{M}$ of a normed linear space $\mathcal{N}$ is a* **best approximation** *to a given element $u \in \mathcal{N}$, if*
$$\|u - u^*\|_{\mathcal{N}} = \sup_{v\in\mathcal{M}} \|u - v\|_{\mathcal{N}} =: E_{\mathcal{M}}(u).$$

Theorem 12.2.4 *(BAT) An element u^* of a subspace $\mathcal{M}$ of a pre-Hilbert space $\mathcal{H}$ is a best approximation to a given element $u \in \mathcal{H}$, iff the variational identity*

(EqVar)
$$(u - u^*, v)_{\mathcal{H}} = 0 \textit{ for all } v \in \mathcal{M} \tag{12.2.5}$$
holds. If it exists, the best approximation is unique. If $\mathcal{M}$ is finite-dimensional and spanned by linearly independent elements $u_1 \dots, u_M$, then the coefficients α^* of the representation
$$u^* = \sum_{j=1}^{M} \alpha_j^* u_j$$
are solutions of the **normal equations**
$$\sum_{j=1}^{M} \alpha_j^*(u_j, u_k)_{\mathcal{H}} = (u, u_k)_{\mathcal{H}},\ 1 \le k \le M,$$
and the symmetric and positive definite matrix with entries $(u_j, u_k)_{\mathcal{H}}$ is called a **Gram matrix**.

Proof: Let u^* be a best approximation to u. Then consider an arbitrary $v \in \mathcal{M}$ and form the quadratic function

$$u_v(\alpha) := \|u - u^* + \alpha v\|_{\mathcal{H}}^2 = \|u - u^*\|_{\mathcal{H}}^2 + 2\alpha(u - u^*, v)_{\mathcal{H}} + \alpha^2\|v\|_{\mathcal{H}}^2$$

whose minimum must be attained at $\alpha = 0$. This implies $(u - u^*, v)_{\mathcal{H}} = 0$. Conversely, assume (12.2.5, *EqVar*) and write any other element $v \in \mathcal{M}$ as $v = u^* + 1 \cdot (v - u^*)$. Then (12.2.5, *EqVar*) implies that the quadratic function u_{u^*-v} is minimal at $\alpha = 0$, proving $u_{u^*-v}(1) = \|u - v\|_{\mathcal{H}} \geq u_{u^*-v}(0) = \|u - u^*\|_{\mathcal{H}}$. If u^* and u^{**} are two best approximations from $\mathcal{M}$ to u, then we can subtract the two variational identities $(u - u^*, v)_{\mathcal{H}} - (u - u^{**}, v)_{\mathcal{H}} = (u^{**} - u^*, v)_{\mathcal{H}} = 0$ for all $v \in \mathcal{M}$ and insert $v = u^{**} - u^*$ to get $u^{**} = u^*$. The third assertion is a specialization of (12.2.5, *EqVar*). □

Corollary 12.2.6 *The first statement of Theorem 12.2.4 (BAT) holds also in the case of a positive semidefinite bilinear form. The Gram matrix in the finite-dimensional case now is only positive semidefinite.* □

Corollary 12.2.7 *(BAC) Let $\lambda_1, \ldots, \lambda_M$ be linear functionals on a pre-Hilbert space $\mathcal{H}$ and let some $u \in \mathcal{H}$ be given. An element u^* of $\mathcal{H}$ solves the problem*

$$\|u^*\|_{\mathcal{H}} = \inf_{\substack{v \in \mathcal{H} \\ \lambda_j(v) = \lambda_j(u) \\ 1 \leq j \leq M}} \|v\|_{\mathcal{H}},$$

iff the variational identity

$$(u^*, v)_{\mathcal{H}} = 0 \text{ for all } v \in \mathcal{H} \text{ with } \lambda_j(v) = 0,\ 1 \leq j \leq M.$$

holds, or iff there are real numbers $\alpha_1, \ldots, \alpha_M$ such that

$$(u^*, v)_{\mathcal{H}} = \sum_{j=1}^{M} \alpha_j \lambda_j(v) \text{ for all } v \in \mathcal{H}.$$

Proof: Consider the subspace

$$\mathcal{M} = \{\, v \in \mathcal{H} \;:\; \lambda_j(v) = 0,\ 1 \leq j \leq M\}$$

and reformulate the problem by writing any $v \in \mathcal{H}$ with $\lambda_j(v) = \lambda_j(u)$, $1 \leq j \leq M$ as $v = u - w$ for $w \in \mathcal{M}$. Then we have a problem of best approximation to u from $\mathcal{M}$ and can simply use Theorem 12.2.4 (*BAT*) to prove the first assertion. We then have to prove that the first variational identity implies the second. But this follows from a standard linear algebra argument that we include for completeness as the next lemma. □

Lemma 12.2.8 *If $A : X \to Y$ and $B : X \to Z$ are linear maps between linear spaces, and if B vanishes on the kernel $\ker A$ of A, then B factorizes over $A(X)$, i.e.: there is a map $C : A(X) \to Z$ such that $B = C \circ A$. If Z is normed and if Y is finite-dimensional, then C is continuous.*

Proof: There is an isomorphism $D : A(X) \to X/\ker A$, and one can define $\widetilde{B} : A/\ker A \to Z$ by $\widetilde{B}(x + \ker A) := B(x)$ because $B(\ker A) = \{0\}$. Then $C := \widetilde{B} \circ D$ does the job, since

$$C(A(x)) = \widetilde{B}(D(A(x))) = \widetilde{B}(x + \ker A) = B(x)$$

for all $x \in X$. If Y is finite-dimensional, the isomorphic spaces $A(X) \subseteq Y$ and $X/\ker A$ must also be finite-dimensional. Since all linear mappings defined on finite-dimensional linear spaces with values in normed linear spaces are continuous, we are finished. □

So far, Theorem 12.2.4 *(BAT)* does not imply existence of best approximations from subspaces of infinite dimension. It just characterizes them. To get existence, we need that certain nice sequences actually have limits:

Definition 12.2.9 *(DefHS) A pre-Hilbert space $\mathcal{H}$ with inner product $(\cdot,\cdot)_{\mathcal{H}}$ is a* **Hilbert space** *over $I\!R$, if $\mathcal{H}$ is* **complete** *under the norm $\|\cdot\|_{\mathcal{H}}$, i.e.: as a normed linear space.*

We now prove the crucial **projection theorem** in Hilbert spaces:

Theorem 12.2.10 *(PTHS) If $\mathcal{H}$ is a Hilbert space with a closed subspace $\mathcal{M}$, then any element $u \in \mathcal{H}$ has a unique best approximation $u^*_{\mathcal{M}}$ from $\mathcal{M}$, and the elements $u^*_{\mathcal{M}}$ and $u - u^*_{\mathcal{M}}$ are orthogonal. The map $\Pi_{\mathcal{M}} : \mathcal{H} \to \mathcal{M}$ with $\Pi_{\mathcal{M}}(u) := u^*_{\mathcal{M}}$ is linear, has norm one if $\mathcal{M}$ is nonzero, and is a* **projector***, i.e.: $\Pi^2_{\mathcal{M}} = \Pi_{\mathcal{M}}$. If Id is the identity mapping, then $Id - \Pi_{\mathcal{M}}$ is another projector, mapping $\mathcal{H}$ onto the* **orthogonal complement**

$$\mathcal{M}^- := \{ u \in \mathcal{H} \ : \ (u,v)_{\mathcal{H}} = 0 \textit{ for all } v \in \mathcal{M} \}.$$

of $\mathcal{M}$. Finally, the decomposition

$$\mathcal{H} = \mathcal{M} + \mathcal{M}^-$$

is a direct and orthogonal sum of spaces.

Proof: The existence proof for approximations from finite-dimensional subspaces is a consequence of Theorem 12.2.4 *(BAT)*, and we postpone the general case for a moment. The orthogonality statement follows in general from Theorem 12.2.4 (*BAT*), and it yields Pythagoras' theorem in the form

$$\|u\|_{\mathcal{H}}^2 = \|u - u^*_{\mathcal{M}}\|_{\mathcal{H}}^2 + \|u^*\|_{\mathcal{H}}^2.$$

This in turn proves that both projectors have a norm not exceeding one. It is easy to prove that $\alpha u^*_{\mathcal{M}} + \beta v^*_{\mathcal{M}}$ is a best approximation to $\alpha u + \beta v$ for all $\alpha, \beta \in I\!R$ and all $u, v \in \mathcal{H}$, using the variational identity in Theorem 12.2.4 (*BAT*). To prove linearity of the projectors, we use uniqueness of the best approximation, as follows from Theorem 12.2.4 (*BAT*). Finally, surjectivity of the projectors is easily proven from the best approximation property of their definition.

Thus we are left with the existence proof for the infinite-dimensional case. The nonnegative real number $E_{\mathcal{M}}(u)$ can be written as the limit of a decreasing sequence $\{\|u - v_n\|_{\mathcal{H}}\}_n$ for certain elements $v_n \in \mathcal{M}$, because it is defined as an infimum. Forming the subspaces

$$\mathcal{M}_n := \text{span}\,\{v_1, \ldots, v_n\} \subseteq \mathcal{M}$$

and unique best approximations w_n to u from $\mathcal{M}_n$, we get

$$E_{\mathcal{M}}(u) \leq \|u - w_n\|_{\mathcal{H}} \leq \|u - v_n\|_{\mathcal{H}},$$

such that the sequence $\{\|u - w_n\|_{\mathcal{H}}\}_n$ converges to $E_{\mathcal{M}}(u)$, too. We now fix indices $m \geq n$ and use that $(u - w_m, w_m - w_n)_{\mathcal{H}} = 0$ follows from the best approximation property of w_m. Then we have

$$\begin{aligned}
\|u - w_n\|_{\mathcal{H}}^2 - \|u - w_m\|_{\mathcal{H}}^2 &= \|u - w_m + w_m - w_n\|_{\mathcal{H}}^2 - \|u - w_m\|_{\mathcal{H}}^2 \\
&= \|u - w_m\|_{\mathcal{H}}^2 + 2(u - w_m, w_m - w_n)_{\mathcal{H}} \\
&\quad + \|w_m - w_n\|_{\mathcal{H}}^2 - \|u - w_m\|_{\mathcal{H}}^2 \\
&= \|w_m - w_n\|_{\mathcal{H}}^2,
\end{aligned}$$

and since the sequence $\{\|u - w_n\|_{\mathcal{H}}^2\}_n$ is convergent and thus a Cauchy sequence, we get that $\{w_n\}_n \subset \mathcal{M}$ is a Cauchy sequence in $\mathcal{M} \subseteq \mathcal{H}$. Now the completeness of $\mathcal{H}$ assures the existence of a limit $w^* \in \mathcal{H}$ of this sequence, and since $\mathcal{M}$ was ssumed to be closed, the element w^* must belong to $\mathcal{M}$. The above identity can be used to let m tend to infinity, and then we get

$$\|u - w_n\|_{\mathcal{H}}^2 - \|u - w^*\|_{\mathcal{H}}^2 = \|w^* - w_n\|_{\mathcal{H}}^2.$$

This proves

$$E_{\mathcal{M}}(u) \leq \|u - w^*\|_{\mathcal{H}} \leq \|u - w_n\|_{\mathcal{H}},$$

and since the right-hand side converges to $E_{\mathcal{M}}(u)$, the element w^* must be the best approximation to u. □

We now proceed towards the completion theorem for pre-Hilbert spaces.

Theorem 12.2.11 *(HSCT) Let $\mathcal{H}$ be a pre-Hilbert space with inner product $(\cdot,\cdot)_{\mathcal{H}}$. Then there is a Hilbert space $\mathcal{J}$ and an isometric embedding $J : \mathcal{H} \to \mathcal{J}$ such that the following is true:*

1. *$J(\mathcal{H})$ is dense in $\mathcal{J}$.*
2. *Any continuous mapping $A : \mathcal{H} \to \mathcal{N}$ with values in a Banach space $\mathcal{N}$ has a unique continuous extension $B : \mathcal{J} \to \mathcal{N}$ such that $B \circ J = A$.*

Proof: We first form the space of all Cauchy sequences in $\mathcal{H}$, which clearly is a linear space over $I\!R$. Two such sequences are called equivalent, if their difference is a sequence in $\mathcal{H}$ converging to zero. The space $\mathcal{J}$ now is defined as the space of equivalence classes of Cauchy sequences in $\mathcal{H}$ modulo zero sequences. These classes clearly form a vector space under the usual operations on sequences. If we use an overstrike to stand for "class of", we write an element of $\mathcal{J}$ as $\overline{\{u_n\}_n}$ for a Cauchy sequence $\{u_n\}_n \in \mathcal{H}$. Now it is time to define an inner product

$$(\overline{\{u_n\}_n}, \overline{\{v_n\}_n})_{\mathcal{J}} := \lim_{n\to\infty} (u_n, v_n)_{\mathcal{H}}$$

on $\mathcal{J}$ and the embedding J via the constant Cauchy sequences

$$Ju := \{u\}_n := \{u_n = u\}_n$$

for each $u \subset \mathcal{H}$. Then

$$(Ju, Jv)_{\mathcal{J}} = (u, v)_{\mathcal{H}}$$

makes sure that J is an isometry and injective. But we still have to show that the inner product on $\mathcal{J}$ is well-defined and positive definite. If $\{u_n\}_n$ and $\{v_n\}_n$ are Cauchy sequences in $\mathcal{H}$, then

$$|\|u_n\|_{\mathcal{H}} - \|u_m\|_{\mathcal{H}}| \le \|u_n - u_m\|_{\mathcal{H}}$$

implies that the sequences $\{\|u_n\|_{\mathcal{H}}\}_n$ and $\{\|v_n\|_{\mathcal{H}}\}_n$ are Cauchy sequences in $I\!R$, and thus convergent and bounded by constants C_u and C_v. But then

$$\begin{array}{rcl}(u_n, v_n)_{\mathcal{H}} - (u_m, v_m)_{\mathcal{H}} & = & (u_n, v_n)_{\mathcal{H}} - (u_n, v_m)_{\mathcal{H}} - (u_m, v_m)_{\mathcal{H}} + (u_n, v_m)_{\mathcal{H}} \\ & = & (u_n, v_n - v_m)_{\mathcal{H}} - (u_m - u_n, v_m)_{\mathcal{H}} \\ & \le & C_u\|v_n - v_m\|_{\mathcal{H}} + C_v\|u_m - u_n\|_{\mathcal{H}}\end{array}$$

proves that $\{(u_n, v_n)_{\mathcal{H}}\}_n$ is a Cauchy sequence in $I\!R$ and thus convergent. Two representatives of a class $\overline{\{u_n\}_n}$ differ just by a zero sequence that does not affect the inner product's value. The proof of definiteness again uses that zero sequences represent zero in $\mathcal{J}$. This finishes the proof of well-definedness of the new inner product.

Thus $\mathcal{J}$ is another pre-Hilbert space that contains an isometric image of $\mathcal{H}$, and we first want to prove that $J(\mathcal{H})$ is dense in $\mathcal{J}$. Let us take an element $\overline{\{u_n\}_n} \in \mathcal{J}$ and use the fact that for each $\epsilon > 0$ there is some $K(\epsilon)$ such that for all $n, m \geq K(\epsilon)$ we have

$$\|u_n - u_m\|_{\mathcal{H}} \leq \epsilon.$$

Now take $m \geq K(\epsilon)$ and the fixed Cauchy sequence $\overline{\{u_m\}_n} = J(u_m)$. Then

$$\|J(u_m) - \overline{\{u_n\}_n}\|_{\mathcal{J}} = \lim_{n\to\infty} \|u_m - u_n\|_{\mathcal{H}} \leq \epsilon$$

proves the density assertion.

We now proceed to prove completeness of $\mathcal{J}$. To do this we have to form a Cauchy sequence $\{\overline{\{u_n^{(m)}\}_n}\}_m$ of equivalence classes $\overline{\{u_n^{(m)}\}_n}$ of Cauchy sequences $\{u_n^{(m)}\}_n \subset \mathcal{H}$. For each $m \in I\!N$ we can use the density property of $\mathcal{H}$ in $\mathcal{J}$ to find an element $v_m \in \mathcal{H}$ such that

$$\|\overline{\{u_n^{(m)}\}_n} - J(v_m)\|_{\mathcal{J}} \leq 1/m.$$

Due to

$$\begin{array}{rcl}
\|v_n - v_m\|_{\mathcal{H}} & = & \|J(v_n) - J(v_m)\|_{\mathcal{J}} \\
& \leq & \|J(v_n) - \overline{\{u_n^{(n)}\}_n}\|_{\mathcal{J}} + \\
& & + \|\overline{\{u_n^{(n)}\}_n} - \overline{\{u_n^{(m)}\}_n}\|_{\mathcal{J}} + \|\overline{\{u_n^{(m)}\}_n} - J(v_m)\|_{\mathcal{J}} \\
& \to & 0
\end{array}$$

for $n, m \to \infty$, the sequence $\{v_m\}_m$ is a Cauchy sequence in $\mathcal{H}$. We now form

$$\begin{array}{rcl}
\|\overline{\{u_n^{(k)}\}_n} - \{v_n\}_n\|_{\mathcal{J}} & \leq & \|\overline{\{u_n^{(k)}\}_n} - J(v_k)\|_{\mathcal{J}} + \|J(v_k) - \{v_n\}_n\|_{\mathcal{J}} \\
& \leq & 1/k + \lim_{n\to\infty} \|v_k - v_n\|_{\mathcal{H}} \\
& \to & 0
\end{array}$$

for $k \to \infty$, proving convergence towards $\{v_n\}_n$.

Now let $A : \mathcal{H} \to \mathcal{N}$ be a linear continuous mapping with values in a complete normed linear space $\mathcal{N}$. If $\overline{\{u_n\}_n}$ is an element of $\mathcal{J}$, we define the extension B on $\overline{\{u_n\}_n}$ by

(*Bmapdef*)

$$B(\overline{\{u_n\}_n}) := \lim_{n\to\infty} A(u_n). \tag{12.2.12}$$

Since A is continuous, it is bounded and due to

$$\|A(u_m) - A(u_n)\|_{\mathcal{N}} \leq \|A\| \|u_m - u_n\|_{\mathcal{H}}$$

the sequence $\{Au_n\}_n$ is a Cauchy sequence in $\mathcal{N}$. But as $\mathcal{N}$ is a Banach space, the sequence is convergent and (12.2.12, *Bmapdef*) is well-defined. Clearly $B \circ J = A$ holds by definition. Any two such extensions must agree on the dense subspace $A(\mathcal{H})$ of $\mathcal{J}$, and since they are continuous, they must agree on all of $\mathcal{J}$. □We add a little application:

Lemma 12.2.13 *If $\mathcal{M}$ is a dense subspace of a Hilbert space $\mathcal{H}$, then the closure of $\mathcal{M}$ is isometrically isomorphic to $\mathcal{H}$.*

Proof: The closure of $\mathcal{M}$ can be identified with a closed subspace $\mathcal{N}$ of $\mathcal{H}$, and we assert that $\mathcal{N} = \mathcal{H}$. To this end, decompose $\mathcal{H}$ into $\mathcal{H} = \mathcal{N} + \mathcal{N}^{-}$ and take an element u from $\mathcal{N}^{-}$. It must be orthogonal to all elements from $\mathcal{M}$, and by continuity of the functional $v \mapsto (u, v)_{\mathcal{H}}$ it must be orthogonal to all of $\mathcal{H}$. Thus it must be zero. □

We further need the **Riesz representation theorem** for continuous linear functionals:

Theorem 12.2.14 *(RieszT) Any continuous linear real-valued functional λ on a Hilbert space $\mathcal{H}$ can be written as*

(RieszRep)

$$\lambda = (\cdot, g_\lambda)_{\mathcal{H}} \qquad (12.2.15)$$

with a unique element $g_\lambda \in \mathcal{H}$. The map $\lambda \mapsto g_\lambda$ is an isometric isomorphism between the **dual Hilbert space** *$\mathcal{H}^*$ of $\mathcal{H}$, consisting of all continuous linear real-valued functionals on $\mathcal{H}$, and $\mathcal{H}$ itself.*

Proof: If $\lambda = 0$, then $g_\lambda = 0$ does the job and is unique. If $\lambda \neq 0$, the kernel $\mathcal{L}$ of λ is not the full space $\mathcal{H}$. It is, however, a closed linear subspace, and thus there is some element $h_\lambda \in \mathcal{L}^{-}$ with $\|h_\lambda\|_{\mathcal{H}} = 1$. Now for each $u \in \mathcal{H}$ the element $\lambda(u)h_\lambda - \lambda(h_\lambda)u$ must necessarily be in $\mathcal{L}$ and thus orthogonal to h_λ. This means

$$\begin{array}{rcl} 0 &=& (h_\lambda, \lambda(u)h_\lambda - \lambda(h_\lambda)u)_{\mathcal{H}}, \\ \lambda(u)(h_\lambda, h_\lambda)_{\mathcal{H}} &=& \lambda(h_\lambda)(u, h_\lambda)_{\mathcal{H}}, \\ \lambda(u) &=& (u, \lambda(h_\lambda)h_\lambda)_{\mathcal{H}}. \end{array}$$

The norm of λ is bounded by

$$\begin{array}{rcl} \|\lambda\|_{\mathcal{H}^*} &:=& \sup\limits_{u \in \mathcal{H} \setminus \{0\}} \dfrac{|\lambda(u)|}{\|u\|_{\mathcal{H}}} \\ &\leq& |\lambda(h_\lambda)| \end{array}$$

due to Cauchy-Schwarz, but using $u = h_\lambda$ in the definition of the norm yields equality. Since we set $g_\lambda := \lambda(h_\lambda)h_\lambda$, we get $\|\lambda\|_{\mathcal{H}^*} = \|g_\lambda\|_{\mathcal{H}}$. Uniqueness of g_λ satisfying (12.2.15, *RieszRep*) is easy to prove, and equally easy is the proof of injectivity and surjectivity of the map $\lambda \mapsto g_\lambda$. □

12.3 Special Functions and Transforms

(*SecSFT*) This is intended as a reference and tutorial for classical formulas involving special functions (e.g.: Gamma, Beta, and Bessel functions) and their transforms. Results on Fourier transforms in general are in section 12.5 (*SecFTRd).* This section, so far, is in raw and unsorted form, because all required formulae are just collected here.

12.3.1 Gamma Function

(*SecGammaFunction*) The **Gamma function** is defined by

(*GammaDef*)

$$? (z) = \int_0^\infty t^{z-1}e^{-t}dt \tag{12.3.1}$$

and has the properties

$$\begin{array}{rcl} ?(z+1) & = & z?(z), \quad z \notin -I\!N \\ ?(k+1) & = & k!, \quad k \in I\!N \\ ?(1/2) & = & \sqrt{\pi}. \end{array}$$

The equation

(*EqGxy*)

$$\int_0^1 u^{x-1}(1-u)^{y-1}du = \frac{?(x)?(y)}{?(x+y)} \tag{12.3.2}$$

for any $x, y > 0$ will be useful.

12.3.2 Volumes and Surface Integrals

(*SecVSI*) The volume of the d-dimensional ball

$$B_r(0) := \{\ x \in I\!R^d\ :\ \|x\|_2 \leq r\ \}$$

of radius r is

(*EqVolBall*)

$$\text{vol } B_r(0) = \frac{r^d\pi^{d/2}}{?(1+d/2)}. \tag{12.3.3}$$

The surface area σ_{d-1} of the $d-1$-dimensional sphere in $I\!R^d$ for $d \geq 1$ is (*VolS*)

$$\sigma_{d-1} = \text{ vol}\,(S^{d-1}) = 2\pi^{d/2}/?\,(d/2). \tag{12.3.4}$$

This follows for $d > 2$ from the representation

$$d\sigma = \prod_{j=1}^{d-1} (\sin \varphi_j)^{d-1-j} d\varphi_j$$

of the surface element $d\sigma$ in terms of the angles

$$\varphi_j \in [0, \pi],\ 1 \leq j \leq d-2,\ \varphi_{d-1} \in [0, 2\pi]$$

and univariate integration, while $d = 1, 2$ are standard.

12.3.3 Bessel Functions

(*SecBesF*) We now consider the function $F(r\|\omega\|_2, d)$ defined by the integral (*EqDefFtd*)

$$F(t, d) := \int_{\|y\|_2=1} e^{-ity\cdot z} dy \tag{12.3.5}$$

over the surface of the unit ball in $I\!R^d$ for $t \geq 0, d \geq 2$, and some $\|z\|_2 = 1,\ z \in I\!R^d$. This integral is invariant under orthogonal transformations Q of $I\!R^d$, as is easily obtainable from replacement of z by Qz. Thus the integral is independent of z, as already indicated by the notation, and we can assume $z = (-1, 0, \ldots, 0)$ for its evaluation. Let σ_{d-1} be the surface area of the $d-1$-sphere, i.e.: the boundary of the unit ball in $I\!R^d$. We now assume $d \geq 3$ and integrate over the surface of the $d-1$-sphere by summing up the integrals over surfaces of $(d-2)$-spheres, splitting $y = (y_1, u)$ and setting $z \cdot y = \cos \varphi$. This yields

$$\begin{aligned} F(t, d) &= \int_{\|y\|_2=1} e^{ity\cdot z} dy \\ &= \int_0^\pi e^{it\cos\varphi} \int_{\|u\|_2^2=1-y_1^2} du d\varphi \\ &= \sigma_{d-2} \int_0^\pi e^{it\cos\varphi} (\sin(\varphi))^{d-2} d\varphi \\ &= \sigma_{d-2} \int_{-1}^1 e^{its} (1-s^2)^{(d-3)/2} ds \end{aligned}$$

and contains an instance of the **Bessel function**

(*JBF*)

$$J_\nu(t) = \frac{(t/2)^\nu}{?\left(\frac{2\nu+1}{2}\right)?\left(\frac{1}{2}\right)}\int_{-1}^{1} e^{its}(1-s^2)^{\frac{2\nu\perp 1}{2}}ds \tag{12.3.6}$$

which is well-defined for $\Re(\nu) > -\frac{1}{2}$. We end up with $\nu = \frac{d-2}{2}$ and get

(*EqFtdRep*)

$$\begin{aligned} F(t,d) &= \sigma_{d-2}\frac{?\left(\frac{d-1}{2}\right)?\left(\frac{1}{2}\right)}{(t/2)^{(d-2)/2}}J_{(d-2)/2}(t) \\ &= 2\pi^{d/2}(t/2)^{-(d-2)/2}J_{(d-2)/2}(t). \end{aligned} \tag{12.3.7}$$

Direct integration shows that this is also valid for $d = 2$ or $\nu = 0$, using $\sigma_0 = 2$.

12.3.4 Power Series of Bessel Functions

(*SecPSBF*) The Bessel function of (12.3.6, *JBF*) has the power series representation

(*JBFP*)

$$J_\nu(t) = \left(\frac{t}{2}\right)^\nu \sum_{j=0}^{\infty}\frac{\left(-\frac{t^2}{4}\right)^j}{j!?\,(\nu+j+1)} \tag{12.3.8}$$

that is valid for all $t \in \mathbb{C}\setminus\{0\}$ and all $\nu \in \mathbb{C}$. The integral representation (12.3.6, *JBF*) is first proven to be identical to the power series representation (12.3.8, *JBFP*) on its domain of definition. Since the power series is convergent everywhere, the general definition of J_ν can then be done by (12.3.8, *JBFP*). We first expand the exponential in

$$\begin{aligned} \int_{-1}^{1} e^{its}(1-s^2)^{(2\nu-1)/2}ds &= \sum_{j=0}^{\infty}\frac{(it)^j}{j!}\int_{-1}^{1} s^j(1-s^2)^{(2\nu-1)/2}ds \\ &= \sum_{j=0}^{\infty}\frac{(it)^{2j}}{2j!}\int_{-1}^{1} s^{2j}(1-s^2)^{(2\nu-1)/2}ds \end{aligned}$$

and use symmetry to cancel the odd powers. The equation (12.3.2, *EqGxy*) will come in handy after the substitution $s^2 = u$. Then

$$\begin{aligned} \sum_{j=0}^{\infty}\frac{(it)^{2j}}{2j!}\int_{-1}^{1} s^{2j}(1-s^2)^{(2\nu-1)/2}ds &= \sum_{j=0}^{\infty}\frac{(it)^{2j}}{2j!}\int_{0}^{1} u^{j-1/2}(1-u)^{(2\nu-1)/2}du \\ &= \sum_{j=0}^{\infty}\frac{?\left(j+\frac{1}{2}\right)?\left(\frac{2\nu+1}{2}\right)}{?\,(j+\nu+1)}\frac{(it)^{2j}}{2j!} \\ &= \sum_{j=0}^{\infty}\frac{?\left(\frac{1}{2}\right)?\left(\frac{2\nu+1}{2}\right)}{j!?\,(j+\nu+1)}\left(-\frac{t^2}{4}\right)^j \end{aligned}$$

uses the same split of ? $(j + \frac{1}{2})$ as before. This can be put int (12.3.6, *JBF*) to yield the power series representation.

Looking at (12.3.8, *JBFP*), we can define a function H_ν by

(*EqHnuDef*)

$$\left(\frac{z}{2}\right)^{-\nu} J_\nu(z) =: H_\nu(z^2/4) = \sum_{k=0}^{\infty} \frac{(-z^2/4)^k}{k!?\,(k+\nu+1)} \tag{12.3.9}$$

for $\nu \in \mathbb{C}$. This function often occurs in the text.

In a very special situation the power series representation (12.3.8, *JBFP*) implies

(*JBh2*)

$$\begin{aligned} J_{-1/2}(t) &= \left(\frac{t}{2}\right)^{-1/2} \sum_{j=0}^{\infty} \frac{\left(-\frac{t^2}{4}\right)^j}{j!?\,(j+1/2)} \\ &= \left(\frac{t}{2}\right)^{-1/2} \sum_{j=0}^{\infty} \frac{(-1)^j t^{2j}}{2^{2j} j!((j-1)/2)((j-3)/2)\dots(1/2)\sqrt{\pi}} \\ &= \left(\frac{t}{2}\right)^{-1/2} \sum_{j=0}^{\infty} \frac{(-1)^j t^{2j}}{(2j)!\sqrt{\pi}} \\ &= \left(\frac{t}{2}\right)^{-1/2} \frac{1}{\sqrt{\pi}} \cos(t) \\ &= \sqrt{\frac{2}{\pi}} \frac{\cos(t)}{\sqrt{t}}, \end{aligned} \tag{12.3.10}$$

and the other Bessel functions with half-integer order are similarly obtainable as linear combinations of elementary functions.

12.3.5 Relations Between Bessel Functions

(*SecRBBF*) By differentiation of the H_ν function from (12.3.9, *EqHnuDef*) we get

(*EqHnuDer*)

$$-\frac{d}{dt}H_\nu(rt) = -\tfrac{d}{dt}\sum_{k=0}^{\infty}\frac{(-rt)^k}{k!?\,(\nu+k+1)} = r\sum_{k=1}^{\infty}\frac{(-rt)^{k-1}}{(k-1)!?\,(\nu+k+1)} = r\sum_{k=0}^{\infty}\frac{(-rt)^k}{k!?\,(\nu+k+2)} = H_{\nu+1}(rt). \tag{12.3.11}$$

and

(*EqHnuDerNu*)

$$\frac{d}{dt}t^\nu H_\nu(rt) = \tfrac{d}{dt}\sum_{k=0}^{\infty}\frac{(-rt)^k t^\nu}{k!?\,(\nu+k+1)} = r\sum_{k=0}^{\infty}\frac{(-r)^k(\nu+k)t^{\nu+k-1}}{k!?\,(\nu+k+1)} = \sum_{k=0}^{\infty}\frac{(-rt)^k t^{\nu-1}}{k!?\,(\nu-1+k+1)} = t^{\nu-1}H_{\nu-1}(rt). \tag{12.3.12}$$

We further need a special identity for Bessel functions:

(*JBI*)

$$J_{\mu+\nu+1}(t) = \frac{t^{\nu+1}}{2^\nu?\,(\nu+1)}\int_0^1 J_\mu(ts)s^{\mu+1}(1-s^2)^\nu ds,\ t>0, \nu>-1, \mu>-\frac{1}{2}. \tag{12.3.13}$$

Since the integral is finite, we can simply insert the power series and get

$$\begin{aligned}
\int_0^1 J_\mu(ts)s^{\mu+1}(1-s^2)^\nu ds &= \int_0^1 \left(\left(\frac{ts}{2}\right)^\mu \sum_{j=0}^\infty \frac{\left(-\frac{(ts)^2}{4}\right)^j}{j!?\,(\mu+j+1)}\right) s^{\mu+1}(1-s^2)^\nu ds \\
&= \sum_{j=0}^\infty \frac{(-1)^j(\frac{t}{2})^{\mu+2j}}{j!?\,(\mu+j+1)} \int_0^1 s^{2\mu+2j+1}(1-s^2)^\nu ds \\
&= \sum_{j=0}^\infty \frac{(-1)^j(\frac{t}{2})^{\mu+2j}}{j!?\,(\mu+j+1)} \frac{1}{2} \int_0^1 r^{\mu+j}(1-r)^\nu dr \\
&= \sum_{j=0}^\infty \frac{(-1)^j(\frac{t}{2})^{\mu+2j}}{j!?\,(\mu+j+1)} \frac{1}{2} \frac{?\,(\mu+j+1)?\,(\nu+1)}{?\,(\mu+\nu+j+2)} \\
&= \left(\sum_{j=0}^\infty \frac{(-1)^j(\frac{t}{2})^{\mu+\nu+1+2j}}{j!?\,(\mu+\nu+j+2)}\right) \frac{2?\,(\nu+1)}{t^{\nu+1}} \\
&= \frac{2?\,(\nu+1)}{t^{\nu+1}} J_{\mu+\nu+1}(t).
\end{aligned}$$

There is a special application in the text for $\nu = 0$ and $\mu = (d-2)/2$, with

(*CSPnuFT*)

$$J_{d/2}(t) = t\int_0^1 J_{(d-2)/2}(ts)s^{d/2}ds. \tag{12.3.14}$$

12.3.6 Bounds on Bessel Functions

(*SecBBF*) We continue with two properties of Bessel functions from [35](*narcowich-ward:92-1*):

(*EqJsqBound*)

$$J_{d/2}^2(z) \;\le\; \frac{2^{d+2}}{\pi z}\,, \quad z > 0 \tag{12.3.15}$$

(*EqJsqInfty*)

$$\lim_{z\to 0} z^{-d} J_{d/2}^2(z) \;=\; \frac{1}{2^d ?^2\,(1+d/2)}. \tag{12.3.16}$$

The second of these follows easily from the power series expansion, since

$$\begin{aligned}
\lim_{z\to 0}\left(\frac{z}{2}\right)^{-\nu} J_\nu(z) &= \frac{1}{?\,(1+\nu)} \\
\lim_{z\to 0} z^{-\nu} J_\nu(z) &= \frac{2^{-\nu}}{?\,(1+\nu)} \\
\lim_{z\to 0}\left(z^{-\nu} J_\nu(z)\right)^2 &= \frac{2^{-2\nu}}{?\,(1+\nu)^2}.
\end{aligned}$$

Unfortunately, equation (12.3.15, *EqJsqBound*) is much more difficult and must (for now) be left to the cited literature. Similarly, there is a weaker, but more general bound

(*EqBFBound*)

$$|J_\nu(x)| \leq 1 \tag{12.3.17}$$

for all $x \in I\!R$ and $\nu \geq 0$ ([1](*abramowitz-stegun:70-1*), 9.1.60, p. 362). Both of the above bounds should combine into the general inequality

(*EqBFBound2*)

$$|J_\nu(|x|)| \leq \mathcal{O}(|x|^{-1/2}), \; x \to \infty \tag{12.3.18}$$

in view of [1](*abramowitz-stegun:70-1*), 9.2.1, p. 364. These things will be added later.

12.3.7 Integrals Involving Bessel Functions

(*SecWSI*) From [1](*abramowitz-stegun:70-1*) 11.4.16, p. 486 we take the moment equations

(*EqMomJnu*)

$$\int_0^\infty t^\mu J_\nu(t)dt = 2^\mu \frac{?\,((\nu+\mu+1)/2)}{?\,((\nu-\mu+1)/2)} \tag{12.3.19}$$

which are valid for $\Re(\nu+\mu) > -1$, $\Re(\mu) < 1/2$. We now use these to derive similar equations for the H_ν functions by

(*EqHnuMom*)

$$\begin{aligned} \int_0^\infty s^\rho H_\nu(s)ds &= \int_0^\infty (z^2/4)^\rho H_\nu(z^2/4)(z/2)dz \\ &= \int_0^\infty (z^2/4)^\rho (z/2)^{-\nu} J_\nu(z)(z/2)dz \\ &= 2^{\nu-1-2\rho}\int_0^\infty z^{2\rho-\nu+1} J_\nu(z)dz \\ &= \frac{?\,(\rho+1)}{?\,(\nu-\rho)} \end{aligned} \tag{12.3.20}$$

for $\rho > -1$ and $\nu > 2\rho + \frac{1}{2}$.

Another citation from [1](*abramowitz-stegun:70-1*) 11.4.41, p. 487 is the **Weber-Schafheitlin integral**

(*EqWeSchaf*)

$$\int_0^\infty t^{\mu-\nu+1} J_\mu(at) J_\nu(bt) dt$$

$$= \left\{ \begin{array}{cc} 0 & 0 < b < a \\ \dfrac{2^{\mu-\nu+1} a^\mu (b^2 - a^2)^{\nu-\mu-1}}{b^\nu ?\,(\nu-\mu)} & 0 < a < b \end{array} \right\} \tag{12.3.21}$$

for $\Re\nu > \Re\mu > -1$ and $a \neq b > 0$.

12.3.8 Bessel Functions of Third Kind

(*SecBFTK*) The Bessel function K_ν of third kind (alias Mcdonald function) is

(*KnuDef*)

$$K_\nu(z) = \frac{\pi^{1/2}(z/2)^\nu}{?\,(\nu+1/2)} \int_1^\infty e^{-zt}(t^2-1)^{\nu-1/2} dt \tag{12.3.22}$$

for $|\arg z| < \pi/2$ and $\Re\nu > -1/2$, and its asymptotics near zero is

(*KnuAsyZero*)

$$K_\nu(z) = \frac{(z/2)^{-\nu}}{?\,(\nu)} + o(1), \tag{12.3.23}$$

while it behaves like

(*KnuAsyInf*)

$$K_\nu(z) = \frac{\sqrt{\pi}}{\sqrt{2z}} e^{-z}(1 + \mathcal{O}(z^{-1})), \tag{12.3.24}$$

near infinity. Due to [1](*abramowitz-stegun:70-1*), 11.4.44, p.488 it is related to the J_ν Bessel functions via the identity

(*EqKJ*)

$$\int_0^\infty \frac{t^{\nu+1} J_\nu(at)}{(t^2+z^2)^{\mu+1}} = \frac{a^\mu z^{\nu-\mu}}{2^\mu ?\,(\mu+1)} K_{\nu-\mu}(az) \tag{12.3.25}$$

for $a, z > 0,\ -1 < \nu < 2\mu + 3/2$. It satisfies the differential equation

(*EqKnuDif*)

$$K_\nu'(z) = K_{\nu-1}(z) - \frac{\nu}{z} K_\nu(z) \tag{12.3.26}$$

due to [1](*abramowitz-stegun:70-1*), 9.6.26, p. 376.

12.4 Lebesgue Integration

(*SecLI*) This section covers some results from Lebesgue integration. We assume some basic knowledge and concentrate on certain specific questions that may not be covered by every course on Lebesgue integration.

12.4.1 L_2 spaces

We want to look at density questions for subspaces of $L_2(\mathbb{R}^d)$. For each continuous function f on $\mathbb{R}^d$ we define the **support** as

$$\text{supp}\, f := \text{ clos}\,\{\, x \in \mathbb{R}^d \ : \ f(x) \neq 0 \,\}$$

and note that any continuous function with compact support clearly is in $L_2(\mathbb{R}^D)$.

Lemma 12.4.1 *(LemBSDense) The space $C_0(\mathbb{R}^d)$ of continuous functions with compact support is a dense subspace of $L_2(\mathbb{R}^d)$.*

Proof: Let a function $g \in \tilde{L}_2(\mathbb{R}^d)$ be given. For any $n \in \mathbb{N}$ we can restrict g to $[-n,n]^d$ and cut off extremely large values to get a function

$$g_n \in L_2[-n,n]^d,\ g(x) = \left\{ \begin{array}{cl} g(x) & x \in [-n,n],\ |g(x)| \leq n \\ 0 & \text{else} \end{array} \right\}$$

with

$$\|g - g_n\|^2_{L_2(\mathbb{R}^d)} \leq \int_{|g(x)|>n} |g(x)|^2 dx \int_{x \notin [-n,n]^d} |g(x)|^2 dx \to 0 \text{ for } n \to \infty.$$

This proves that the bounded L_2 functions with compact support are dense in $L_2(\mathbb{R}^d)$. But in each of the spaces $L_2[-n,n]$ we have density of continuous functions. This can either be proven by Weierstraß type theorems or by approximation with smoothed step functions. □

Lemma 12.4.2 *(LemContShift) The shift operator $S_z \ : \ f(\cdot) \mapsto f(\cdot - z)$ is a continuous function of z near zero in the following sense: for each given $u \in L_2(\mathbb{R}^d)$ and each given $\epsilon > 0$ there is some $\delta > 0$ such that*

$$\|S_z(u) - u\|_{L_2(\mathbb{R}^d)} \leq \epsilon$$

for all $\|z\|_2 \leq \delta$.

Proof: It is easy to see that it suffices to prove the result for functions in $g \in C_0(I\!R^d)$. Let g be supported on $[-K, K]^d$ and use uniform continuity by picking some $\delta > 0$ such that for all $\|x - y\|_2 < \delta < 1$ we have

$$|g(x) - g(y)| < \epsilon(2K + 2)^{-d/2}.$$

Then for $\|z\|_2 < \delta$ we get

$$\begin{array}{rcl} \|Sg - g\|^2_{L_2(I\!R^d)} & = & \int_{I\!R^d} |S_z(g)(x) - g(x)|^2 dx \\ & = & \int_{I\!R^d} |g(x - z) - g(x)|^2 dx \\ & \leq & (\epsilon(2K + 2)^{-d/2})^2 (2K + 2)^d \\ & = & \epsilon^2. \end{array}$$

□

We now want to prove that the space $\mathcal{S}$ of tempered test functions is dense in $L_2(I\!R^d)$. For this, we have to study functions like (12.5.11, *deltaschar*) in some more detail. They are in $\mathcal{S}$ for all positive values of ϵ, and Lemma 12.5.12 (*LemRepro*) tells us that the operation

$$f \mapsto M_\epsilon(f) := \int_{I\!R^d} f(y)\varphi(\epsilon, \cdot - y) dy$$

maps each continuous L_1 function f to a "mollified" function $M_\epsilon(f)$ such that

$$\lim_{\epsilon \to 0} M_\epsilon(f)(x) = f(x)$$

uniformly on compact subsets of $I\!R^d$.

It is common to replace the Gaussian in (12.5.14, *deltarep*) by an infinitely differentiable function with compact support, e.g.

(*Friedmoll*)

$$\varphi(\epsilon, x) = \left\{ \begin{array}{rl} c(\epsilon)\exp(-1/(\epsilon^2 - \|x\|_2^2)) & \|x\|_2 < \epsilon \\ 0 & \|x\|_2 \geq \epsilon \end{array} \right\} \tag{12.4.3}$$

where the constant $c(\epsilon)$ is such that

$$\int_{I\!R^d} \varphi(\epsilon, x) dx = 1$$

holds for all $\epsilon > 0$. This **Friedrich's mollifier** can also be used in the definition of M_ϵ. It has the advantage that Lemma 12.5.12 (*LemRepro*) holds for more general functions, i.e.: for functions which are in L_1 only locally.

We now want to study the action of M_ϵ on L_2 functions. Let $u \in L_2(I\!R^d)$ be given, and apply the Cauchy-Schwarz inequality to

$$M_\epsilon(f)(x) = \int_{I\!R^d} (f(y)\sqrt{\varphi(\epsilon, x-y)})\sqrt{\varphi(\epsilon, x-y)}dy$$

to get

$$\begin{array}{rcl} |M_\epsilon(f)(x)|^2 & \leq & \int_{I\!R^d} |f(y)|^2\varphi(\epsilon, x-y)dy \int_{I\!R^d} \varphi(\epsilon, x-y)dy \\ & = & \int_{I\!R^d} |f(y)|^2\varphi(\epsilon, x-y)dy \end{array}$$

and

$$\int_{I\!R^d} |M_\epsilon(f)(x)|^2 dx \leq \int_{I\!R^d}\int_{I\!R^d} |f(y)|^2\varphi(\epsilon, z)dydz = \int_{I\!R^d} |f(y)|^2 dy$$

such that M_ϵ has norm less than or equal to one in the L_2 norm. It is even more simple to prove the identity

$$(f, M_\epsilon g)_{L_2(R^d)} = (M_\epsilon f, g)_{L_2(R^d)}$$

for all $f, g \in L_2(I\!R^d)$ by looking at the integrals. Here, we used the Fubini theorem on $I\!R^d$ which requires some care, but there are no problems because everything can either be done with a Friedrich's mollifier, or be done on sufficiently large compact sets first, letting the sets tend to $I\!R^d$ later.

We now use a Friedrich's mollifier to study the L_2 error of the mollification. This is very similar to the arguments we already know. The error is representable pointwise as

$$f(x) - M_\epsilon(f)(x) = \int_{I\!R^d} (f(x) - f(y))\varphi(\epsilon, x-y)dy$$

and we can use the Cauchy-Schwarz inequality to get

$$|f(x) - M_\epsilon(f)(x)|^2 \leq \int_{\|x-y\|_2 < \epsilon} |f(x) - f(y)|^2\varphi(\epsilon, x-y)dy.$$

This can be integrated to get

$$\int_{I\!R^d} |f(x) - M_\epsilon(f)(x)|^2 dx \leq \int_{\|z\|_2 < \epsilon} \varphi(\epsilon, z)\int_{I\!R^d} |f(y+z) - f(y)|^2 dydz,$$

and we use the continuity of the shift operator as proven in Lemma 12.4.2 (*LemContShift)* to make this as small as we want by picking a suitably small ϵ. This shows

$$\lim_{\epsilon\to 0} \|f - M_\epsilon(f)\|_{L_2(I\!R^d)} = 0$$

and proves

Lemma 12.4.4 *(FTD) The space $\mathcal{S}$ of test functions is dense in $L_2(I\!R^d)$.*
$\square$

Lemma 12.4.5 *(FTDC) The space $C_0^\infty(I\!R^d)$ of compactly supported infinitely differentiable functions is dense in $L_2(I\!R^d)$.*

Proof: We can use Lemma 12.4.1 (*LemBSDense)* to go over from an $f \in \tilde{L}_2(I\!R^d)$ to a compactly supported function, and then we can use Friedrich's mollifier to generate an infinitely differentiable function. Both processes work with arbitrarily small L_2 errors. $\square$

12.5 Fourier Transforms on $I\!R^d$

(*SecFTRd*) This section contains the necessary definitions and results on Fourier transforms in $I\!R^d$ together with their generalizations. Since we do not want to rely on books on distributions, we develop the relevant machinery here.

12.5.1 Fourier Transforms of Tempered Test Functions

There are two major possibilities to pick a space $\mathcal{S}$ of test functions on $I\!R^d$ to start with, and we take Laurent Schwartz's **tempered test functions** that are verbally defined as complex-valued functions on $I\!R^d$ whose partial derivatives exist for all orders and decay faster than any polynomial towards infinity. Such functions clearly define a linear space over $\mathbb{C}$ of functions on $I\!R^d$, and the standard notation is $\mathcal{S}$.

Definition 12.5.1 *(DefFT) For a tempered test function $u \in \mathcal{S}$, the* **Fourier transform** *is*

(FT)

$$\widehat{u}(\omega) := (2\pi)^{-d/2} \int_{I\!R^d} u(x)e^{-ix\cdot\omega}dx, \tag{12.5.2}$$

where ω varies in $I\!R^d$ and $x\cdot\omega$ is shorthand for the scalar product $x^T\omega = \omega^T x$ to avoid the ^{T} symbol in the exponent. Since the definition even works for general $u \in L_1(I\!R^d)$, it is well-defined on $\mathcal{S}$ and clearly linear. Note that we use the **symmetric** form of the transform and do not introduce a factor 2π in the exponent of the exponential. This sometimes makes comparisons to other presentations somewhat difficult; in particular, our notation induces a factor in the usual convolution theorem for Fourier transforms, as proven in the next section

12.5.2 Convolutions

(*SecFTConv*) As far as the integral is well-defined, the **convolution** $f * g$ of two complex-valued functions f, g on $I\!R^d$ is defined as the function

$$(f * g)(x) := \int_{I\!R^d} f(y)g(x-y)dy.$$

This is a linear operation in both f and g, and it is well defined for all functions in $\mathcal{S}$ and all functions in $L_2(I\!R^d)$. We check the Fourier transform of the convolution of tempered test functions $f, g \in \mathcal{S}$ and apply Fubini's theorem for this:

(*EqFTC*)

$$\begin{aligned}
\widehat{f * g}(\omega) &= (2\pi)^{-d/2}\int_{I\!R^d} (f * g)(x)e^{-ix\cdot\omega}dx \\
&= (2\pi)^{-d/2}\int_{I\!R^d}\int_{I\!R^d} f(y)g(x-y)dye^{-ix\cdot\omega}dx \\
&= (2\pi)^{-d/2}\int_{I\!R^d}\int_{I\!R^d} f(y)g(x-y)e^{-iy\cdot\omega}e^{-i(x-y)\cdot\omega}dydx \\
&= (2\pi)^{+d/2}(2\pi)^{-d/2}\int_{I\!R^d} f(y)e^{-iy\cdot\omega}dy(2\pi)^{-d/2}\int_{I\!R^d} g(z)e^{-iz\cdot\omega}dz \\
&= (2\pi)^{d/2}(\widehat{f}\widehat{g})(\omega).
\end{aligned} \tag{12.5.3}$$

12.5.3 Identities for Fourier Transforms

(*SecFTIFT*) Here are some handy identities that are easily provable:

(*EqFTShift*)

$$\widehat{f(\cdot - y)}(\omega) = e^{-iy\cdot\omega}\widehat{f(\cdot)}(\omega) \tag{12.5.4}$$

(*EqFTScale*)

$$\widehat{f(\cdot/r)}(\omega) = r^d\widehat{f(\cdot)}(r\omega) \tag{12.5.5}$$

They do not only hold for tempered test functions from $\mathcal{S}$, but usually generalize to all function spaces to which the Fourier transform can be extended.

12.5.4 Fourier Transforms of Gaussians

(*SecPDG*) To get used to calculations of Fourier transforms, let us start with the **Gaussian** $u_\gamma(x) = \exp(-\gamma\|x\|_2^2)$ for $\gamma > 0$, which clearly is in the space of test functions, since all derivatives are polynomials multiplied with the Gaussian itself. As a byproduct we shall get that the Gaussian is positive definite on $\mathbb{R}^d$. Fortunately, the Gaussian can be written as a d-th power of the entire analytic function $\exp(-\gamma z^2)$, and we can thus work on $\mathbb{C}^d$ instead of $\mathbb{R}^d$. We simply use substitution in

$$\begin{aligned}
\widehat{u_\gamma}(i\omega) &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-\gamma\|x\|_2^2} e^{x\cdot\omega} dx \\
&= (2\pi)^{-d/2} e^{\|\omega\|_2^2/4\gamma} \int_{\mathbb{R}^d} e^{-\|\sqrt{\gamma}x-\omega/2\sqrt{\gamma}\|_2^2} dx \\
&= (2\pi\gamma)^{-d/2} e^{\|\omega\|_2^2/4\gamma} \int_{\mathbb{R}^d} e^{-\|y\|_2^2} dy
\end{aligned}$$

and are done up to the evaluation of the dimension-dependent constant

$$\int_{\mathbb{R}^d} e^{-\|y\|_2^2} dy =: c^d$$

which is a d-th power, because the integrand factorizes nicely. We calculate c^2 by using polar coordinates and get

$$\begin{aligned}
c^2 &= \int_{\mathbb{R}^2} e^{-\|y\|_2^2} dy \\
&= \int_0^{2\pi} \int_0^\infty e^{-r^2} r\, dr\, d\varphi \\
&= 2\pi \int_0^\infty e^{-r^2} r\, dr \\
&= -\pi \int_0^\infty (-2r) e^{-r^2}\, dr \\
&= \pi.
\end{aligned}$$

This proves the first assertion of

Theorem 12.5.6 *(GaussPD) The Gaussian*

$$u_\gamma(x) = \exp(-\gamma\|x\|_2^2)$$

has Fourier transform

(GFT)

$$\widehat{u_\gamma}(\omega) = (2\gamma)^{-d/2} e^{-\|\omega\|_2^2/4\gamma} \tag{12.5.7}$$

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

Proof: Let us first invert the Fourier transform by setting $\beta := 1/4\gamma$ in (12.5.7, *GFT*):

$$\begin{aligned}\exp(-\beta\|\omega\|_2^2) &= (4\pi\beta)^{-d/2}\int_{I\!R^d} e^{-\|x\|_2^2/4\beta}e^{-ix\cdot\omega}dx\\ &= (2\pi)^{-d/2}\int_{I\!R^d}(2\beta)^{-d/2}e^{-\|x\|_2^2/4\beta}e^{+ix\cdot\omega}dx.\end{aligned}$$

Then take any set $X = \{x_1, \ldots, x_M\} \subset I\!R^d$ of M distinct points and any vector $\alpha \in I\!R^M$ to form

$$\begin{aligned}\alpha^T A_X\alpha &= \sum_{j,k=1}^{M}\alpha_j\alpha_k\exp(-\beta\|x_j-x_k\|_2^2)\\ &= \sum_{j,k=1}^{M}\alpha_j\alpha_k(4\pi\beta)^{-d/2}\int_{I\!R^d} e^{-\|x\|_2^2/4\beta}e^{-ix\cdot(x_j-x_k)}dx\\ &= (4\pi\beta)^{-d/2}\int_{I\!R^d} e^{-\|x\|_2^2/4\beta}\sum_{j,k=1}^{M}\alpha_j\alpha_k e^{-ix\cdot(x_j-x_k)}dx\\ &= (4\pi\beta)^{-d/2}\int_{I\!R^d} e^{-\|x\|_2^2/4\beta}\left|\sum_{j=1}^{M}\alpha_j e^{-ix\cdot x_j}\right|^2 dx \geq 0.\end{aligned}$$

This proves positive semidefiniteness of the Gaussian. To prove definiteness, we can assume

$$f(x) := \sum_{j=1}^{M}\alpha_j e^{-ix\cdot x_j} = 0$$

for all $x \in I\!R^d$ and have to prove that all coefficients α_j vanish. Taking derivatives at zero, we get

$$0 = D^\beta f(0) = \sum_{j=1}^{M}\alpha_j(-ix_j)^\beta,$$

and this is a homogeneous system for the coefficients α_j whose coefficient matrix is a generalized Vandermonde matrix, possibly transposed and with scalar multiples for rows or columns. This proves the assertion in one dimension, where the matrix corresponds to the classical Vandermonde matrix. The multivariate case reduces to the univariate case by picking a nonzero vector $y \in I\!R^d$ that is not orthogonal to any of the finitely many differences $x_j - x_k$ for $j \neq k$. Then the real values $y \cdot x_j$ are all distinct for $j = 1, \ldots, M$ and one can consider the univariate function

$$g(t) := f(ty) = \sum_{j=1}^{M}\alpha_j e^{-ity\cdot x_j} = 0$$

which does the job in one dimension. □

Note that the Gaussian is mapped to itself by the Fourier transform, if we pick $\gamma = 1/2$ in (12.5.7, *GFT*). We shall use the Gaussian's Fourier transform in the proof of the fundamental **Fourier Inversion Theorem**:

Theorem 12.5.8 *(FTTS) The Fourier transform is bijective on $\mathcal{S}$, and its inverse is the transform*

(IFT)

$$\check{u}(x) := (2\pi)^{-d/2} \int_{I\!R^d} u(\omega) e^{ix\cdot\omega} d\omega. \tag{12.5.9}$$

Proof: The multivariate derivative D^α of $\widehat{u}$ can be taken under the integral sign, because u is in $\mathcal{S}$. Then

$$(D^\alpha \widehat{u})(\omega) = (2\pi)^{-d/2} \int_{I\!R^d} u(x)(-ix)^\alpha e^{-ix\cdot\omega} dx,$$

and we multiply this by ω^β and use integration by parts

$$\begin{array}{rcl}
\omega^\beta (D^\alpha \widehat{u})(\omega) & = & (2\pi)^{-d/2} \int_{I\!R^d} u(x)(-ix)^\alpha (i)^\beta (-i\omega)^\beta e^{-ix\cdot\omega} dx \\
& = & (2\pi)^{-d/2} \int_{I\!R^d} u(x)(-ix)^\alpha (i)^\beta \frac{d^\beta}{dx^\beta} e^{-ix\cdot\omega} dx \\
& = & (2\pi)^{-d/2} (-1)^{|\alpha|+|\beta|} i^{\alpha+\beta} \int_{I\!R^d} e^{-ix\cdot\omega} \frac{d^\beta}{dx^\beta} (u(x) x^\alpha) dx
\end{array}$$

to prove that $\widehat{u}$ lies in $\mathcal{S}$, because all derivatives decay faster than any polynomial towards infinity. The second assertion follows from the Fourier inversion formula

(IFT2)

$$u(x) := (2\pi)^{-d/2} \int_{I\!R^d} \widehat{u}(\omega) e^{ix\cdot\omega} d\omega \tag{12.5.10}$$

that we now prove for all $u \in \mathcal{S}$. This does not work directly if we naively put the definition of $\widehat{u}$ into the right-hand-side, because the resulting multiple integral does not satisfy the assumptions of Fubini's theorem. We have to do a regularization of the integral, and since this is a very useful trick, we write it out in some detail:

$$\begin{array}{rcl}
(2\pi)^{-d/2} \int_{I\!R^d} \widehat{u}(\omega) e^{ix\cdot\omega} d\omega & = & (2\pi)^{-d} \int_{I\!R^d} \int_{I\!R^d} u(y) e^{i(x-y)\cdot\omega} dy d\omega \\
& = & \lim\limits_{\epsilon \searrow 0} (2\pi)^{-d} \displaystyle\int_{I\!R^d} \int_{I\!R^d} u(y) e^{i(x-y)\cdot\omega - \epsilon\|\omega\|_2^2} dy d\omega \\
& = & \lim\limits_{\epsilon \searrow 0} (2\pi)^{-d} \displaystyle\int_{I\!R^d} \left(\int_{I\!R^d} e^{i(x-y)\cdot\omega - \epsilon\|\omega\|_2^2} d\omega \right) u(y) dy \\
& = & \lim\limits_{\epsilon \searrow 0} \displaystyle\int_{I\!R^d} \varphi(\epsilon, x-y) u(y) dy
\end{array}$$

with

(*deltaschar*)

$$\varphi(\epsilon, z) := (2\pi)^{-d} \int_{I\!R^d} e^{iz\cdot\omega - \epsilon\|\omega\|_2^2} d\omega. \tag{12.5.11}$$

The proof is completed by application of the following result that is handy in many contexts: □

Lemma 12.5.12 *(LemRepro) The family of functions $\varphi(\epsilon, z)$ of (12.5.11, deltaschar) approximates the point evaluation functional in the sense*

(Repro)

$$u(x) = \lim_{\epsilon\searrow 0} \int_{I\!R^d} \varphi(\epsilon, x-y)u(y)dy \tag{12.5.13}$$

for all functions u that are in $L_1(I\!R^d)$ and continuous around x.

Proof: We first remark that the definition of φ is a disguised form of the inverse Fourier transform equation of the Gaussian. Thus we get

(*deltarep*)

$$\varphi(\epsilon, x) = (4\pi\epsilon)^{-d/2} e^{-\|x\|_2^2/4\epsilon} \tag{12.5.14}$$

and

$$\int_{I\!R^d} \varphi(\epsilon, x)dx = (4\pi\epsilon)^{-d/2} \int_{I\!R^d} e^{-\|x\|_2^2/4\epsilon} dx = 1.$$

To prove (12.5.13, *Repro*), we start with some given $\delta > 0$ and first find some ball $B_\rho(x)$ of radius $\rho(\delta)$ around x such that $|u(x) - u(y)| \le \delta/2$ holds uniformly for all $y \in B_\rho(x)$. Then we split the integral in

$$\begin{array}{rcl} |u(x) - \int_{I\!R^d} \varphi(\epsilon, x-y)u(y)dy| & = & |\int_{I\!R^d} \varphi(\epsilon, x-y)(u(x)-u(y))dy| \\ & \le & \int_{\|y-x\|_2 \le \rho} \varphi(\epsilon, x-y)|u(x)-u(y)|dy \\ & & + \int_{\|y-x\|>\rho} \varphi(\epsilon, x-y)|u(x)-u(y)|dy \\ & \le & \delta/2 + (4\pi\epsilon)^{-d/2} e^{-\rho^2/4\epsilon} 2\|u\|_1 \\ & \le & \delta \end{array}$$

for all sufficiently small ϵ. □

Due to the Fourier inversion formula, we now know that the Fourier transform is bijective on $\mathcal{S}$. We want relate the Fourier transform to the L_2 inner product, but we have to use the latter over $C\!\!\!\!I$ to account for the possibly complex values of the Fourier transform. Furthermore, we have good reasons to define the inner product as

$$(f,g)_{L_2(\mathbb{R}^d)} := (2\pi)^{-d/2}\int_{\mathbb{R}^d} f(x)\overline{g(x)}dx \qquad (Ltwodef) \qquad (12.5.15)$$

with a factor that simplifies some of the subsequent formulae. Fubini's theorem easily proves the identity

$$(v,\widehat{u})_{L_2(\mathbb{R}^d)} = (2\pi)^{-d}\int_{\mathbb{R}^d} v(x)\int_{\mathbb{R}^d}\overline{u(y)}e^{+ix\cdot y}dydx = (\check{v},u)_{L_2(\mathbb{R}^d)}$$

for all test functions $u,v\in\mathcal{S}$. Setting $v=\widehat{w}$ we get **Plancherel's equation**

$$(\widehat{w},\widehat{u})_{L_2(\mathbb{R}^d)} = (w,u)_{L_2(\mathbb{R}^d)} \qquad (PlanEq) \qquad (12.5.16)$$

for the Fourier transform on $\mathcal{S}$, proving that the Fourier transform is isometric on $\mathcal{S}$ as a subspace of $L_2(\mathbb{R}^d)$.

The Fourier transform clearly exists pointwise for functions in $L_1(\mathbb{R}^d)$, and we have

Lemma 12.5.17 *(FTLoneLem) The Fourier transform maps L_1 functions into continuous L_∞ functions on $\mathbb{R}^d$. The Fourier transform of a test function f is real-valued, if and only if f satisfies $\overline{f(-\cdot)} = f(\cdot)$.*

Proof. : It is easy to see that

$$|\widehat{u}(\omega)| \le (2\pi)^{-d/2}\int_{\mathbb{R}^d}|u(x)|dx,$$

and the continuity follows from the theorem on majorized convergence of Lebesgue integrals when applied to $\omega_n\to\omega$ and

$$\widehat{u}(\omega_n)-\widehat{u}(\omega) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} u(x)\left(e^{-ix\cdot\omega_n}-e^{-ix\cdot\omega}\right)dx,$$

because the integrand is in L_1. The final assertion is a consequence of

$$\begin{array}{rcl} \overline{\widehat{u(\cdot)}(\omega)} & = & (2\pi)^{-d/2}\int_{\mathbb{R}^d} u(x)e^{ix\cdot\omega}dx \\ & = & (2\pi)^{-d/2}\int_{\mathbb{R}^d} u(-x)e^{-ix\cdot\omega}dx \\ & = & \widehat{u(-\cdot)}(\omega) \end{array}$$

and its counterpart for the inverse Fourier transform. □

12.5.5 Fourier Transforms of Functionals

With Plancherel's equation in mind, let us look at the linear functional

$$\lambda_u(v) := (u, v)_{L_2(\mathbb{R}^d)}$$

on $\mathcal{S}$. We see that

$$\lambda_{\widehat{u}}(v) = (\widehat{u}, v)_{L_2(\mathbb{R}^d)} = (u, \check{v})_{L_2(\mathbb{R}^d)} = \lambda_u(\check{v})$$

holds. A proper definition of the Fourier transform for functionals λ_u should be in line with the functions u that represent them, and thus we should define

$$\widehat{\lambda_u} := \lambda_{\widehat{u}}$$

or in more generality

$$\widehat{\lambda}(v) := \lambda(\check{v})$$

for all $v \in \mathcal{S}$. Since the space $\mathcal{S}$ of test functions is quite small, its dual, the space of linear functionals on $\mathcal{S}$, is quite large. In particular, the functionals of the form λ_u are defined on all of $\mathcal{S}$, if u is a **tempered function**. The latter form the space $\mathcal{K}$ of all continuous functions on $\mathbb{R}^d$ that grow at most polynomially for arguments tending to infinity.

Definition 12.5.18 *The Fourier transform of a linear functional λ on $\mathcal{S}$ is the linear functional $\widehat{\lambda}$ on $\mathcal{S}$ defined by*

$$\widehat{\lambda}(v) := \lambda(\check{v}) \text{ or } \widehat{\lambda}(\widehat{v}) := \lambda(v)$$

for all $v \in \mathcal{S}$. If the latter can be represented in the form λ_w with a tempered function $w \in \mathcal{K}$, we say that w is the Fourier transform of λ and write $w = \widehat{\lambda}$. The **generalized Fourier transform** *of a tempered function $u \in \mathcal{K}$ is the Fourier transform $\widehat{\lambda_u}$ of the functional λ_u.*

Example 12.5.19 *(ExDelta) The functional $\delta_x(v) := v(x)$ has the form*

$$\delta_x(v) = v(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{v}(\omega) e^{+ix\cdot\omega} d\omega,$$

and its Fourier transform is of the form λ_{u_x} with

$$u_x(\omega) = \widehat{\delta_x}(\omega) = e^{-ix\cdot\omega}.$$

Here, the normalization of the L_2 inner product (12.5.15, Ltwodef) pays off. Note that the Fourier transform is not a test function, but rather a tempered function from $\mathcal{K}$ and in particular a bounded function. The functional $\delta := \delta_0$ has the Fourier transform $u_0 = 1$.

Example 12.5.20 *(Exlxma) A very important class of functionals for our purposes consists of the space $\mathcal{P}_\Omega^- = (\mathbb{P}_m^d)_{\mathbb{R}^d}^-$ of functionals of the form (3.3.1, Deflxma) that vanish on $\mathbb{P}_m^d$. Their action on a test function v is*

$$\begin{aligned}
\lambda_{X,M\alpha}(v) &= \sum_{j=1}^{M} \alpha_j v(x_j) \\
&= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{v}(\omega) \sum_{j=1}^{M} \alpha_j e^{ix_j \cdot \omega} d\omega \\
&= \widehat{\lambda}_{X,M\alpha}(\widehat{v})
\end{aligned}$$

such that the Fourier transform of the functional $\lambda_{X,M\alpha}$ is the functional generated by the bounded function

$$\widehat{\lambda}_{X,M,\alpha}(\omega) = \sum_{j=1}^{M} \alpha_j e^{-ix_j \cdot \omega}.$$

If we expand the exponential into its power series, we see that

$$\begin{aligned}
\widehat{\lambda}_{X,M,\alpha}(\omega) &= \sum_{k=0}^{\infty} \sum_{j=1}^{M} \alpha_j (-ix_j \cdot \omega)^k / k! \\
&= \sum_{k=m}^{\infty} \sum_{j=1}^{M} \alpha_j (-ix_j \cdot \omega)^k / k!
\end{aligned}$$

since the functional vanishes on $\mathbb{P}_m^d$. Thus $\widehat{\lambda}_{X,M,\alpha}(\omega)$ has a zero of order at least m in the origin. If the functional $\lambda_{X,M\alpha}$ itself were representable by a function u, the function u should be L_2-orthogonal to all polynomials from $\mathbb{P}_m^d$. We shall use both of these facts later.

Example 12.5.21 *(ExFTPol) The monomials x^α are in the space $\mathcal{K}$, and thus they should at least have generalized Fourier transforms in the sense of functionals. This can easily be verified via*

$$\begin{aligned}
\left(-i\tfrac{d}{dx}\right)^\alpha v(x) &= \left(-i\tfrac{d}{dx}\right)^\alpha (2\pi)^{-d/2} \textstyle\int_{\mathbb{R}^d} \widehat{v}(\omega) e^{+ix\cdot\omega} d\omega \\
&= (2\pi)^{-d/2} \textstyle\int_{\mathbb{R}^d} \widehat{v}(\omega)(-i \cdot i\omega)^\alpha e^{+ix\cdot\omega} d\omega \\
&= (2\pi)^{-d/2} \textstyle\int_{\mathbb{R}^d} \widehat{v}(\omega) \omega^\alpha e^{+ix\cdot\omega} d\omega,
\end{aligned}$$

and the associated functional is

$$v \mapsto \left(-i\frac{d}{dx}\right)^\alpha v(x)$$

at $x = 0$.

12.5.6 Fourier Transform in $L_2(I\!R^d)$

The test functions from $\mathcal{S}$ are dense in $L_2(I\!R^d)$ (see Lemma 12.4.4 (*FTD*) for details), and thus we have

Theorem 12.5.22 *(FLtwoT) The Fourier transform has an L_2-isometric extension from the space $\mathcal{S}$ of tempered test functions to $L_2(I\!R^d)$. The same holds for the inverse Fourier transform, and both extensions are inverses of each other in $L_2(I\!R^d)$. Furthermore, Plancherel's equation (12.5.16, PlanEq) holds in $L_2(I\!R^d)$.* □

Note that this result does not allow to use the Fourier transform formula (or its inverse) in the natural pointwise form. For any $f \in L_2(I\!R^d)$ one first has to provide a sequence of test functions $v_n \in \mathcal{S}$ that converges to f in the L_2 norm for $n \to \infty$, and then, by continuity, the image $\hat{f}$ of the Fourier transform is uniquely defined almost everywhere by

$$\lim_{n\to\infty} \|\hat{f} - \widehat{v_n}\|_{L_2(I\!R^d)} = 0.$$

This can be done via Friedrich's mollifiers as defined in (12.4.3, *Friedmoll*), replacing the Gaussian in the representation (12.5.14, *deltarep*) by a compactly supported infinitely differentiable function.

A more useful characterization of $\hat{f}$ is the variational equation

$$(\hat{f}, v)_{L_2(I\!R^d)} = (f, \check{v})_{L_2(I\!R^d)}$$

for all test functions $v \in \mathcal{S}$, or, by continuity, all functions $v \in L_2(I\!R^d)$. This is an equivalent form of Plancherel's equation

(*PlEqL*)

$$(\hat{f}, \hat{v})_{L_2(I\!R^d)} = (f, v)_{L_2(I\!R^d)} \qquad (12.5.23)$$

for all $f, v \in \tilde{L}_2(I\!R^d)$. Some definitions of generalized Fourier transforms use such variational equations to define $\hat{f}$ by (12.5.23, *PlEqL*) for all v from a subspace of $\mathcal{S}$.

12.6 Sobolev Spaces

(*SecSob*) This section contains definitions of Sobolev spaces and proves Sobolev's embedding theorems.

13 Appendix

13.1 Basis Functions

(*SecBF*) Here we try to give a complete list (up to this date) of the available conditionally positive definite functions with their transforms and their recursion formulas. Proofs are either in the main text or in section 12.3 (*SecSFT*) of the appendix.

$\phi(r)$	Parameters	m
r^β	$\beta > 0,\ \beta \notin 2\mathbb{N}$	$m \geq \lceil \beta/2 \rceil$
$r^\beta \log r$	$\beta > 0,\ \beta \in 2\mathbb{N}$	$m > \beta/2$
$(r^2+c^2)^{\beta/2}$	$\beta > 0,\ \beta \notin 2\mathbb{N}$	$m \geq \lceil \beta/2 \rceil$

Table 8: Conditionally Positive Definite Functions (*TCPDFct2*)

$\phi(r)$	Parameters	Smoothness	Dimension	Name/Reference
$e^{-\beta r^2}$	$\beta > 0$	$C^\infty(\mathbb{R}^d)$	$d < \infty$	Gaussian
$(r^2+c^2)^{\beta/2}$	$\beta < 0$	$C^\infty(\mathbb{R}^d)$	$d < \infty$	inv. Multiquadric
$r^\nu K_\nu(r)$	$\nu > 0$	$C^{\lfloor \nu \rfloor}$	$d < \infty$	Sobolev spline
$(1-r)_+^2(2+r)$		C^0	$d \leq 3$	Wu [47](*wu:95-2*)
$(1-r)_+^4(1+4r)$		C^2	$d \leq 3$	Wendland [46](*wendland:95-1*)

Table 9: Unconditionally Positive Definite Functions (*TPDFct2*)

$\phi(r)$	Transform
$e^{-\beta r^2}$	
$(r^2+c^2)^{\beta/2}$	
$r^\nu K_\nu(r)$	
$(1-r)_+^2(2+r)$	
$(1-r)_+^4(1+4r)$	

Table 10: Transforms (*TFT*)

13.2 MATLAB routines

Here we provide the MATLAB sources required to do the examples of section 2.5 (*SecExamples*).

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions.* Dover, New York, 1970.

[2] J.H. Ahlberg, E.N. Nilson, and J.L. Walsh. *The theory of splines and their applications*, volume 38 of *Mathemtics in science and engineering.* Academic Press, 1986.

[3] R. Askey. Radial characteristic functions. Technical Report 1262, Univ. of Wisconsin, 1973. MRC Technical Sum. Report.

[4] M. Atteia. *Hilbertian Kernels and Spline Functions.* North-Holland, Amsterdam, 1992. Studies in Computational Mathematics (4).

[5] R.K. Beatson and G.N. Newsam. Fast evaluation of radial basis functions: I. Advances in the theory and applications of radial basis functions. *Comput. Math. Appl.*, 24(12):7–19, 1992.

[6] B. Bojanov and H. Woźniakowski, editors. *Optimal Recovery.* Nova Science Publishers, New York, 1992.

[7] C. de Boor, R.A. DeVore, and A. Ron. The structure of finitely generated shift–invariant spaces in $L_2(\mathbb{R}^d)$. *J. Funct. Anal.*, pages 37–78, 1994.

[8] D. Braess. *Nonlinear Approximation Theory.* Springer, Berlin, 1986.

[9] E.W. Cheney. *Introduction to Approximation Theory.* McGraw–Hill, Maidenhead, 1966.

[10] 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.

[11] J. Duchon. Sur l'erreur d'interpolation des fonctions de plusieurs variables pas les D^m–splines. *Rev. Française Automat. Informat. Rech. Opér. Anal. Numer.*, 12:325–334, 1978.

[12] J. Duchon. Splines minimizing rotation–invariate semi–norms in sobolev spaces. In W. Schempp and K. Zeller, editors, *Constructive Theory of Functions of Several Variables*, pages 85–100. Springer, Berlin–Heidelberg, 1979.

[13] M.S. Floater and A. Iske. Multistep scattered data interpolation using compactly supported radial basis functions. To appear in *Journal of Computational and Applied Mathematics*, 1995.

[14] M.S. Floater and A. Iske. Scattered data analysis and industrial applications. In F. Fontanella, K. Jetter, and P.-J. Laurent, editors, *Advanced Topics in Multivariate Approximation*, pages 1–10. World Scientific Publishing, 1996.

[15] R. Franke. Scattered data interpolation. Test of some methods. *Mathematics of Computation*, pages 181–200, 1982.

[16] M. Golomb and H.F. Weinberger. Optimal approximation and error bounds. In R.E. Langer, editor, *On Numerical Approximation*, pages 117–190. The University of Wisconsin Press, Madison, 1959.

[17] R.L. Hardy. Multiquadric equations of topography and other irregular surfaces. *J. Geophys. Res.*, 76:1905–1915, 1971.

[18] A. Iske. *Charakterisierung bedingt positiv definiter Funktionen für multivariate Interpolationsmethoden mit radialen Basisfunktionen.* Dissertation, Universität Göttingen, 1994.

[19] A. Iske. Characterization of function spaces associated with conditionally positive definite functions. In T. Lyche M. Daehlen and L.L. Schumaker, editors, *Mathematical Methods for Curves and Surfaces*, pages 265–270. Vanderbilt University Press, Nashville, TN, 1995.

[20] W.R. Madych and S.A. Nelson. Multivariate interpolation: a variational theory. Manuscript, 1983.

[21] W.R. Madych and S.A. Nelson. Multivariate interpolation and conditionally positive definite functions. *Approximation Theory and its Applications*, 4:77–89, 1988.

[22] W.R. Madych and S.A. Nelson. Error bounds for multiquadric interpolation. In C.K. Chui, L.L. Schumaker, and J.D. Ward, editors, *Approximation Theory VI*, pages 413–416. Academic Press, Boston, 1989.

[23] W.R. Madych and S.A. Nelson. Multivariate interpolation and conditionally positive definite functions II. *Mathematics of Computation*, 54:211–230, 1990.

[24] W.R. Madych and S.A. Nelson. Bounds on multivariate polynomials and exponential error estimates for multiquadric interpolation. *Journal of Approximation Theory*, 70:94–114, 1992.

[25] G. Meinardus. *Approximation of functions: theory and numerical methods*. Springer, Berlin, 1967.

[26] H. Meschkowski. *Hilbertsche Räume mit Kernfunktion*. Springer, Berlin, 1962.

[27] C.A. Micchelli. Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constructive Approximation*, 2:11–22, 1986.

[28] C.A. Micchelli and T.J. Rivlin. A survey of optimal recovery. In C.A. Micchelli and T.J. Rivlin, editors, *Optimal Estimation in Approximation Theory*, pages 1–54. Plenum Press, 1977.

[29] C.A. Micchelli and T.J. Rivlin. Optimal recovery of best approximations. *Resultate der Mathematik*, 3:25–32, 1978.

[30] C.A. Micchelli and T.J. Rivlin. Lectures on optimal recovery. In P.R. Turner, editor, *Numerical Analysis, Lancaster 1984*, pages 12–93. Springer Verlag, Lecture Notes in Mathematics 1129, 1984.

[31] C.A. Micchelli, T.J. Rivlin, and S. Winograd. Optimal recovery of smooth functions approximations. *Numerische Mathematik*, 260:191–200, 1976.

[32] F.J. Narcowich and J.D. Ward. Norm of inverses and condition numbers for matrices associated with scattered data. *Journal of Approximation Theory*, 64:69–94, 1991.

[33] F.J. Narcowich and J.D. Ward. Norms of inverses for matrices associated with scattered data. In P.J. Laurent, A. Le Méhauté, and L.L. Schumaker, editors, *Curves and Surfaces*, pages 341–348. Academic Press, Boston, 1991.

[34] F.J. Narcowich and J.D. Ward. Generalized Hermite interpolation via matrix–valued conditionally positive definite functions. Technical Report 280, Department of Mathematics, Texas A & M University, 1992.

[35] F.J. Narcowich and J.D. Ward. Norm estimates for the inverses of a general class of scattered–data radial–function interpolation matrices. *Journal of Approximation Theory*, 69:84–109, 1992.

[36] I.P. Natanson. *Konstruktive Funktionentheorie.* Akademie–Verlag, Berlin, 1955.

[37] 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.

[38] M.J.D. Powell. Univariate multiquadric approximation: reproduction of linear polynomials. Technical Report DAMTP 1990/NA2, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, 1992.

[39] M.J.D. Powell. Truncated Laurent expansions for the fast evaluation of thin plate splines. *Numer. Algorithms*, 5(1–4):99–120, 1993. Algorithms for Approximation, III (Oxford, 1992).

[40] R. Schaback. Improved error bounds for scattered data interpolation by radial basis functions. Submitted to *Mathematics of Computation*, 1996.

[41] R. Schaback. Error estimates and condition numbers for radial basis function interpolation. *Advances in Computational Mathematics*, 3:251–264, 1995.

[42] R. Schaback. Optimal recovery in translation–invariant spaces of functions. *Annals of Numer. Math.*, 1, 1996.

[43] R. Schaback and Z. Wu. Operators on radial basis functions. To appear in the proceedings of the 1995 Cancun conference.

[44] J. Stewart. Positive definite functions and generalizations, an historical survey. *Rocky Mountain J. Math.*, 6:409–434, 1976.

[45] M. Weinrich. *Charakterisierung von Funktionenräumen bei der Interpolation mit radialen Basisfunktionen.* Dissertation, Universität Göttingen, 1994.

[46] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4:389–396, 1995.

[47] Z. Wu. Multivariate compactly supported positive definite radial functions. *Advances in Computational Mathematics*, 4:2833–292, 1995.

These were the references from the usual radial basis function database kept in Göttingen. The following are preliminary, and will be updated later.

References

[1] (*DNW*) Dyn, Narcowich, Ward, new paper for general recovery on manifolds

[2] (*Light-Wayne:96-1*) W. Light, H. Wayne, recent preprint

[3] (*muehlbach:??-??*) Mühlbach, papers on recursive interpolation, somewhere

[4] (*PowellEffTPSSystem*) Powell, efficiency of tps system enhanced

[5] (*NRSW*) Narcowich, Schaback, Ward, new paper multilevel

Index

This is a preliminary version: no subitems, just some keywords as provided by the \kw macro.

almost everywhere 143
approximation 7
autocorrelation function 148
Banach space 178
barycentric coordinates 111
Beppo-Levi space 49
Bessel function 187
best approximation 179
bounded mapping 179
Cauchy-Schwarz inequality 178
closed subspace 178
complete 178
complete 181
condition number 98
conditionally positive definite function 14
conditionally positive definite function 46
cone 147
continuous mapping 179
convergent sequence 178
convolution 198
dense 178
dual Hilbert space 185
Euclidean norm 7
fill distance 18
finite elements 10
finite interpolation property 44
finite separation property 44
Fourier Inversion Theorem 201
Fourier transform 197
Friedrich's mollifier 195
Gamma function 186
Gaussian 13
Gaussian 199
generalized Fourier transform 204
generalized optimal power function 41
generalized optimal recovery problem 30
generalized transform 145
Gram matrix 179
Hankel transform 149
harmonic functions 50
Hilbert space 181
inner product 178
interior cone condition 118
interpolation 7
inverse multiquadric 13
inverse multiquadrics 169
kernel expansion 97
k-nearest neighbor problem 176
Laplace transform 154
Lipschitz continuity 107
locations 9
monotone operator 155
multiquadric 13
native space 54
native space 55
native space 63
natural cubic spline 79
Newton–type interpolation formula 96
next-neighbour-rule 87
nondegenerate 9
normal equations 179
normalization 63
operator norm 179
optimal linear recovery process 82
optimal recovery problem 29
orthogonal complement 181
orthogonal 179

Plancherel's equation	144
Plancherel's equation	203
polyharmonic functions	50
positive definite function	12
positive definite function	43
power function	80
pre-Hilbert space	177
projection theorem	181
projector	181
query problems	176
radial basis functions	10
recovery problem	7
Riesz representation theorem	185
rotational invariance	10
seminorm	178
separation distance	18
Sobolev inequality	49
Sobolev space	51
sphere	7
Stirling's inequality	113
support	194
Taylor's Formula	176
tempered function	204
tempered radial functions	151
tempered test functions	197
test functions	143
thin-plate spline	13
thin-plate splines	50
transform mapping	143
translation invariance	10
truncated power function	176
unit ball	7
Weber–Schafheitlin integral	158
Weber-Schafheitlin integral	192
zero sequence	178
zonal	10
zonal	53