Chapter 1

Introduction

1.1 History and Outline

Originally, the motivation for the basic meshfree approximation methods (radial basis functions and moving least squares methods) came from applications in geodesy, geophysics, mapping, or meteorology. Later, applications were found in many areas such as in the numerical solution of PDEs, artificial intelligence, learning theory, neural networks, signal processing, sampling theory, statistics (kriging), finance, and optimization. It should be pointed out that (meshfree) local regression methods have been used (independently) in statistics for more than 100 years (see, e.g., [146]).

"Standard" multivariate approximation methods (splines or finite elements) require an underlying mesh (e.g. triangulation) for the definition of basis functions or elements. This is very difficult in space dimensions > 2.

Some historical landmarks for meshfree methods in approximation theory:

- D. Shepard, Shepard functions, late 1960s (application, surface modelling)
- Rolland Hardy (Iowa State Univ.), multiquadrics (MQs), early 1970s (application, geodesy)
- Jean Duchon (Université Joseph Fourier, Grenoble, France), variational approach (in $\mathbb{R}^2$ minimize integral of $\nabla^2 s$), leads to thin plate splines (TPSs), mid 1970s (mathematics)
- Jean Meinguet (Université Catholique de Louvain, Louvain, Belgium), surface splines, late 1970s (mathematics)
- Peter Lancaster and Kes Šalkauskas (Univ. of Calgary, Canada): Surfaces generated by moving least squares methods, 1981, generalizes Shepard functions.
- Richard Franke (NPG, Montery), in 1982 compared scattered data interpolation methods, and concluded MQs and TPs were best. Franke conjectured interpolation matrix for MQs is invertible.
- Wally Madych (Univ. Connecticut), and S. A. Nelson (Iowa State Univ.), Multivariate interpolation: A variational theory, unpublished manuscript, 1983 (proved Franke's conjecture).

- Charles Micchelli (IBM), Interpolation of scattered data: Distance matrices and conditionally positive definite functions, 1986.

Topics to be covered:

- radial basis functions (multiquadrics, thin plate splines, Gaussians
- moving least squares methods (element-free Galerkin (EFG), *hp*-clouds, meshless local Petrov-Galerkin (MLPG), radial point interpolation method (RPIM), reproducing kernel particle method (RKPM), smooth particle hydrodynamics (SPH))
- partition of unity methods
- quasi-interpolation methods
- dual reciprocity method (DRM)

Applications discussed:

- scattered data fitting
- solution of PDEs (collocation, Galerkin; elliptic, parabolic, hyperbolic)
- surface reconstruction
- machine learning
- optimization

1.2 Motivation: Scattered Data Interpolation

In this section we will describe the general process of scattered data fitting, which is one of the fundamental problems in approximation theory and data modelling in general. Our desire to have a well-posed problem formulation will naturally lead to the concepts of positive definite matrices, and strictly positive definite functions.

1.2.1 Scattered Data Interpolation.

In many scientific disciplines one faces the following problem. We have a set of data (measurements, and locations at which these measurements were obtained), and we want to find a rule which allows us to deduce information about the process we are studying also at locations different from those at which we obtained our measurements. Thus, we are trying to find a function s which is a "good" fit to the given data. There are many ways to decide what we mean by "good", and the only criterion we will consider now is that we want the function s to exactly match the given measurements at the corresponding locations. This approach is called *interpolation*, and if the locations at which the measurements are taken are not on a uniform or regular grid, then the process is called *scattered data interpolation*. More precisely, we are considering the following

Problem 1.2.1 *Given data* $(\boldsymbol{x}_j, y_j)$, $j = 1, \ldots, N$ *with* $\boldsymbol{x}_j \in \mathbb{R}^s$, $y_j \in \mathbb{R}$ *find a (continuous) function* $\mathcal{P}f$ *such that* $\mathcal{P}f(\boldsymbol{x}_j) = y_j$, $j = 1, \ldots, N$.

Here the $\boldsymbol{x}_j$ are the measurement locations (or *data sites*), and the y_j are the corresponding measurements (or *data values*). We will often assume that these values are obtained by sampling a data function f at the data sites, i.e., $y_j = f(\boldsymbol{x}_j)$, $j = 1, \ldots, N$. The fact that we allow $\boldsymbol{x}_j$ to lie in s-dimensional space $\mathbb{R}^s$ means that the formulation of Problem 1.2.1 allows us to cover many different types of problems. If $s = 1$ the data could be a series of measurements taken over a certain time period, thus the "data sites" $\boldsymbol{x}_j$ would correspond to certain time instances. For $s = 2$ we can think of the data being obtained over a planar region, and so $\boldsymbol{x}_j$ corresponds to the two coordinates in the plane. For instance, we might want to produce a map which shows the rainfall in the state we live in based on the data collected at weather station located throughout the state. For $s = 3$ we might think of a similar situation in space. One possibility is that we could be interested in the temperature distribution inside some solid body. Higher-dimensional examples might not be that intuitive, but a multitude of them exist, e.g., in finance, economics or statistics, but also in artificial intelligence or learning.

A convenient and common approach to solving the scattered data problem is to make the assumption that the function $\mathcal{P}f$ is a linear combination of certain *basis functions* B_k, i.e.,

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{k=1}^{N} c_k B_k(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s\,. \tag{1.1}$$

Solving the interpolation problem under this assumption leads to a system of linear equations of the form

$$A\boldsymbol{c} = \boldsymbol{y},$$

where the entries of the *interpolation matrix* A are given by $A_{jk} = B_k(\boldsymbol{x}_j)$, $j, k = 1, \ldots, N$, $\boldsymbol{c} = [c_1, \ldots, c_N]^T$, and $\boldsymbol{y} = [y_1, \ldots, y_N]^T$.

Problem 1.2.1 will be well-posed, i.e., a solution to the problem will exist and be unique, if and only if the matrix A is non-singular.

In the univariate setting it is well known that one can interpolate to arbitrary data at N distinct data sites using a polynomial of degree $N - 1$. For the multivariate setting, however, there is the following negative result due to Mairhuber and Curtis in 1956 [425].

Theorem 1.2.2 *If* $\Omega \subset \mathbb{R}^s$, $s \geq 2$, *contains an interior point, then there exist no Haar spaces of continuous functions except for one-dimensional ones.*

In order to understand this theorem we need

Definition 1.2.3 *Let the linear finite-dimensional function space* $\mathcal{B} \subseteq C(\Omega)$ *have a basis* $\{B_1, \ldots, B_N\}$. *Then* $\mathcal{B}$ *is a* Haar space *on* Ω *if*

$$\det\left(B_k(\boldsymbol{x}_j)\right) \neq 0$$

for any set of distinct $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ *in* Ω.

Remarks:

1. Note that existence of a Haar space guarantees invertibility of the interpolation matrix $(B_k(\boldsymbol{x}_j))$, i.e., existence and uniqueness of an interpolant to data specified at $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, from the space $\mathcal{B}$.

2. As mentioned above, univariate polynomials of degree $N-1$ form an N-dimensional Haar space for data given at $x_1, \ldots, x_N$.

3. The Mairhuber-Curtis Theorem implies that in the multivariate setting we can no longer expect this to be the case. E.g., it is not possible to perform unique interpolation with (multivariate) polynomials of degree N to data given at arbitrary locations in $\mathbb{R}^2$.

4. The Mairhuber-Curtis Theorem tells us that if we want to have a well-posed multivariate scattered data interpolation problem, then the basis needs to depend on the data locations.

Proof of Theorem 1.2.2: Let $s \geq 2$ and suppose $\mathcal{B}$ is a Haar space with basis $\{B_1, \ldots, B_N\}$ with $N \geq 2$. Then, by the definition of a Haar space

$$\det\left(B_k(\boldsymbol{x}_j)\right) \neq 0 \tag{1.2}$$

for any distinct $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$.

Now consider a closed path P in Ω connecting only $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. This is possible since – by assumption – Ω contains an interior point. We can exchange the positions of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ by moving them continuously along the path P (without interfering with any of the other $\boldsymbol{x}_j$). This means, however, that rows 1 and 2 of the determinant (1.2) have been exchanged, and so the determinant has changed sign.

Since the determinant is a continuous function of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ we must have had $\det = 0$ at some point along P. This is a contradiction. □

In order to obtain such data dependent approximation spaces we now consider positive definite matrices and functions.

1.2.2 Positive Definite Matrices and Functions

A common concept in linear algebra is that of a *positive definite matrix*.

Definition 1.2.4 *A real symmetric matrix A is called* positive semi-definite *if its associated quadratic form is non-negative, i.e.,*

$$\sum_{j=1}^{N}\sum_{k=1}^{N} c_j c_k A_{jk} \geq 0 \tag{1.3}$$

for $\boldsymbol{c} = [c_1, \ldots, c_N]^T \in \mathbb{R}^N$. If the only vector $\boldsymbol{c}$ that turns (1.3) into an equality is the zero vector, then A is called positive definite.

An important property of positive definite matrices is that all their eigenvalues are positive, and therefore a positive definite matrix is non-singular (but certainly not vice versa).

If we therefore had basis functions B_k in the expansion (1.1) above which generate a positive definite interpolation matrix, we would always have a well-posed interpolation problem. To this end we introduce the concept of a *positive definite function* from classical analysis.

Historically, in the 1920s and 30s, only positive definite functions were introduced. However, in order to meet our goal of having a well-posed interpolation problem it is necessary to sharpen the classical notion of a positive definite function to that of a **strictly** positive definite one. This leads to an unfortunate difference in terminology used in the context of matrices and functions. Unfortunately, in the course of history it has turned out that a positive definite function is associated with a positive **semi**-definite matrix.

Definition 1.2.5 *A complex-valued continuous function* Φ *is called* positive definite on $\mathbb{R}^s$ *if*

$$\sum_{j=1}^{N}\sum_{k=1}^{N} c_j\overline{c_k}\Phi(\boldsymbol{x}_j-\boldsymbol{x}_k)\geq 0 \tag{1.4}$$

for any N *pairwise different points* $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\in\mathbb{R}^s$, *and* $\boldsymbol{c}=[c_1,\ldots,c_N]^T\in\mathbb{C}^N$. *The function* Φ *is called* strictly positive definite on $\mathbb{R}^s$ *if the only vector* $\boldsymbol{c}$ *that turns (1.4) into an equality is the zero vector.*

We note that an extension of the notion of positive definiteness to cover complex coefficients $\boldsymbol{c}$ and complex-valued functions Φ as done in Definition 1.2.5 will be helpful when deriving some properties of (strictly) positive definite functions later on. Moreover, the celebrated *Bochner's Theorem* (see the next chapter) characterizes exactly the positive definite functions of Definition 1.2.5. In all practical circumstances, however, we will be concerned with real-valued functions only, and a characterization of such functions appears below as Theorem 1.2.7.

Definition 1.2.5 and the discussion preceding it suggest that we should use strictly positive definite functions as basis functions in (1.1), i.e., $B_k(\boldsymbol{x})=\Phi(\boldsymbol{x}-\boldsymbol{x}_k)$, or

$$\mathcal{P}f(\boldsymbol{x})=\sum_{k=1}^{N} c_k\Phi(\boldsymbol{x}-\boldsymbol{x}_k),\qquad \boldsymbol{x}\in\mathbb{R}^s\,. \tag{1.5}$$

Remarks:

1. The function $\mathcal{P}f$ of (1.5) will yield an interpolant that is *translation invariant*, i.e., the interpolant to translated data is the same as the translated interpolant to the original data.

2. Definition 1.2.5 can be generalized to the notion of strictly positive definite kernels of the form $\Phi(\boldsymbol{x},\boldsymbol{y})$.

3. Positive definite functions were first considered in classical analysis early in the 20th century. In the 1920s Mathias [432] seems to have been the first to define and study positive definite functions. An overview of the development up to the mid 1970s can be found in [590]. There seems to have been no need to study strictly positive functions until Micchelli [456] made the connection between scattered data interpolation and positive definite functions. We will discuss some of the most important properties and characterizations of (strictly) positive definite functions in the next chapter.

4. We would like to point out that when reading recent articles (especially in the radial basis function literature) dealing with (strictly) positive definite functions one has to be aware of the fact that some authors have tried to "correct" history, and now refer to strictly positive definite functions as positive definite functions.

We close this section with a list of some basic properties of (strictly) positive definite functions and some examples.

Theorem 1.2.6 *Some basic properties of positive definite functions are*

(1) If $\Phi_1, \ldots, \Phi_n$ are positive definite on $\mathbb{R}^s$ and $c_i \geq 0$, $i = 1, \ldots, n$, then

$$\Phi(\boldsymbol{x}) = \sum_{i=1}^{n} c_i \Phi_i(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s,$$

is also positive definite. Moreover, if one of the Φ_i is strictly positive definite and the corresponding $c_i > 0$, then Φ is strictly positive definite.

(2) $\Phi(-\boldsymbol{x}) = \overline{\Phi(\boldsymbol{x})}$.

(3) $\Phi(\mathbf{0}) \geq 0$.

(4) Any positive definite function is bounded, in fact,

$$|\Phi(\boldsymbol{x})| \leq \Phi(\mathbf{0}).$$

(5) If Φ is positive definite with $\Phi(\mathbf{0}) = 0$ then $\Phi \equiv 0$.

(6) The product of (strictly) positive definite functions is (strictly) positive definite.

Proof: Properties (1) and (3) follow immediately from Definition 1.2.5.

To show (2) we let $N = 2$, $\boldsymbol{x}_1 = \mathbf{0}$, $\boldsymbol{x}_2 = \boldsymbol{x}$, and choose $c_1 = 1$ and $c_2 = c$. Then the quadratic form in Definition 1.2.5 becomes

$$\sum_{j=1}^{2} \sum_{k=1}^{2} c_j \overline{c_k} \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) = (1 + |c|^2)\Phi(\mathbf{0}) + c\Phi(\boldsymbol{x}) + \overline{c}\Phi(-\boldsymbol{x}) \geq 0$$

for every $c \in \mathbb{C}$. Taking $c = 1$ and $c = i$ (where $i = \sqrt{-1}$), respectively, we can see that both $\Phi(\boldsymbol{x}) + \Phi(-\boldsymbol{x})$ and $i\,(\Phi(\boldsymbol{x}) - \Phi(-\boldsymbol{x}))$ must be real. This, however, is only possible if $\Phi(-\boldsymbol{x}) = \overline{\Phi(\boldsymbol{x})}$.

For the proof of (4) we let $N = 2$, $\boldsymbol{x}_1 = \mathbf{0}$, $\boldsymbol{x}_2 = \boldsymbol{x}$, and choose $c_1 = |\Phi(\boldsymbol{x})|$ and $c_2 = -\overline{\Phi(\boldsymbol{x})}$. Then the quadratic form in Definition 1.2.5 is

$$\sum_{j=1}^{2}\sum_{k=1}^{2} c_j\overline{c_k}\Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) = 2\Phi(\mathbf{0})|\Phi(\boldsymbol{x})|^2 - \Phi(-\boldsymbol{x})\Phi(\boldsymbol{x})|\Phi(\boldsymbol{x})| - \Phi^2(\boldsymbol{x})|\Phi(\boldsymbol{x})| \geq 0.$$

Since $\Phi(-\boldsymbol{x}) = \overline{\Phi(\boldsymbol{x})}$ by Property 2, this gives

$$2\Phi(\mathbf{0})|\Phi(\boldsymbol{x})|^2 - 2|\Phi(\boldsymbol{x})|^3 \geq 0.$$

If $|\Phi(\boldsymbol{x})| > 0$, we divide by $|\Phi(\boldsymbol{x})|^2$ and the statement follows immediately. In case $|\Phi(\boldsymbol{x})| \equiv 0$ the statement holds trivially.

Property (5) follows immediately from (4), and Property (6) is a consequence of a theorem by Schur, which states that the elementwise (or Hadamard) product of positive (semi-)definite matrices is positive (semi-)definite (see [132] or [634] for details). □

Remarks:

1. Property (1) states that in particular the sum of two (strictly) positive definite functions is (strictly) positive definite.

2. Property (2) shows that any real-valued (strictly) positive definite function has to be even. However, it is also possible to characterize real-valued (strictly) positive definite functions using only *real* coefficients (see [634] for details), i.e.,

Theorem 1.2.7 *A real-valued continuous function Φ is positive definite on $\mathbb{R}^s$ if and only if it is even and*

$$\sum_{j=1}^{N}\sum_{k=1}^{N} c_j c_k \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) \geq 0 \tag{1.6}$$

for any N pairwise different points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^s$, and $\boldsymbol{c} = [c_1, \ldots, c_N]^T \in \mathbb{R}^N$. The function Φ is strictly positive definite on $\mathbb{R}^s$ if the only vector $\boldsymbol{c}$ that turns (1.6) into an equality is the zero vector.

Examples:

1. The function $\Phi(\boldsymbol{x}) = e^{i\boldsymbol{x}\cdot\boldsymbol{y}}$, $\boldsymbol{y} \in \mathbb{R}^s$, is positive definite on $\mathbb{R}^s$ since the quadratic form in Definition 1.2.5 becomes

$$\begin{aligned}\sum_{j=1}^{N}\sum_{k=1}^{N} c_j\overline{c_k}\Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) &= \sum_{j=1}^{N}\sum_{k=1}^{N} c_j\overline{c_k}e^{i(\boldsymbol{x}_j - \boldsymbol{x}_k)\cdot\boldsymbol{y}} \\ &= \sum_{j=1}^{N} c_j e^{i\boldsymbol{x}_j\cdot\boldsymbol{y}} \sum_{k=1}^{N} \overline{c_k} e^{-i\boldsymbol{x}_k\cdot\boldsymbol{y}} \\ &= \left|\sum_{j=1}^{N} c_j e^{i\boldsymbol{x}_j\cdot\boldsymbol{y}}\right|^2 \geq 0.\end{aligned}$$

2. The cosine function is positive definite on $\mathbb{R}$ since, for $x \in \mathbb{R}$, we have $\cos x = \frac{1}{2}\left(e^{ix} + e^{-ix}\right)$, and Property (1) and the previous example can be invoked.

1.2.3 Radial Functions

Of particular interest in applications are positive definite functions which are also radial. Radial functions have the nice property that they are invariant under all Euclidean transformations (i.e., translations, rotations, and reflections). This is an immediate consequence of the fact that Euclidean transformations are characterized by orthogonal transformation matrices and are therefore norm-invariant. Invariance under translation, rotation and reflection is often desirable in applications. We therefore define

Definition 1.2.8 *A function* $\Phi : \mathbb{R}^s \to \mathbb{R}$ *is called* radial *provided there exists a* univariate *function* $\varphi : [0, \infty) \to \mathbb{R}$ *such that*

$$\Phi(\boldsymbol{x}) = \varphi(r), \quad \textit{where} \quad r = \|\boldsymbol{x}\|,$$

and $\|\cdot\|$ *is some norm on* $\mathbb{R}^s$ *– usually the Euclidean norm.*

Definition 1.2.8 says that for a radial function Φ

$$\|\boldsymbol{x}_1\| = \|\boldsymbol{x}_2\| \quad \Longrightarrow \quad \Phi(\boldsymbol{x}_1) = \Phi(\boldsymbol{x}_2), \qquad \boldsymbol{x}_1,\ \boldsymbol{x}_2 \in \mathbb{R}^d\,.$$

However, what makes radial functions most useful for applications is the fact that the interpolation problem becomes insensitive to the dimension s of the space in which the data sites lie. Instead of having to deal with a multivariate function Φ (whose complexity will increase with increasing space dimension s) we can work with the same univariate function φ for all choices of s.

We call the univariate function φ a *(strictly) positive definite radial function on* $\mathbb{R}^s$ if and only if the associated multivariate function Φ is (strictly) positive definite on $\mathbb{R}^s$ in the sense of Definition 1.2.5 and radial in the sense of Definition 1.2.8.

Chapter 2

Positive Definite and Completely Monotone Functions

Below we will first summarize facts about positive definite functions, and closely related completely and multiply monotone functions. Most of these facts are integral characterizations and were established in the 1930s by Bochner and Schoenberg. In the second part of this chapter we will mention the more recent extensions to strictly positive definite and strictly completely/multiply monotone functions. Integral characterizations are an essential ingredient in the theoretical analysis of radial basis functions.

2.1 A Brief Summary of Integral Transforms

Before we get into the details of the integral representations we summarize some formulas for various integral transforms to be used later.

The Fourier transform conventions we will adhere to are laid out in

Definition 2.1.1 *The* Fourier transform *of* $f \in L_1(\mathbb{R}^s)$ *is given by*

$$\hat{f}(\boldsymbol{\omega}) = \frac{1}{\sqrt{(2\pi)^s}} \int_{\mathbb{R}^s} f(\boldsymbol{x}) e^{-i\boldsymbol{\omega}\cdot\boldsymbol{x}} d\boldsymbol{x}, \qquad \boldsymbol{\omega} \in \mathbb{R}^s, \tag{2.1}$$

and its inverse Fourier transform *is given by*

$$\check{f}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^s}} \int_{\mathbb{R}^s} f(\boldsymbol{\omega}) e^{i\boldsymbol{x}\cdot\boldsymbol{\omega}} d\boldsymbol{\omega}, \qquad \boldsymbol{x} \in \mathbb{R}^s\,.$$

Remark: This definition of the Fourier transform can be found in Rudin [537]. Another, just as common, definition uses

$$\hat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}^s} f(\boldsymbol{x}) e^{-2\pi i\boldsymbol{\omega}\cdot\boldsymbol{x}} d\boldsymbol{x}, \tag{2.2}$$

and can be found in Stein and Weiss [589]. The form we use can also be found in Wendland's book [634], whereas (2.2) is used in the book by Cheney and Light [132].

Similarly, we can define the Fourier transform of a finite (signed) measure μ on $\mathbb{R}^s$ by

$$\hat{\mu}(\boldsymbol{\omega}) = \frac{1}{\sqrt{(2\pi)^s}} \int_{\mathbb{R}^s} e^{-i\boldsymbol{\omega}\cdot\boldsymbol{x}} d\mu(\boldsymbol{x}), \qquad \boldsymbol{\omega} \in \mathbb{R}^s .$$

Since we will be interested in positive definite radial functions, we note that the Fourier transform of a radial function is again radial. Indeed,

Theorem 2.1.2 *Let $\Phi \in L_1(\mathbb{R}^s)$ be continuous and radial, i.e., $\Phi(\boldsymbol{x}) = \varphi(\|\boldsymbol{x}\|)$. Then its Fourier transform $\hat{\Phi}$ is also radial, i.e., $\hat{\Phi}(\boldsymbol{\omega}) = \mathcal{F}_s\varphi(\|\boldsymbol{\omega}\|)$ with*

$$\mathcal{F}_s\varphi(r) = \frac{1}{\sqrt{r^{s-2}}} \int_0^\infty \varphi(t) t^{\frac{s}{2}} J_{(s-2)/2}(rt) dt,$$

where $J_{(s-2)/2}$ is the classical Bessel function of the first kind of order $(s-2)/2$.

Remark: The integral transform appearing in Theorem 2.1.2 is also referred to as a *B*essel transform.

A third integral transform to play an important role in the following is the *Laplace transform*. We have

Definition 2.1.3 *The* Laplace transform *of a piecewise continuous function f that satisfies $|f(t)| \leq Me^{at}$ for some constants a and M is given by*

$$\mathcal{L}f(s) = \int_0^\infty f(t) e^{-st} dt, \qquad s > a.$$

Similarly, the Laplace transform of a Borel measure μ on $[0, \infty)$ is given by

$$\mathcal{L}\mu(s) = \int_0^\infty e^{-st} d\mu(t).$$

The Laplace transform is continuous at the origin if and only if μ is finite.

2.2 Bochner's Theorem

One of the most celebrated results on positive definite functions is their characterization in terms of Fourier transforms established by Bochner in 1932 (for $s = 1$) and 1933 (for general s).

Theorem 2.2.1 *(Bochner's Theorem) A (complex-valued) function $\Phi \in C(\mathbb{R}^s)$ is positive definite on $\mathbb{R}^s$ if and only if it is the Fourier transform of a finite non-negative Borel measure μ on $\mathbb{R}^s$, i.e.,*

$$\Phi(\boldsymbol{x}) = \hat{\mu}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^s}} \int_{\mathbb{R}^s} e^{-i\boldsymbol{x}\cdot\boldsymbol{y}} d\mu(\boldsymbol{y}), \qquad \boldsymbol{x} \in \mathbb{R}^s .$$

Proof: There are many proofs of this theorem. Bochner's original proof can be found in [58], p. 407. Other proofs can be found, e.g., in the books by Cuppens ([147], p. 41) or Gelfand and Vilenkin ([250], p. 155). A nice proof using the Riesz Representation Theorem to interpret the Borel measure as a distribution, and then taking advantage of distributional Fourier transforms can be found in the book by Wendland [634].

We will prove only the one (easy) direction which is important for the application to scattered data interpolation. We assume Φ is the Fourier transform of a finite non-negative Borel measure and show Φ is positive definite. Thus,

$$\begin{aligned}
\sum_{j=1}^{N}\sum_{k=1}^{N} c_j\overline{c_k}\Phi(\boldsymbol{x}_j-\boldsymbol{x}_k) &= \frac{1}{\sqrt{(2\pi)^s}}\sum_{j=1}^{N}\sum_{k=1}^{N}\left[c_j\overline{c_k}\int_{\mathbb{R}^s} e^{-i(\boldsymbol{x}_j-\boldsymbol{x}_k)\cdot\boldsymbol{y}}d\mu(\boldsymbol{y})\right] \\
&= \frac{1}{\sqrt{(2\pi)^s}}\int_{\mathbb{R}^s}\left[\sum_{j=1}^{N} c_j e^{-i\boldsymbol{x}_j\cdot\boldsymbol{y}}\sum_{k=1}^{N}\overline{c_k}e^{i\boldsymbol{x}_k\cdot\boldsymbol{y}}\right]d\mu(\boldsymbol{y}) \\
&= \frac{1}{\sqrt{(2\pi)^s}}\int_{\mathbb{R}^s}\left|\sum_{j=1}^{N} c_j e^{-i\boldsymbol{x}_j\cdot\boldsymbol{y}}\right|^2 d\mu(\boldsymbol{y}) \geq 0.
\end{aligned}$$

The last inequality holds because of the conditions imposed on the measure μ. □

2.3 Strictly Positive Definite Functions

In order to accomplish our goal of guaranteeing a well-posed interpolation problem, we have to extend (if possible) Bochner's characterization to *strictly* positive definite functions.

We begin with a sufficient condition for a function to be strictly positive definite on $\mathbb{R}^s$.

For this result we require the notion of the *carrier* of a (non-negative) Borel measure defined on some topological space X. This set is given by

$$X \setminus \bigcup\{O : \ O \text{ is open and } \mu(O) = 0\}.$$

Theorem 2.3.1 *Let μ be a non-negative finite Borel measure on $\mathbb{R}^s$ whose carrier is not a set of Lebesgue measure zero. Then the Fourier transform of μ is strictly positive definite on $\mathbb{R}^s$.*

Proof: As in the proof of Bochner's Theorem we have

$$\begin{aligned}
\sum_{j=1}^{N}\sum_{k=1}^{N} c_j\overline{c_k}\hat{\mu}(\boldsymbol{x}_j-\boldsymbol{x}_k) &= \frac{1}{\sqrt{(2\pi)^s}}\sum_{j=1}^{N}\sum_{k=1}^{N} c_j\overline{c_k}\left[\int_{\mathbb{R}^s} e^{-i(\boldsymbol{x}_j-\boldsymbol{x}_k)\cdot\boldsymbol{y}}d\mu(\boldsymbol{y})\right] \\
&= \frac{1}{\sqrt{(2\pi)^s}}\int_{\mathbb{R}^s}\left[\sum_{j=1}^{N} c_j e^{-i\boldsymbol{x}_j\cdot\boldsymbol{y}}\sum_{k=1}^{N}\overline{c_k}e^{i\boldsymbol{x}_k\cdot\boldsymbol{y}}\right]d\mu(\boldsymbol{y})
\end{aligned}$$

$$= \frac{1}{\sqrt{(2\pi)^s}} \int_{\mathbb{R}^s} \left| \sum_{j=1}^{N} c_j e^{-i\boldsymbol{x}_j \cdot \boldsymbol{y}} \right|^2 d\mu(\boldsymbol{y}) \geq 0.$$

Now let

$$g(\boldsymbol{y}) = \sum_{j=1}^{N} c_j e^{-i\boldsymbol{x}_j \cdot \boldsymbol{y}},$$

and assume that the points $\boldsymbol{x}_j$ are all distinct and $\boldsymbol{c} \neq \boldsymbol{0}$. In this case the functions $\boldsymbol{y} \mapsto e^{-i\boldsymbol{x}_j \cdot \boldsymbol{y}}$ are linearly independent, and thus the zero set of g, i.e., $\{\boldsymbol{y} \in \mathbb{R}^s : \ g(\boldsymbol{y}) = 0\}$ has Lebesgue measure zero. Therefore, the only remaining way to make the above inequality an equality is if the carrier of μ is contained in the zero set of g, i.e., has Lebesgue measure zero. □

The following corollary gives us a way to *construct* strictly positive definite functions.

Corollary 2.3.2 *Let f be a continuous non-negative function in $L_1(\mathbb{R}^s)$ which is not identically zero. Then the Fourier transform of f is strictly positive definite on $\mathbb{R}^s$.*

Proof: We use the measure μ defined for any Borel set B by

$$\mu(B) = \int_B f(\boldsymbol{x}) d\boldsymbol{x}.$$

Then the carrier of μ is equal to the closed support of f. However, since f is non-negative and not identically equal to zero, its support has positive Lebesgue measure, and hence the Fourier transform of f is strictly positive definite by the preceding theorem. □

Remark: Work toward an analog of Bochner's Theorem, i.e., an integral characterization for functions which are strictly positive definite on $\mathbb{R}^s$, is given in [112] for $s = 1$.

Example: The *Gaussian*

$$\Phi(\boldsymbol{x}) = e^{-\alpha \|\boldsymbol{x}\|^2}, \quad \alpha > 0, \tag{2.3}$$

is strictly positive definite on $\mathbb{R}^s$ for any s. This is essentially due to the fact that the Fourier transform of a Gaussian is again a Gaussian. In particular, for $\alpha = \frac{1}{2}$ we have $\hat{\Phi} = \Phi$ which can be verified by direct calculation. The general statement follows from the properties of the Fourier transform (complete details are given in the book by Wendland on pp. 50 and 69). An easier argument (using completely monotone functions) will become available later.

Remark: Since Gaussians play a central role in statistics, this is a good place to mention that positive definite functions are – up to a normalization $\Phi(0) = 1$ – identical with characteristic functions of distribution functions in statistics.

Finally, a criterion to check whether a given function is strictly positive definite is given in [634].

Theorem 2.3.3 *Let Φ be a continuous function in $L_1(\mathbb{R}^s)$. Φ is strictly positive definite if and only if Φ is bounded and its Fourier transform is non-negative and not identically equal to zero.*

Remark: The proof of Theorem 2.3.3 shows that – if $\Phi \not\equiv 0$ (which implies that $\hat{\Phi} \not\equiv 0$) – we need to ensure only that $\hat{\Phi}$ be non-negative in order for Φ to be strictly positive definite.

Example: Theorem 2.3.3 can be used to show that the so-called *inverse multiquadrics*

$$\Phi(\boldsymbol{x}) = \left(\|\boldsymbol{x}\|^2 + \alpha^2\right)^{-\beta}, \quad \alpha > 0,\ \beta > \frac{s}{2}, \tag{2.4}$$

are strictly positive definite on $\mathbb{R}^s$ (complete details are given in [634]). By using another argument based on completely monotone functions we will be able to show that in fact we need to require only $\beta > 0$, and therefore the inverse multiquadrics are strictly positive definite on any $\mathbb{R}^s$.

2.4 Positive Definite Radial Functions

We now turn our attention to positive definite radial functions. Theorem 2.1.2 can be used to prove the following characterization due to Schoenberg (see [569], p.816).

Theorem 2.4.1 *A continuous function $\varphi : [0, \infty) \to \mathbb{R}$ is positive definite and radial on $\mathbb{R}^s$ if and only if it is the Bessel transform of a finite non-negative Borel measure μ on $[0, \infty)$, i.e.,*

$$\varphi(r) = \int_0^\infty \Omega_s(rt) d\mu(t),$$

where

$$\Omega_s(r) = \begin{cases} \cos r & \text{for } s = 1, \\ \Gamma\left(\frac{s}{2}\right)\left(\frac{2}{r}\right)^{(s-2)/2} J_{(s-2)/2}(r) & \text{for } s \geq 2, \end{cases}$$

and $J_{(s-2)/2}$ is the classical Bessel function of the first kind of order $(s-2)/2$.

Since any function which is positive definite and radial on $\mathbb{R}^{s_1}$ is also positive definite and radial on $\mathbb{R}^{s_2}$ as long as $s_2 \leq s_1$, those functions which are positive definite and radial on $\mathbb{R}^s$ *for all* s are of particular interest. This latter class of functions was also characterized by Schoenberg ([569], pp. 817–821.). We saw above that the Gaussians and inverse multiquadrics provide examples of such functions.

Theorem 2.4.2 *A continuous function $\varphi : [0, \infty) \to \mathbb{R}$ is positive definite and radial on $\mathbb{R}^s$ for all s if and only if it is of the form*

$$\varphi(r) = \int_0^\infty e^{-r^2 t^2} d\mu(t),$$

where μ is a finite non-negative Borel measure on $[0, \infty)$.

Remark: Schoenberg referred to the functions which are positive definite and radial on $\mathbb{R}^s$ for all s as positive definite radial functions on ℓ_2.

We end this section with examples of functions that are strictly positive definite and radial on $\mathbb{R}^s$ with restrictions on the space dimension s. Moreover, the following functions differ from the previous ones in that they have *compact support*.

Examples:

1. The *truncated power function*

$$\varphi_\ell(r) = (1-r)_+^\ell \tag{2.5}$$

is strictly positive definite and radial on $\mathbb{R}^s$ provided ℓ satisfies $\ell \geq \lfloor \frac{s}{2} \rfloor + 1$. For details see [634]. Here we have used the cutoff function $(\cdot)_+$ which is defined by

$$(x)_+ = \begin{cases} x, & \text{for } x \geq 0, \\ 0, & \text{for } x < 0. \end{cases}$$

2. Let $f \in C[0,\infty)$ be non-negative and not identically equal to zero, and define the function φ by

$$\varphi(r) = \int_0^\infty (1-rt)_+^{k-1} f(t)dt. \tag{2.6}$$

Then φ is strictly positive definite and radial on $\mathbb{R}^s$ provided $k \geq \lfloor \frac{s}{2} \rfloor + 2$. This can been verified by considering the quadratic form

$$\sum_{j=1}^N \sum_{k=1}^N c_j c_k \varphi(\|\boldsymbol{x}_j - \boldsymbol{x}_k\|) = \int_0^\infty \sum_{j=1}^N \sum_{k=1}^N c_j c_k \varphi_{k-1}(t\|\boldsymbol{x}_j - \boldsymbol{x}_k\|) f(t)dt$$

which is non-negative since φ_{k-1} is strictly positive definite by the first example, and f is non-negative. Since f is also assumed to be not identically equal to zero, the only way for the quadratic form to equal zero is if $\boldsymbol{c} = \boldsymbol{0}$.
Note that (2.6) amounts to another integral transform of f with the compactly supported truncated power function as integration kernel. We will take another look at these functions in the context of multiply monotone functions below.

The Schoenberg characterization of positive definite radial functions on $\mathbb{R}^s$ for all s implies that we have a finite non-negative Borel measure μ on $[0,\infty)$ such that

$$\varphi(r) = \int_0^\infty e^{-r^2t^2} d\mu(t).$$

If we want to find a zero r_0 of φ then we have

$$\varphi(r_0) = \int_0^\infty e^{-r_0^2t^2} d\mu(t) = 0.$$

Since the exponential function is positive and the measure is non-negative, it follows that μ must be the zero measure. However, then ϕ is identically equal to zero. Therefore, a non-trivial function φ that is positive definite and radial on $\mathbb{R}^s$ for all s can have no zeros. This implies in particular that *there are no compactly supported univariate continuous functions that are positive definite and radial on* $\mathbb{R}^s$ *for all* s.

2.5 Completely Monotone Functions

We now introduce a class of functions which is very closely related to positive definite radial functions and leads to a simple characterization of such functions.

Definition 2.5.1 *A function* $\varphi : [0,\infty) \to \mathbb{R}$ *which is in* $C[0,\infty) \cap C^\infty(0,\infty)$ *and which satisfies*

$$(-1)^\ell \varphi^{(\ell)}(r) \geq 0, \qquad r > 0,\ \ell = 0, 1, 2, \ldots,$$

is called completely monotone on $[0,\infty)$.

Example: Some examples of completely monotone functions are

1. $\varphi(r) = \alpha, \qquad \alpha \geq 0,$
2. $\varphi(r) = e^{-\alpha r}, \qquad \alpha \geq 0,$
3. $\varphi(r) = \dfrac{\alpha}{r^{1-\alpha}}, \quad \alpha \leq 1,$
4. $\varphi(r) = \dfrac{1}{(r+\alpha^2)^\beta}, \qquad \alpha > 0,\ \beta \geq 0.$

The following theorem gives an integral characterization of completely monotone functions.

Theorem 2.5.2 *(Hausdorff-Bernstein-Widder Theorem) A function* $\varphi : [0,\infty) \to \mathbb{R}$ *is completely monotone on* $[0,\infty)$ *if and only if it is the Laplace transform of a finite non-negative Borel measure* μ *on* $[0,\infty)$*, i.e.,* φ *is of the form*

$$\varphi(r) = \mathcal{L}\mu(r) = \int_0^\infty e^{-rt} d\mu(t).$$

Remark: Widder's proof of this theorem can be found in [644], p. 160, where he reduces the proof of this theorem to another theorem by Hausdorff on completely monotone sequences. A detailed proof can also be found in the books by Cheney and Light [132] and Wendland [634]. □

Remark: Some properties of completely monotone functions are:

1. A non-negative finite linear combination of completely monotone functions is completely monotone.
2. The product of two completely monotone functions is completely monotone.

The following connection between positive definite radial and completely monotone functions was first pointed out by Schoenberg in 1938.

Theorem 2.5.3 *A function* φ *is completely monotone on* $[0,\infty)$ *if and only if* $\Phi = \varphi(\|\cdot\|^2)$ *is positive definite and radial on* $\mathbb{R}^s$ *for all* s.

Remark: Note that the function Φ is now defined via the *square* of the norm. This is different from our earlier definition of radial functions (see Definition 1.2.8).

Proof: One possibility is to use a change of variables to combine Schoenberg's characterization of functions that are positive definite and radial on any $\mathbb{R}^s$, Theorem 2.4.2, with the Hausdorff-Bernstein-Widder characterization of completely monotone functions. To get more insight we present an alternative proof of the claim that the completely monotone function φ gives rise to a Φ that is positive definite and radial on any $\mathbb{R}^s$. Details for the other direction can be found, e.g., in [634].

The Hausdorff-Bernstein-Widder Theorem implies that we can write φ as

$$\varphi(r) = \int_0^\infty e^{-rt} d\mu(t)$$

with a finite non-negative Borel measure μ. Therefore, $\Phi(\boldsymbol{x}) = \varphi(\|\boldsymbol{x}\|^2)$ has the representation

$$\Phi(\boldsymbol{x}) = \int_0^\infty e^{-\|\boldsymbol{x}\|^2 t} d\mu(t).$$

To see that this function is positive definite on any $\mathbb{R}^s$ we consider the quadratic form

$$\sum_{j=1}^N \sum_{k=1}^N c_j c_k \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) = \int_0^\infty \sum_{j=1}^N \sum_{k=1}^N c_j c_k e^{-t\|\boldsymbol{x}_j - \boldsymbol{x}_k\|^2} d\mu(t).$$

Since we saw earlier that the Gaussians are strictly positive definite and radial on any $\mathbb{R}^s$ it follows that the quadratic form is non-negative. $\square$.

We can see from the previous proof that if the measure μ is not concentrated in the origin, then Φ is even strictly positive definite and radial on any $\mathbb{R}^s$. This condition on the measure is equivalent with ϕ not being constant. With this additional restriction on φ we can apply the notion of a completely monotone function to the scattered data interpolation problem. The following *interpolation theorem* was already proved by Schoenberg in 1938 ([569], p. 823).

Theorem 2.5.4 *If the function $\varphi : [0, \infty) \to \mathbb{R}$ is completely monotone but not constant, then $\varphi(\|\cdot\|^2)$ is strictly positive definite and radial on $\mathbb{R}^s$ for any s.*

Proof: Very similar to earlier proofs. We obtain strictness by using the measure condition, i.e., the property that φ is not constant. $\square$

Example: The following functions are completely monotone and not constant. Therefore, they lead to strictly positive definite radial functions on any $\mathbb{R}^s$, and can be used as basic functions to generate bases for (1.5).

1. The functions $\varphi(r) = (r + \alpha^2)^{-\beta}$, $\alpha,\ \beta > 0$, are completely monotone and not constant since

   $$(-1)^\ell \varphi^{(\ell)}(r) = (-1)^{2\ell} \beta(\beta+1)\cdots(\beta+\ell-1)(r+\alpha^2)^{-\beta-\ell} \geq 0, \qquad \ell = 0, 1, 2, \ldots.$$

   Thus

   $$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^N c_j \left(\|\boldsymbol{x} - \boldsymbol{x}_j\|^2 + \alpha^2\right)^{-\beta}, \qquad \boldsymbol{x} \in \mathbb{R}^s,$$

can be used to solve the scattered data interpolation problem. The associated interpolation matrix is guaranteed to be positive definite. These functions are the inverse multiquadrics encountered earlier. Now it is clear that the earlier restriction $\beta > \frac{s}{2}$ is no longer required.

2. The functions $\varphi(r) = e^{-\alpha r}$, $\alpha > 0$, are completely monotone and not constant since
$$(-1)^\ell \varphi^{(\ell)}(r) = \alpha^\ell e^{-\alpha r} \geq 0, \qquad \ell = 0, 1, 2, \ldots .$$
Thus
$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{N} c_j e^{-\alpha \|\boldsymbol{x} - \boldsymbol{x}_j\|^2}, \qquad \boldsymbol{x} \in \mathbb{R}^s,$$
corresponds to interpolation with Gaussian radial basis functions.

Remarks:

1. A complete characterization of strictly positive definite functions in terms of completely monotone functions, i.e., the converse of Schoenberg's Theorem 2.5.4, is given in Wendland's book [634].

2. We just saw (for the second time) that Gaussians are strictly positive definite and radial on all $\mathbb{R}^s$. Also, Theorem 1.2.6 stating basic properties of positive definite functions shows us that (positive) linear combinations of (strictly) positive definite functions are (strictly) positive definite. The Schoenberg characterization of functions that are (strictly) positive definite and radial on any $\mathbb{R}^s$, Theorem 2.4.2, shows that *all* such functions are given as linear combinations of Gaussians.

2.6 Multiply Monotone Functions

As we will see below, another interesting class of functions is given by

Definition 2.6.1 *A function $\varphi : (0, \infty) \to \mathbb{R}$ which is in $C^{k-2}(0, \infty)$ ($k \geq 2$), and for which $(-1)^l \varphi^{(l)}(r)$ is non-negative, non-increasing, and convex for $l = 0, 1, 2, \ldots, k-2$ is called* k-times monotone on $(0, \infty)$. *In case $k = 1$ we only require $\varphi \in C(0, \infty)$ to be non-negative and non-increasing.*

Since convexity of φ means that $\varphi(\frac{r_1+r_2}{2}) \leq \frac{\varphi(r_1)+\varphi(r_2)}{2}$, or simply $\varphi''(r) \geq 0$ if φ'' exists, a multiply monotone function is in essence just a completely monotone function whose monotonicity is "truncated".

Examples:

1. The truncated power function
$$\varphi_\ell(r) = (1 - r)_+^\ell$$
is ℓ-times monotone for any ℓ.

2. If we define the integral operator I by

$$(If)(r) = \int_r^\infty f(s)ds, \qquad r \geq 0,$$

and f is ℓ-times monotone, then If is $\ell + 1$-times monotone.

Remark: The operator I plays an important role in the construction of compactly supported radial basis functions (more later).

An integral representation for the class of multiply monotone functions was given by Williamson [645] but apparently already known to Schoenberg in 1940.

Theorem 2.6.2 *A continuous function $\varphi : (0, \infty) \to \mathbb{R}$ is k-times monotone on $(0, \infty)$ if and only if it is of the form*

$$\varphi(r) = \int_0^\infty (1 - rt)_+^{k-1} d\mu(t), \tag{2.7}$$

where μ is a non-negative Borel measure on $(0, \infty)$.

Proof: To see that a function of the form 2.7 is indeed multiply monotone we just need to differentiate under the integral (since derivatives up to order $k-2$ of $(1-rt)_+^{k-1}$ are continuous and bounded). The other direction can be found in [645]. □

For $k \to \infty$ this characterization is equivalent to the Hausdorff-Bernstein-Widder characterization Theorem 2.5.2. Williamson also shows that the product of multiply monotone functions is multiply monotone.

We can see from the Examples 1 and 2 of Section 2.4 that certain multiply monotone functions give rise to positive definite radial functions. Such a connection was first noted by Askey [10] using the truncated power functions of Example 1 in Section 2.4 (and in the one-dimensional case by Pólya). In the RBF literature the following theorem was stated in Micchelli's paper [456], and then refined by Buhmann [79]:

Theorem 2.6.3 *Let $k = \lfloor s/2 \rfloor + 2$ be a positive integer. If $\varphi : (0, \infty) \to \mathbb{R}$ is k-times monotone on $(0, \infty)$ but not constant, then $\varphi(\| \cdot \|^2)$ is strictly positive definite and radial on $\mathbb{R}^s$.*

Remark: Most versions of Theorem 2.6.3 contain misprints in the literature. The correct form should be as stated above.

Wu [655] states

Theorem 2.6.4 *A function $\varphi : [0, \infty) \to \mathbb{R}$ is strictly positive definite and radial on $\mathbb{R}^s$ for $s \leq 2k + 1$ if and only if $\varphi(r)r^{2k} \in L_1(0, \infty) \cap C[0, \infty)$ and $\mathcal{F}_1\varphi(\| \cdot \|^2/2)$ is k-times monotone.*

Using this theorem he starts with the truncated power function $f_k(r) = (1 - 2r)_+^k$ (which is k-times monotone) and obtains functions of the form

$$\varphi_k(r) = \mathcal{F}_1 f_k(\cdot^2/2)(r) = \sqrt{\frac{2}{\pi}} \int_0^\infty (1 - t^2)_+^k \cos(rt)dt$$

which are strictly positive definite and radial in $\mathbb{R}^{2k+1}$ and for which $\mathcal{F}_1\varphi_k(\|\cdot\|^2/2)$ is multiply monotone, i.e.,

$$(-1)^\ell \frac{d^\ell}{dr^\ell}\left(\mathcal{F}_1\varphi_k(\cdot^2/2)\right)(r) = \frac{2^{k-\ell}k!}{(k-\ell)!} f_{k-\ell}(r) \geq 0, \qquad 0 \leq \ell \leq k.$$

The special case $k = 0$ yields

$$\varphi_0(r) = \sqrt{\frac{2}{\pi}}\mathrm{sinc}(r),$$

and the family of functions $\{\varphi_k\}$ generalizes the sinc function used in sampling theory. These functions have a *compactly supported Fourier transform.*

However, if we start with the truncated power function $\varphi(r) = (1-2r)_+^{k+1}$, which we know to be strictly positive definite and radial in $\mathbb{R}^s$ for $s \leq 2k+1$, then (as above)

$$\mathcal{F}_1\varphi(\cdot^2/2)(r) = \sqrt{\frac{2}{\pi}}\int_0^\infty (1-t^2)_+^{k+1}\cos(rt)dt.$$

In fact, Wu gives the explicit formula

$$\sqrt{\frac{2}{\pi}}\int_0^\infty (1-t^2)_+^{k+1}\cos(rt)dt = 2^{k+1}\Gamma(k+2)r^{-k-3/2}J_{k+3/2}(r).$$

Clearly, these functions are *not* monotone. This seems to present a contradict the statement of Theorem 2.6.4.

Remark: As a final remark in this chapter we mention we are a long way from having a complete characterization of (radial) functions for which the scattered data interpolation problem has a unique solution. As we will see later, such a characterization will involve also functions which are not strictly positive definite. For example, we will mention a result of Micchelli's according to which *conditionally* positive definite functions of order one can be used for the scattered data interpolation problem. Furthermore, all of the results dealt with so far involve radial basis functions which are centered at the given data sites. There are only limited results addressing the situation in which the centers for the basis functions and the data sites may differ.

Chapter 3

Scattered Data Interpolation with Polynomial Precision and Conditionally Positive Definite Functions

3.1 Scattered Data Interpolation with Polynomial Precision

Sometimes the assumption on the form (1.1) of the solution to the scattered data interpolation Problem 1.2.1 is extended by adding certain polynomials to the expansion, i.e., $\mathcal{P}f$ is now assumed to be of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{k=1}^{N} c_k B_k(\boldsymbol{x}) + \sum_{l=1}^{M} d_l p_l(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s, \tag{3.1}$$

where $p_1, \ldots, p_M$ form a basis for the $M = \binom{s+m-1}{m-1}$-dimensional linear space Π^s_{m-1} of polynomials of total degree less than or equal to $m-1$ in s variables.

Since enforcing the interpolation conditions $\mathcal{P}f(\boldsymbol{x}_i) = f(\boldsymbol{x}_i)$, $i = 1, \ldots, N$, leads to a system of N linear equations in the $N + M$ unknowns c_k and d_l one usually adds the M additional conditions

$$\sum_{k=1}^{N} c_k p_l(\boldsymbol{x}_k) = 0, \qquad l = 1, \ldots, M,$$

to ensure a unique solution.

Example: For $m = s = 2$ we add the space of bivariate linear polynomials, i.e., $\Pi^2_1 = \text{span}\{1, x, y\}$. Using the notation $\boldsymbol{x} = (x, y)$ we get the expansion

$$\mathcal{P}f(x, y) = \sum_{k=1}^{N} c_k B_k(x, y) + d_1 + d_2 x + d_3 y, \qquad \boldsymbol{x} = (x, y) \in \mathbb{R}^2,$$

which we use to solve

$$\mathcal{P}f(x_i, y_i) = f(x_i, y_i), \qquad i = 1, \ldots, N,$$

together with the three additional conditions

$$\begin{aligned} \sum_{k=1}^{N} c_k &= 0, \\ \sum_{k=1}^{N} c_k x_k &= 0, \\ \sum_{k=1}^{N} c_k y_k &= 0. \end{aligned}$$

Remark: While the use of polynomials is somewhat arbitrary (any other set of M linearly independent functions could be used), it is obvious that the addition of polynomials of total degree at most $m-1$ guarantees polynomial precision, i.e., if the data come from a polynomial of total degree less than or equal to $m-1$ they are fitted by that polynomial.

In general, solving the interpolation problem based on the extended expansion (3.1) now amounts to solving a system of linear equations of the form

$$\begin{bmatrix} A & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix}, \tag{3.2}$$

where the pieces are given by $A_{jk} = B_k(\boldsymbol{x}_j)$, $j, k = 1, \ldots, N$, $P_{jl} = p_l(\boldsymbol{x}_j)$, $j = 1, \ldots, N$, $l = 1, \ldots, M$, $\boldsymbol{c} = [c_1, \ldots, c_N]^T$, $\boldsymbol{d} = [d_1, \ldots, d_M]^T$, $\boldsymbol{y} = [y_1, \ldots, y_N]^T$, and $\boldsymbol{0}$ is a zero vector of length M.

It is possible to formulate a theorem concerning the well-posedness of this interpolation problem. As in the previous chapter we begin with an appropriate definition from the linear algebra literature. This, however, covers only the case $m = 1$.

Definition 3.1.1 *A real symmetric matrix A is called* conditionally positive semi-definite of order one *if its associated quadratic form is non-negative, i.e.,*

$$\sum_{j=1}^{N} \sum_{k=1}^{N} c_j c_k A_{jk} \geq 0 \tag{3.3}$$

for all $\boldsymbol{c} = [c_1, \ldots, c_N]^T \in \mathbb{R}^N$ which satisfy

$$\sum_{j=1}^{N} c_j = 0.$$

If $\boldsymbol{c} \neq \boldsymbol{0}$ implies strict inequality in (3.3) then A is called conditionally positive definite of order m.

Remarks:

1. In the linear algebra literature the definition usually uses "$\leq$", and then A is referred to as (conditionally or almost) *negative* definite.

2. Obviously, conditionally positive definite matrices of order one exist only for $N > 1$.

3. Conditional positive definiteness of order one of a matrix A can also be interpreted as A being positive definite on the space of vectors $\boldsymbol{c}$ such that

$$\sum_{j=1}^{N} c_j = 0.$$

   Thus, in this sense, A is positive definite on the space of vectors $\boldsymbol{c}$ "perpendicular" to constant functions.

Since an $N \times N$ matrix which is conditionally positive definite of order one is positive definite on a subspace of dimension $N-1$ it has the interesting property that at least $N-1$ of its eigenvalues are positive. This follows immediately from the Courant-Fischer Theorem of linear algebra (see e.g., [431], Thm. 5.8(a)):

Theorem 3.1.2 *Let A be a symmetric $N \times N$ matrix with eigenvalues*

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N.$$

Let $1 \leq k \leq N$ and for each subspace W with $\dim W = N - k + 1$ set

$$c_k(W) = \max_{\boldsymbol{x} \in W, \|\boldsymbol{x}\|=1} \boldsymbol{x}^T A \boldsymbol{x}$$

and

$$d_k(W) = \min_{\boldsymbol{x} \in W, \|\boldsymbol{x}\|=1} \boldsymbol{x}^T A \boldsymbol{x}.$$

Then

$$c_k(W) \geq \lambda_k, \quad d_k(W) \leq \lambda_{N-k+1}, \quad k = 1, \ldots, N.$$

With an additional hypothesis on A we can make an even stronger statement.

Theorem 3.1.3 *An $N \times N$ matrix A which is conditionally positive definite of order one and has a non-positive trace has 1 negative and $N-1$ positive eigenvalues.*

Proof: From the Courant-Fischer Theorem we get that A has at least $N-1$ positive eigenvalues. But since $\mathrm{tr}(A) = \sum_{i=1}^{N} \lambda_i \leq 0$, where the λ_i denote the eigenvalues of A, A also must have at least one negative eigenvalue. $\square$

3.2 Conditionally Positive Definite Functions

In analogy to the earlier discussion of interpolation with positive definite functions we will now introduce conditionally positive definite and strictly conditionally positive definite functions of order m.

Definition 3.2.1 *A complex-valued continuous function Φ is called* conditionally positive definite of order m on $\mathbb{R}^s$ *if*

$$\sum_{j=1}^{N}\sum_{k=1}^{N} c_j \overline{c_k} \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) \geq 0 \tag{3.4}$$

for any N points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^s$, and $\boldsymbol{c} = [c_1, \ldots, c_N]^T \in \mathbb{C}^N$ satisfying

$$\sum_{j=1}^{N} c_j p(\boldsymbol{x}_j) = 0,$$

for any complex-valued polynomial p of degree at most $m-1$. The function Φ is called strictly conditionally positive definite of order m on $\mathbb{R}^s$ *if the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^s$ are distinct, and $\boldsymbol{c} \neq \boldsymbol{0}$ implies strict inequality in (3.4).*

An immediate observation is that a function which is conditionally positive definite of order m on $\mathbb{R}^s$ also is conditionally positive definite of any higher order. In particular, this definition is more general than that for positive definite functions since the case $m = 0$ yields that class of functions, i.e., (strictly) conditionally positive definite functions of order zero are (strictly) positive definite, and therefore a (strictly) positive definite function is always (strictly) conditionally positive definite of any order.

As for positive definite functions earlier, we can restrict ourselves to real-valued, even functions Φ and real coefficients. A detailed discussion is presented in [634].

Theorem 3.2.2 *A real-valued continuous even function Φ is called* conditionally positive definite of order m on $\mathbb{R}^s$ *if*

$$\sum_{j=1}^{N}\sum_{k=1}^{N} c_j c_k \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) \geq 0 \tag{3.5}$$

for any N points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^s$, and $\boldsymbol{c} = [c_1, \ldots, c_N]^T \in \mathbb{R}^N$ satisfying

$$\sum_{j=1}^{N} c_j \boldsymbol{x}_j^{\boldsymbol{\alpha}} = 0, \qquad |\boldsymbol{\alpha}| < m, \quad \boldsymbol{\alpha} \in \mathbb{N}_0^s\,.$$

The function Φ is called strictly conditionally positive definite of order m on $\mathbb{R}^s$ *if the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^s$ are distinct, and $\boldsymbol{c} \neq \boldsymbol{0}$ implies strict inequality in (3.5).*

Here we have used the usual *multi-integer notation*, i.e.,

$$\boldsymbol{\alpha} \in \mathbb{N}_0^s, \quad |\boldsymbol{\alpha}| = \sum_{i=1}^{s} \alpha_i, \quad \text{and} \quad \boldsymbol{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_s^{\alpha_s}.$$

Remarks:

1. The matrix A with entries $A_{jk} = \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k)$ corresponding to a real and even strictly conditionally positive definite function of order m can also be interpreted as being positive definite on the space of vectors $\boldsymbol{c}$ such that

   $$\sum_{j=1}^{N} c_j \boldsymbol{x}^{\boldsymbol{\alpha}} = 0, \qquad |\boldsymbol{\alpha}| < m.$$

   Thus, in this sense, A is positive definite on the space of vectors $\boldsymbol{c}$ "perpendicular" to polynomials of degree at most $m-1$.

2. The Courant-Fischer Theorem now implies that A has at least $N-m$ positive eigenvalues.

Using Theorem 3.1.3 we can see that interpolation with strictly conditionally positive definite functions of order one is possible even without adding a polynomial term. This was first observed by Micchelli [456].

Theorem 3.2.3 *Suppose Φ is strictly conditionally positive definite of order one and that $\Phi(\mathbf{0}) \le 0$. Then for any distinct points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^s$ the matrix A with entries $A_{jk} = \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k)$ has $N-1$ positive and 1 negative eigenvalue, and is therefore non-singular.*

Proof: Clearly, the matrix A is conditionally positive definite. Moreover, the trace of A is given by $\mathrm{tr}(A) = N\Phi(0) \le 0$. Therefore, Theorem 3.1.3 applies. □

As we will see below, this theorem covers the *multiquadrics* $\Phi(x) = -(\|\boldsymbol{x}\|^2 + \alpha^2)^\beta$, $\alpha \ge 0$, $0 < \beta < 1$.

Another special property of a conditionally positive definite function of order one is

Lemma 3.2.4 *If C is an arbitrary real constant and the real even function Φ is (strictly) conditionally positive definite of order one, then $\Phi + C$ is also (strictly) conditionally positive definite of order one.*

Proof: Simply consider

$$\sum_{j=1}^{N}\sum_{k=1}^{N} c_j c_k [\Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) + C] = \sum_{j=1}^{N}\sum_{k=1}^{N} c_j c_k \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) + \sum_{j=1}^{N}\sum_{k=1}^{N} c_j c_k C.$$

The second term on the right is zero since Φ is conditionally positive definite of order one, i.e., $\sum_{j=1}^{N} c_j = 0$, and thus the statement follows. □

Before we formulate the theorem about the uniqueness of the solution to the interpolation problem based on expansion (3.1), we define a property which forms a very mild restriction on the location of the data sites.

Definition 3.2.5 *We call a set of points* $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{R}^s$ m-unisolvent *if the only polynomial of total degree at most* m *interpolating zero data on* $\mathcal{X}$ *is the zero polynomial.*

This definition comes from polynomial interpolation, in which case it guarantees a unique solution for interpolation to given data at a subset of the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ by a polynomial of degree m. A sufficient condition (to be found in [140], Ch. 9) on the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ to form an m-unisolvent set in $\mathbb{R}^2$ is

Theorem 3.2.6 *Suppose* $\{L_0, \ldots, L_m\}$ *is a set of* $m+1$ *distinct lines in* $\mathbb{R}^2$*, and that* $\mathcal{U} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M\}$ *is a set of* $M = (m+1)(m+2)/2$ *distinct points such that the first point lies on* L_0*, the next two points lie on* L_1 *but not on* L_0*, and so on, so that the last* $m+1$ *points lie on* L_m *but not on any of the previous lines* $L_0, \ldots, L_{m-1}$*. Then there exists a unique interpolation polynomial of total degree at most* m *to arbitrary data given at the points in* $\mathcal{U}$*. Furthermore, if the data sites* $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ *contain* $\mathcal{U}$ *as a subset then they form an* m*-unisolvent set on* $\mathbb{R}^2$*.*

Proof: We use induction on m. For $m = 0$ the result is trivial. Take R to be the matrix arising from polynomial interpolation at the points in $\mathcal{U}$, i.e.,

$$R_{jk} = p_k(\boldsymbol{u}_j), \quad j, k = 1, \ldots, M,$$

where the p_k form a basis of Π_m^2. We want to show that the only possible solution to $R\boldsymbol{c} = \boldsymbol{0}$ is $\boldsymbol{c} = \boldsymbol{0}$. This is equivalent to showing that if $p \in \Pi_m^2$ satisfies

$$p(\boldsymbol{u}_i) = 0, \quad i = 1, \ldots, M,$$

then p is the zero polynomial.

For each $i = 1, \ldots, m$, let the equation of the line L_i be given by

$$\alpha_i x_1 + \beta_i x_2 = \gamma_i.$$

Suppose now that p interpolates zero data at all the points $\boldsymbol{u}_i$ as stated above. Since p reduces to a univariate polynomial of degree m on L_m which vanishes at $m+1$ distinct points on L_m, it follows that p vanishes identically on L_m, and so

$$p(x_1, x_2) = (\alpha_m x_1 + \beta_m x_2 - \gamma_m) q(x_1, x_2),$$

where q is a polynomial of degree $m-1$. But now q satisfies the hypothesis of the theorem with m replaced by $m-1$ and $\mathcal{U}$ replaced by $\widetilde{\mathcal{U}}$ consisting of the first $\binom{m+1}{2}$ points of U. By induction, therefore $q \equiv 0$, and thus $p \equiv 0$. This establishes the uniqueness of the interpolation polynomial. The last statement of the theorem is obvious. □

Remarks:

1. This theorem can be generalized to $\mathbb{R}^s$ by using hyperplanes in $\mathbb{R}^s$, and induction on s. Chui also gives an explicit expression for the determinant of the interpolation matrix associated with polynomial interpolation at the set of points $\mathcal{U}$.

2. A theorem similar to Theorem 3.2.6 is already proved by Chung and Yao [143].

3. $(m-1)$-unisolvency of the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ is equivalent to the fact that the matrix P with

$$P_{jl} = p_l(\boldsymbol{x}_j), \quad j = 1, \ldots, N, \ l = 1, \ldots, M,$$

where M and N are chosen as in (3.1), has full (column-)rank.

Example: As can easily be verified, three collinear points in $\mathbb{R}^2$ are not 1-unisolvent, since a linear interpolant, i.e., a plane through three arbitrary heights at these 3 collinear points is not uniquely determined. On the other hand, if a set of points in $\mathbb{R}^2$ contains 3 non-collinear points, then it is 1-unisolvent.

Now we are ready to formulate and prove

Theorem 3.2.7 *If the real-valued even function Φ is strictly conditionally positive definite of order m on $\mathbb{R}^s$ and the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ form an $(m-1)$-unisolvent set, then the system of linear equations (3.2) is uniquely solvable.*

Proof: Assume $[\boldsymbol{c}, \boldsymbol{d}]^T$ is a solution of the homogeneous linear system, i.e., $\boldsymbol{y} = \boldsymbol{0}$. We show that $[\boldsymbol{c}, \boldsymbol{d}]^T = \boldsymbol{0}$ is the only possible solution.

Multiplication of the top block by $\boldsymbol{c}^T$ yields

$$\boldsymbol{c}^T A \boldsymbol{c} + \boldsymbol{c}^T P \boldsymbol{d} = 0.$$

From the bottom block of (3.2) we know $\boldsymbol{c}^T P = \boldsymbol{0}$, and therefore

$$\boldsymbol{c}^T A \boldsymbol{c} = 0.$$

Since the matrix A is conditionally positive definite by assumption we get that $\boldsymbol{c} = \boldsymbol{0}$. The unisolvency of the data sites, i.e., the linear independence of the columns of P, and the fact that $\boldsymbol{c} = \boldsymbol{0}$ guarantee $\boldsymbol{d} = \boldsymbol{0}$ from the top block

$$A\boldsymbol{c} + P\boldsymbol{d} = \boldsymbol{0}$$

of (3.2). □

3.3 An Analog of Bochner's Theorem

In order to give an analog of Bochner's theorem for conditionally positive definite functions we have to introduce a few concepts from distribution theory. The approach described in this section is essentially due to Madych and Nelson [417].

For the definition of generalized Fourier transforms required below we have to define the *Schwartz space* of rapidly decreasing test functions

$$\mathcal{S} = \{\gamma \in C^\infty(\mathbb{R}^s) : \lim_{\|\boldsymbol{x}\| \to \infty} \boldsymbol{x}^{\boldsymbol{\alpha}} (D^{\boldsymbol{\beta}} \gamma)(\boldsymbol{x}) = 0, \ \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}_0^s\},$$

where

$$D^{\boldsymbol{\beta}} = \frac{\partial^{|\boldsymbol{\beta}|}}{\partial x_1^{\beta_1} \cdots \partial x_s^{\beta_s}}, \qquad |\boldsymbol{\beta}| = \sum_{i=1}^{s} \beta_i.$$

Remarks:

1. The space $\mathcal{S}$ consists of all those functions $\gamma \in C^\infty(\mathbb{R}^s)$ which, together with all their derivatives, decay faster than any power of $1/\|\boldsymbol{x}\|$.

2. The space $\mathcal{S}$ contains the space $C_0^\infty(\mathbb{R}^s)$, the space of all infinitely differentiable functions on $\mathbb{R}^s$ with compact support. We also note that $C_0^\infty(\mathbb{R}^s)$ is a true subspace of $\mathcal{S}$ since, e.g., the function $\gamma(\boldsymbol{x}) = e^{-\|\boldsymbol{x}\|^2}$ belongs to $\mathcal{S}$ but not to $C_0^\infty(\mathbb{R}^s)$.

3. A remarkable fact about the Schwartz space is that $\gamma \in \mathcal{S}$ has a classical Fourier transform $\hat{\gamma}$ which is also in $\mathcal{S}$.

Of particular importance will be the following subspace $\mathcal{S}_m$ of $\mathcal{S}$

$$\mathcal{S}_m = \{\gamma \in \mathcal{S} : \ \gamma(\boldsymbol{x}) = \mathcal{O}(\|\boldsymbol{x}\|^m) \text{ for } \|\boldsymbol{x}\| \to 0, \ m \in \mathbb{N}_0\}.$$

Furthermore, the set $\mathcal{B}$ of slowly increasing functions is given by

$$\mathcal{B} = \{f \in C(\mathbb{R}^s) : \ |f(\boldsymbol{x})| \leq |p(\boldsymbol{x})| \text{ for some polynomial } p \in \Pi^s\}.$$

The generalized Fourier transform is now given by

Definition 3.3.1 *Let $\Phi \in \mathcal{B}$ be complex-valued. A continuous function $\hat{\Phi} : \mathbb{R}^s \setminus \{0\} \to \mathbb{C}$ is called the* generalized Fourier transform *of Φ if there exists an integer $m \in \mathbb{N}_0$ such that*

$$\int_{\mathbb{R}^s} \Phi(\boldsymbol{x})\hat{\gamma}(\boldsymbol{x})d\boldsymbol{x} = \int_{\mathbb{R}^s} \hat{\Phi}(\boldsymbol{x})\gamma(\boldsymbol{x})d\boldsymbol{x}$$

is satisfied for all $\gamma \in \mathcal{S}_{2m}$. The smallest such integer m is called the order *of $\hat{\Phi}$.*

Remarks:

1. Since one can show that the generalized Fourier transform of an s-variate polynomial of degree at most $2m$ is zero, it follows that the inverse generalized Fourier transform is only unique up to addition of such a polynomial.

2. Various definitions of the generalized Fourier transform exist in the literature. A classical reference is the book by Gelfand and Vilenkin [250].

3. The order of the generalized Fourier transform is nothing but the order of the singularity at the origin of the generalized Fourier transform.

4. For functions in $L_1(\mathbb{R}^s)$ the generalized Fourier transform coincides with the classical Fourier transform, and for functions in $L_2(\mathbb{R}^s)$ it coincides with the distributional Fourier transform.

We now immediately give a characterization of strictly conditionally positive definite functions on $\mathbb{R}^s$ due to Iske (see [314] or [634] for details).

Theorem 3.3.2 *Suppose the complex-valued function $\Phi \in \mathcal{B}$ possesses a generalized Fourier transform $\hat{\Phi}$ of order m which is continuous on $\mathbb{R}^s \setminus \{0\}$. Then Φ is strictly conditionally positive definite of order m if and only if $\hat{\Phi}$ is non-negative and non-vanishing.*

Remarks:

1. Theorem 3.3.2 states that strictly conditionally positive definite functions on $\mathbb{R}^s$ are characterized by the order of the singularity of their generalized Fourier transform at the origin, provided that this generalized Fourier transform is non-negative and non-zero.

2. An integral characterization of conditionally positive definite functions of order m also exists. It can be found in a paper by Sun [597] (see also [634]).

Examples: Wendland [634] explicitly computes the generalized Fourier transforms for various popular basis functions.

1. The *multiquadrics*

$$\Phi(\boldsymbol{x}) = (\|\boldsymbol{x}\|^2 + \alpha^2)^\beta, \qquad \boldsymbol{x} \in \mathbb{R}^s,\ \alpha > 0,\ \beta \in \mathbb{R} \setminus \mathbb{N}_0,$$

have generalized Fourier transforms

$$\hat{\Phi}(\boldsymbol{\omega}) = \frac{2^{1+\beta}}{\Gamma(-\beta)} \left(\frac{\|\boldsymbol{\omega}\|}{\alpha} \right)^{-\beta - s/2} K_{\beta+s/2}(\alpha\|\boldsymbol{\omega}\|), \qquad \boldsymbol{\omega} \neq \mathbf{0},$$

of order $m = \max(0, \lceil \beta \rceil)$. Here K_ν is the *modified Bessel function of the second kind* (sometimes also called modified Bessel function of the third kind, or MacDonald's function) *of order* ν. Therefore, the functions

$$\Phi(\boldsymbol{x}) = (-1)^{\lceil \beta \rceil} (\|\boldsymbol{x}\|^2 + \alpha^2)^\beta, \qquad \beta > 0,\ \beta \notin \mathbb{N},$$

are strictly conditionally positive definite of order $m = \lceil \beta \rceil$ (and higher). In particular, we can use

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{k=1}^{N} c_k \sqrt{\|\boldsymbol{x} - \boldsymbol{x}_k\|^2 + \alpha^2} + d, \qquad \boldsymbol{x} \in \mathbb{R}^s, \alpha > 0,$$

together will the constraint

$$\sum_{k=1}^{N} c_k = 0$$

to solve the scattered data interpolation problem. The resulting interpolant will be exact for constant data. By Theorem 3.2.3 we can also use

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{k=1}^{N} c_k \sqrt{\|\boldsymbol{x} - \boldsymbol{x}_k\|^2 + \alpha^2}, \qquad \boldsymbol{x} \in \mathbb{R}^s, \alpha > 0.$$

Also, the inverse multiquadrics

$$\Phi(\boldsymbol{x}) = (\|\boldsymbol{x}\|^2 + \alpha^2)^\beta, \beta < 0,$$

are again shown to be strictly conditionally positive definite of order $m = 0$, i.e., strictly positive definite.

2. The *powers*

$$\Phi(\boldsymbol{x}) = \|\boldsymbol{x}\|^\beta, \qquad \boldsymbol{x} \in \mathbb{R}^s,\ \beta > 0,\ \beta \notin 2\,\mathbb{N},$$

have generalized Fourier transforms

$$\hat{\Phi}(\boldsymbol{\omega}) = \frac{2^{\beta+s/2}\Gamma(\frac{s+\beta}{2})}{\Gamma(-\beta/2)}\|\boldsymbol{\omega}\|^{-\beta-s}, \qquad \boldsymbol{\omega} \neq \boldsymbol{0},$$

of order $m = \lceil \beta/2 \rceil$. Therefore, the functions

$$\Phi(\boldsymbol{x}) = (-1)^{\lceil \beta/2 \rceil}\|\boldsymbol{x}\|^\beta, \qquad \beta > 0,\ \beta \notin 2\,\mathbb{N},$$

are strictly conditionally positive definite of order $m = \lceil \beta/2 \rceil$ (and higher).

3. The *thin plate splines* (or surface splines)

$$\Phi(\boldsymbol{x}) = \|\boldsymbol{x}\|^{2k}\log\|\boldsymbol{x}\|, \qquad \boldsymbol{x} \in \mathbb{R}^s,\ k \in \mathbb{N},$$

have generalized Fourier transforms

$$\hat{\Phi}(\boldsymbol{\omega}) = (-1)^{k+1}2^{2k-1+s/2}\Gamma(k+s/2)k!\|\boldsymbol{\omega}\|^{-s-2k}$$

of order $m = k + 1$. Therefore, the functions

$$\Phi(\boldsymbol{x}) = (-1)^{k+1}\|\boldsymbol{x}\|^{2k}\log\|\boldsymbol{x}\|, \qquad k \in \mathbb{N},$$

are strictly conditionally positive definite of order $m = k + 1$. In particular, we can use

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{k=1}^{N} c_k\|\boldsymbol{x} - \boldsymbol{x}_k\|^2\log\|\boldsymbol{x} - \boldsymbol{x}_k\| + d_1 + d_2x + d_3y, \qquad \boldsymbol{x} = (x, y) \in \mathbb{R}^2,$$

together will the constraints

$$\begin{aligned}
\sum_{k=1}^{N} c_k &= 0,\\
\sum_{k=1}^{N} c_k x_k &= 0,\\
\sum_{k=1}^{N} c_k y_k &= 0,
\end{aligned}$$

to solve the scattered data interpolation problem provided the data sites are not all collinear. The resulting interpolant will be exact for data coming from a linear function.

Remark: As for strictly positive definite radial functions, we will be able to connect strictly conditionally positive definite radial functions to completely monotone functions, and thus be able to obtain a simpler criterion for checking conditional positive definiteness.

3.4 Conditionally Positive Definite Radial Functions

In analogy to the discussion in Chapter 2 we now focus on conditionally positive definite functions which are radial on $\mathbb{R}^s$ for all s. The paper [273] by Guo, Hu and Sun contains an integral characterization for such functions. This characterization is too technical to be included here.

The main result in [273] is a characterization of conditionally positive definite radial functions on $\mathbb{R}^s$ for all s in terms of completely monotone functions.

Theorem 3.4.1 *Let $\varphi \in C[0,\infty) \cap C^\infty(0,\infty)$. Then the function $\Phi = \varphi(\|\cdot\|^2)$ is conditionally positive definite of order m and radial on $\mathbb{R}^s$ for all s if and only if $(-1)^m\varphi^{(m)}$ is completely monotone on $(0,\infty)$.*

Proof: Micchelli [456] proved that complete monotonicity implies conditional positive definiteness. He also conjectured that the converse holds, and gave a simple proof for this in the case $m = 1$. For $m = 0$ this is Schoenberg's characterization of positive definite radial functions on $\mathbb{R}^s$ for all s in terms of completely monotone functions (Theorem 2.5.3). The remaining part of the theorem is shown in [273]. □

In order to get strict conditional positive definiteness we need to generalize Theorem 2.5.4, i.e., the fact that φ not be constant.

Theorem 3.4.2 *If φ is as in Theorem 3.4.1 and not a polynomial of degree at most m, then Φ is strictly conditionally positive definite of order m and radial on $\mathbb{R}^s$ for all s.*

Examples: We can now more easily verify the conditional positive definiteness of the functions listed in the previous example.

1. The functions

$$\varphi(r) = (-1)^{\lceil\beta\rceil}(r+\alpha^2)^\beta, \qquad \alpha > 0,\ \beta > 0, \beta \notin \mathbb{N}$$

imply

$$\varphi^{(k)}(r) = (-1)^{\lceil\beta\rceil}\beta(\beta-1)\cdots(\beta-k+1)(r+\alpha^2)^{\beta-k}$$

so that

$$(-1)^{\lceil\beta\rceil}\varphi^{(\lceil\beta\rceil)}(r) = \beta(\beta-1)\cdots(\beta-\lceil\beta\rceil+1)(r+\alpha^2)^{\beta-\lceil\beta\rceil}$$

is completely monotone. Moreover, $m = \lceil\beta\rceil$ is the smallest possible m such that $(-1)^m\varphi^{(m)}$ is completely monotone. Therefore, the multiquadrics

$$\Phi(r) = (-1)^{\lceil\beta\rceil}(r^2+\alpha^2)^\beta, \qquad \alpha > 0,\ \beta > 0,$$

are strictly conditionally positive definite of order $m \geq \lceil\beta\rceil$ and radial on $\mathbb{R}^s$ for all s.

2. The functions
$$\varphi(r) = (-1)^{\lceil \beta/2 \rceil} r^{\beta/2}, \qquad \beta > 0,\ \beta \notin 2\,\mathbb{N},$$
imply
$$\varphi^{(k)}(r) = (-1)^{\lceil \beta/2 \rceil} \frac{\beta}{2} \left(\frac{\beta}{2} - 1 \right) \cdots \left(\frac{\beta}{2} - k + 1 \right) r^{\beta/2-k}$$
so that $(-1)^{\lceil \beta/2 \rceil} \varphi^{(\lceil \beta/2 \rceil)}$ is completely monotone and $m = \lceil \beta/2 \rceil$ is the smallest possible m such that $(-1)^m \varphi^{(m)}$ is completely monotone. Therefore, the powers
$$\Phi(r) = (-1)^{\lceil \beta/2 \rceil} r^{\beta}, \qquad \beta > 0,\ \beta \notin 2\,\mathbb{N},$$
are strictly conditionally positive definite of order $m \geq \lceil \beta/2 \rceil$ and radial on $\mathbb{R}^s$ for all s.

3. The thin plate splines
$$\Phi(\|\boldsymbol{x}\|) = (-1)^{k+1} \|\boldsymbol{x}\|^{2k} \log \|\boldsymbol{x}\|, \qquad k \in \mathbb{N},$$
are strictly conditionally positive definite of order $m \geq k+1$ and radial on $\mathbb{R}^s$ for all s. To see this we observe that
$$2\Phi(\|\boldsymbol{x}\|) = (-1)^{k+1} \|\boldsymbol{x}\|^{2k} \log(\|\boldsymbol{x}\|^2).$$
Therefore, we let
$$\varphi(r) = (-1)^{k+1} r^k \log r, \qquad k \in \mathbb{N},$$
and get
$$\varphi^{(\ell)}(r) = (-1)^{k+1} k(k-1) \cdots (k-\ell+1) r^{k-\ell} \log r + p_\ell(r), \qquad 1 \leq \ell \leq k,$$
with p_ℓ a polynomial of degree $k - \ell$. Therefore,
$$\varphi^{(k)}(r) = (-1)^{k+1} k! \log r + C$$
and
$$\varphi^{(k+1)}(r) = (-1)^{k+1} \frac{k!}{r},$$
which is completely monotone on $(0, \infty)$.

We can also apply the integral representation of completely monotone functions from the Hausdorff-Bernstein-Widder Theorem to the previous result. Then we get

Theorem 3.4.3 *A necessary and sufficient condition that the function $\Phi = \varphi(\|\cdot\|^2)$ be conditionally positive definite of order m and radial on $\mathbb{R}^s$ for all s is that its $\varphi^{(m)}$ satisfy*
$$(-1)^m \varphi^{(m)}(r) = \int_0^\infty e^{-rt} d\mu(t), \quad r > 0,$$
where μ is a non-negative Borel measure on $(0, \infty)$ such that
$$\int_0^1 d\mu(t) < \infty \quad \textit{and} \quad \int_1^\infty \frac{d\mu(t)}{t^m} < \infty.$$

The following examples of functions which are conditionally positive definite of order $m = 0$ or $m = 1$ and radial on $\mathbb{R}^s$ for all s are taken from [521]. They are listed with the associated measures corresponding to the formulation of Theorem 3.4.3.

Example:

1. $\Phi(r) = -r$: $\quad m = 1, \quad d\mu(t) = -\dfrac{1}{2\sqrt{\pi t}}dt,$
2. $\Phi(r) = -\sqrt{1+r^2}$: $\quad m = 1, \quad d\mu(t) = -\dfrac{e^{-t}}{2\sqrt{\pi t}}dt,$
3. $\Phi(r) = \dfrac{1}{\sqrt{1+r^2}}$: $\quad m = 0, \quad d\mu(t) = \dfrac{e^{-t}}{\sqrt{\pi t}}dt,$
4. $\Phi(r) = e^{-\alpha r^2}$, $\alpha > 0$: $\quad m = 0, \quad d\mu(t) = \delta(t-\rho)dt$, i.e., point evaluation at ρ.

Finally, Micchelli proved a more general version of Theorem 2.6.3 theorem relating conditionally positive definite radial functions of order m on $\mathbb{R}^s$ and multiply monotone functions. We state a stronger version due to Buhmann [79] which ensures strict conditional positive definiteness.

Theorem 3.4.4 *Let $k = \lfloor s/2 \rfloor - m + 2$ be a positive integer, and suppose $\varphi \in C^{m-1}[0,\infty)$ is not a polynomial of degree at most m. If $(-1)^m \varphi^{(m)}$ is k-times monotone on $(0,\infty)$ but not constant, then $\Phi = \varphi(\|\cdot\|^2)$ is strictly conditionally positive definite of order m and radial on $\mathbb{R}^s$.*

Remark: The converse of the above result is open.

Just as we showed earlier that compactly supported radial function cannot be strictly positive definite on $\mathbb{R}^s$ for all s, it is important to note that there are no truly conditionally positive definite functions with compact support. More precisely,

Theorem 3.4.5 *Assume that the complex-valued function $\Phi \in C(\mathbb{R}^s)$ has compact support. If Φ is strictly conditionally positive definite of (minimal) order m, then m is necessarily zero, i.e., Φ is already strictly positive definite.*

Proof: The hypotheses on Φ ensure that it is integrable, and therefore it possesses a classical Fourier transform $\hat{\Phi}$ which is continuous. For integrable functions the generalized Fourier transform coincides with the classical Fourier transform. Theorem 3.3.2 ensures that $\hat{\Phi}$ is non-negative in $\mathbb{R}^s \setminus \{\mathbf{0}\}$ and not identically equal to zero. By continuity we also get $\hat{\Phi}(\mathbf{0}) \geq 0$, and Theorem 2.3.3 shows that Φ is strictly positive definite. $\square$.

Remark: Theorem 3.4.4 together with Theorem 3.4.5 implies that if we perform m-fold anti-differentiation on a non-constant k-times monotone function, then we obtain a function that is strictly positive definite and radial on $\mathbb{R}^s$ for $s \leq 2(k+m) - 3$.

Example: The function $\varphi_k(r) = (1-r)_+^k$ is k-times monotone. To avoid the integration constant for the compactly supported truncated power function we compute the anti-derivative via

$$I\varphi_k(r) = \int_r^\infty \varphi_k(s)ds = \int_r^\infty (1-s)_+^k ds = \frac{(-1)^k}{k+1}(1-r)_+^{k+1}.$$

m-fold anti-differentiation yields

$$I^m\varphi_k(r) = II^{m-1}\varphi_k(r) = \frac{(-1)^{mk}}{(k+1)(k+2)\cdots(k+m)}(1-r)_+^{k+m}.$$

Therefore, by the Buhmann-Micchelli Theorem, the function

$$\varphi(r) = (1-r)_+^{k+m}$$

is strictly conditionally positive definite of order m and radial on $\mathbb{R}^s$ for $s \leq 2(k+m)-3$, and by Theorem 3.4.5 it is even strictly positive definite and radial on $\mathbb{R}^s$. This was also observed in Example 1 at the end of Section 2.4. In fact, we saw there that φ is strictly positive definite and radial on $\mathbb{R}^s$ for $s \leq 2(k+m)-1$.

We see that we can construct strictly positive definite compactly supported radial functions by anti-differentiating the truncated power function. This is essentially the approach taken by Wendland to construct his popular compactly supported radial basis functions. We describe this construction in the next chapter.

3.5 Composition of Conditionally Positive Definite Functions

When Schoenberg first studied conditionally positive definite matrices of order one it was in connection with isometric embeddings. Based on earlier work by Karl Menger [453] he had the following result characterizing a conditionally positive definite matrix as a certain *distance matrix* (see [568]).

Theorem 3.5.1 *Let A be a real symmetric $N \times N$ matrix with all diagonal entries zero and all other elements positive. Then $-A$ is conditionally positive semi-definite if and only if there exist N points $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N \in \mathbb{R}^N$ for which*

$$A_{jk} = \|\boldsymbol{y}_j - \boldsymbol{y}_k\|^2.$$

These points are the vertices of a simplex in $\mathbb{R}^N$.

There is also a close connection between conditionally positive semi-definite matrices and those which are positive semi-definite. This is a classical result from linear algebra called Schur's theorem. We state a stronger version due to Micchelli [456] that also covers the strict case.

Theorem 3.5.2 *A symmetric matrix $-A$ is conditionally positive semi-definite if and only if the Schur exponential $(e^{-\alpha A_{jk}})_{j,k=1}^N$ is positive semi-definite for all $\alpha > 0$. Moreover, it is positive definite if and only if*

$$A_{jk} > \frac{A_{jj} + A_{kk}}{2}, \qquad j \neq k.$$

A proof of the classical (non-strict) Schur Theorem can be found, e.g., in the book by Horn and Johnson [308].

As an immediate corollary we get an earlier result by Schoenberg (see [569], Thm. 5). We have translated Schoenberg's embedding language into that of conditionally positive definite and completely monotone functions.

Corollary 3.5.3 *A function $\varphi(\cdot)$ is conditionally positive definite of order one and radial on $\mathbb{R}^s$ for all s if and only if the functions $e^{-\alpha\varphi(\cdot^2)}$ are positive definite and radial on $\mathbb{R}^s$ for all s and for all $\alpha > 0$, i.e., $e^{-\alpha\varphi(\cdot)}$ is completely monotone for all $\alpha > 0$.*

Example: The matrix B defined by

$$B_{jk} = e^{-\|\boldsymbol{x}_j - \boldsymbol{x}_k\|^\alpha}, \quad 0 < \alpha \leq 2, \quad j,k = 1,\ldots,N,$$

is positive semi-definite, and if the points $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N$ are distinct B is positive definite. This is true since Schoenberg [569] showed that the matrix A defined by

$$A_{jk} = -\|\boldsymbol{x}_j - \boldsymbol{x}_k\|^\alpha, \quad 0 < \alpha \leq 2, \quad j,k = 1,\ldots,n,$$

is conditionally positive semi-definite, and conditionally positive definite for distinct points.

A more general result regarding the composition of conditionally positive definite functions is given by Baxter [26].

Theorem 3.5.4 *Suppose φ and ψ are functions that are conditionally positive definite of order one are radial on $\mathbb{R}^s$ with $\varphi(0) = 0$. Then $\psi \circ \varphi$ is also conditionally positive definite of order one and radial on $\mathbb{R}^s$. Indeed, if ψ is strictly conditionally positive definite of order one and radial and φ vanishes only at zero, then $\psi \circ \varphi$ is strictly conditionally positive definite of order one and radial.*

We close with some remarks.

Remarks:

1. More results with a similar flavor can be found in [26], [456], and [445].
2. Many of the results given in the previous sections can be generalized to vector-valued or even matrix-valued functions. Some work is done in [407, 408], [474], [484], and [548].

3. Another possible generalization is to consider (strictly) (conditionally) positive definite *kernels* on $X \times X$, where X is some abstract point set and $k_1, \ldots, k_m$ are given real-valued functions governing the order m of conditional positive definiteness.

4. We point out that the approach to solving the interpolation problems taken in the previous section always assumes that the *knots*, i.e., the points $\boldsymbol{x}_k$, $k = 1, \ldots, N$, at which the basis functions are centered, coincide with the data sites. This is a fairly severe restriction, and it has been shown in examples in the context of least squares approximation of scattered data (see e.g., [237, 238], or [192]) that better results can be achieved if the knots are chosen different from the data sites. Theoretical results in this direction are very limited, and are reported in [521] and in [596].

Chapter 4

Compactly Supported Radial Basis Functions

As we saw earlier, compactly supported functions Φ that are truly strictly conditionally positive definite of order $m > 0$ do not exist. The compact support automatically ensures that Φ is strictly positive definite. Another observation was that compactly supported radial functions can be strictly positive definite on $\mathbb{R}^s$ only for a fixed maximal s-value. It is not possible for a function to be strictly positive definite and radial on $\mathbb{R}^s$ for all s and also have a compact support. Therefore we focus our attention on the characterization and construction of functions that are compactly supported, strictly positive definite and radial on $\mathbb{R}^s$ for some fixed s.

According to our earlier work (Bochner's Theorem and generalizations thereof), a function is strictly positive definite and radial on $\mathbb{R}^s$ if its s-variate Fourier transform is non-negative. Theorem 2.1.2 gives the Fourier transform of $\Phi = \varphi(\|\cdot\|)$ as

$$\hat{\Phi}(\boldsymbol{x}) = \mathcal{F}_s\varphi(r) = r^{-(s-2)/2}\int_0^\infty \varphi(t)t^{s/2}J_{(s-2)/2}(rt)dt.$$

4.1 Operators for Radial Functions and Dimension Walks

Schaback and Wu [564] defined an integral operator and its inverse differential operator, and discussed an entire calculus for how these operators act on radial functions. These operators will facilitate the construction of compactly supported radial functions.

Definition 4.1.1 *1. Let φ be such that $t \mapsto t\varphi(t) \in L_1[0,\infty)$, then we define*

$$(\mathcal{I}\varphi)(r) = \int_r^\infty t\varphi(t)dt, \qquad r \geq 0.$$

2. For even $\varphi \in C^2(\mathbb{R})$ we define

$$(\mathcal{D}\varphi)(r) = -\frac{1}{r}\varphi'(r), \qquad r \geq 0.$$

In both cases the resulting functions are to be interpreted as even functions using even extension.

Remark: Note that the operator $\mathcal{I}$ differs from the operator I introduced earlier by a factor t in the integrand. However, the two operators are related. In fact, we have $\mathcal{I}\varphi(\cdot^2/2) = I\varphi(\cdot)$, i.e.,

$$\int_r^\infty t\varphi(t^2/2)dt = \int_{r^2/2}^\infty \varphi(t)dt.$$

The most important properties of these operators are (see, e.g., [564] or [627]):

Theorem 4.1.2 *1. Both $\mathcal{D}$ and $\mathcal{I}$ preserve compact support, i.e., if φ has compact support, then so do $\mathcal{D}\varphi$ and $\mathcal{I}\varphi$.*

2. If $\varphi \in C(\mathbb{R})$ and $t \mapsto t\phi(t) \in L_1[0,\infty)$, then $\mathcal{D}\mathcal{I}\varphi = \varphi$.

3. If $\varphi \in C^2(\mathbb{R})$ is even and $\varphi' \in L_1[0,\infty)$, then $\mathcal{I}\mathcal{D}\varphi = \varphi$.

4. If $t \mapsto t^{s-1}\varphi(t) \in L_1[0,\infty)$ and $s \geq 3$, then $\mathcal{F}_s(\varphi) = \mathcal{F}_{s-2}(\mathcal{I}\varphi)$.

5. If $\varphi \in C^2(\mathbb{R})$ is even and $t \mapsto t^s\varphi'(t) \in L_1[0,\infty)$, then $\mathcal{F}_s(\varphi) = \mathcal{F}_{s+2}(\mathcal{D}\varphi)$.

The operators $\mathcal{I}$ and $\mathcal{D}$ allow us to express s-variate Fourier transforms as $(s-2)$- or $(s+2)$-variate Fourier transforms, respectively. In particular, a direct consequence of the above properties and the characterization of strictly positive definite radial functions (Theorem 2.4.1) is

Theorem 4.1.3 *1. Suppose $\varphi \in C(\mathbb{R})$. If $t \mapsto t^{s-1}\varphi(t) \in L_1[0,\infty)$ and $s \geq 3$, then φ is strictly positive definite and radial on $\mathbb{R}^s$ if and only if $\mathcal{I}\varphi$ is strictly positive definite and radial on $\mathbb{R}^{s-2}$.*

2. If $\varphi \in C^2(\mathbb{R})$ is even and $t \mapsto t^s\varphi'(t) \in L_1[0,\infty)$, then φ is strictly positive definite and radial on $\mathbb{R}^s$ if and only if $\mathcal{D}\varphi$ is strictly positive definite and radial on $\mathbb{R}^{s+2}$.

This allows us to construct new strictly positive definite radial functions from given ones by a "dimension-walk" technique that steps through multivariate Euclidean space in even increments.

4.2 Wendland's Compactly Supported Functions

In [627] Wendland constructed a popular family of compactly supported radial functions by starting with the truncated power function (which we know to be strictly positive definite and radial on $\mathbb{R}^s$ for $s \leq 2\ell - 1$), and then walking through dimensions by repeatedly applying the operator I.

Definition 4.2.1 *With $\varphi_\ell(r) = (1-r)_+^\ell$ we define*

$$\varphi_{s,k} = \mathcal{I}^k\varphi_{\lfloor s/2\rfloor+k+1}.$$

It turns out that the functions $\varphi_{s,k}$ are all supported on $[0,1]$ and have a polynomial representation there. More precisely,

Theorem 4.2.2 *The functions $\varphi_{s,k}$ are strictly positive definite and radial on $\mathbb{R}^s$ and are of the form*

$$\varphi_{s,k}(r) = \begin{cases} p_{s,k}(r), & r \in [0,1], \\ 0, & r > 1, \end{cases}$$

with a univariate polynomial $p_{s,k}$ of degree $\lfloor s/2 \rfloor + 3k + 1$. Moreover, $\varphi_{s,k} \in C^{2k}(\mathbb{R})$ are unique up to a constant factor, and the polynomial degree is minimal for given space dimension s and smoothness $2k$.

Wendland gave recursive formulas for the functions $\varphi_{s,k}$ for all s, k. We instead list the explicit formulas of [195]

Theorem 4.2.3 *The functions $\varphi_{s,k}$, $k = 0, 1, 2, 3$, have the form*

$$\begin{aligned}
\varphi_{s,0}(r) &= (1-r)_+^{\ell}, \\
\varphi_{s,1}(r) &\doteq (1-r)_+^{\ell+1} \left[(\ell+1)r + 1\right], \\
\varphi_{s,2}(r) &\doteq (1-r)_+^{\ell+2} \left[(\ell^2 + 4\ell + 3)r^2 + (3\ell + 6)r + 3\right], \\
\varphi_{s,3}(r) &\doteq (1-r)_+^{\ell+3} \left[(\ell^3 + 9\ell^2 + 23\ell + 15)r^3 + (6\ell^2 + 36\ell + 45)r^2 + (15\ell + 45)r + 15\right],
\end{aligned}$$

where $\ell = \lfloor s/2 \rfloor + k + 1$, and the symbol $\doteq$ denotes equality up to a multiplicative positive constant.

Proof: The case $k = 0$ follows directly from the definition. Application of the definition for the case $k = 1$ yields

$$\begin{aligned}
\varphi_{s,1}(r) &= (I\varphi_{\ell})(r) = \int_r^{\infty} t\varphi_{\ell}(t)dt \\
&= \int_r^{\infty} t(1-t)_+^{\ell} dt \\
&= \int_r^{1} t(1-t)^{\ell} dt \\
&= \frac{1}{(\ell+1)(\ell+2)}(1-r)^{\ell+1}\left[(\ell+1)r + 1\right],
\end{aligned}$$

where the compact support of φ_{ℓ} reduces the improper integral to a definite integral which can be evaluated using integration by parts. The other two cases are obtained similarly by repeated application of I. □

Examples: For $s = 3$ we get some of the most commonly used functions as

$$\begin{aligned}
\varphi_{3,0}(r) &= (1-r)_+^2, && \in C^0 \cap SPD(\mathbb{R}^3) \\
\varphi_{3,1}(r) &\doteq (1-r)_+^4 (4r+1), && \in C^2 \cap SPD(\mathbb{R}^3) \\
\varphi_{3,2}(r) &\doteq (1-r)_+^6 \left(35r^2 + 18r + 3\right), && \in C^4 \cap SPD(\mathbb{R}^3) \\
\varphi_{3,3}(r) &\doteq (1-r)_+^8 \left(32r^3 + 25r^2 + 8r + 1\right), && \in C^6 \cap SPD(\mathbb{R}^3).
\end{aligned}$$

4.3 Wu's Compactly Supported Functions

In [656] Wu presents another way to construct strictly positive definite radial functions with compact support. He starts with the function

$$\psi(r) = (1-r^2)_+^\ell, \qquad \ell \in \mathbb{N},$$

which is strictly positive definite and radial since we know that the truncated power function $\psi(\sqrt{\cdot})$ is multiply monotone. Wu then constructs another function that is strictly positive definite and radial on $\mathbb{R}$ by convolution, i.e.,

$$\begin{aligned}
\psi_\ell(r) &= (\psi * \psi)(2r) \\
&= \int_{-\infty}^{\infty} (1-t^2)_+^\ell (1-(2r-t)^2)_+^\ell dt \\
&= \int_{-1}^{1} (1-t^2)^\ell (1-(2r-t)^2)^\ell dt.
\end{aligned}$$

This function is strictly positive definite since its Fourier transform is essentially the square of the Fourier transform of ψ. Just like the Wendland functions, this function is a polynomial on its support. In fact, the degree of the polynomial is $4\ell+1$, and $\psi_\ell \in C^{2\ell}(\mathbb{R})$.

Now, a family of strictly positive definite radial functions is constructed by a dimension walk using the $\mathcal{D}$ operator, i.e.,

$$\psi_{k,\ell} = \mathcal{D}^k \psi_\ell.$$

The functions $\psi_{k,\ell}$ are strictly positive definite and radial in $\mathbb{R}^s$ for $s \le 2k+1$, are polynomials of degree $4\ell - 2k + 1$ on their support and in $C^{2(\ell-k)}$ in the interior of the support. On the boundary the smoothness increases to $C^{2\ell-k}$.

Example: For $\ell = 3$ we can compute the three functions

$$\psi_{k,3}(r) = \mathcal{D}^k \psi_3(r) = \mathcal{D}^k((1-\cdot^2)_+^3 * (1-\cdot^2)_+^3)(2r), \qquad k = 0,1,2,3.$$

This results in

$$\begin{aligned}
\psi_{0,3}(r) &\doteq \left(5 - 39r^2 + 143r^4 - 429r^6 + 429r^7 - 143r^9 + 39r^{11} - 5r^{13}\right)_+ \\
&\doteq (1-r)_+^7 (5 + 35r + 101r^2 + 147r^3 + 101r^4 + 35r^5 + 5r^6) \qquad \in C^6 \cap SPD(\mathbb{R}) \\
\psi_{1,3}(r) &\doteq \left(6 - 44r^2 + 198r^4 - 231r^5 + 99r^7 - 33r^9 + 5r^{11}\right)_+ \\
&\doteq (1-r)_+^6 (6 + 36r + 82r^2 + 72r^3 + 30r^4 + 5r^5) \qquad \in C^4 \cap SPD(\mathbb{R}^3) \\
\psi_{2,3}(r) &\doteq \left(8 - 72r^2 + 105r^3 - 63r^5 + 27r^7 - 5r^9\right)_+ \\
&\doteq (1-r)_+^5 (8 + 40r + 48r^2 + 25r^3 + 5r^4) \qquad \in C^2 \cap SPD(\mathbb{R}^5) \\
\psi_{3,3}(r) &\doteq \left(16 - 35r + 35r^3 - 21r^5 + 5r^7\right)_+ \\
&\doteq (1-r)_+^4 (16 + 29r + 20r^2 + 5r^3) \qquad \in C^0 \cap SPD(\mathbb{R}^7).
\end{aligned}$$

Remarks:

1. For a prescribed smoothness the polynomial degree of Wendland's functions is lower than that of Wu's functions. For example, both Wendland's function $\varphi_{3,2}$ and Wu's function $\psi_{1,3}$ are C^4 smooth and strictly positive definite and radial in $\mathbb{R}^3$. However, the polynomial degree of Wendland's function is 8, whereas that of Wu's function is 11.

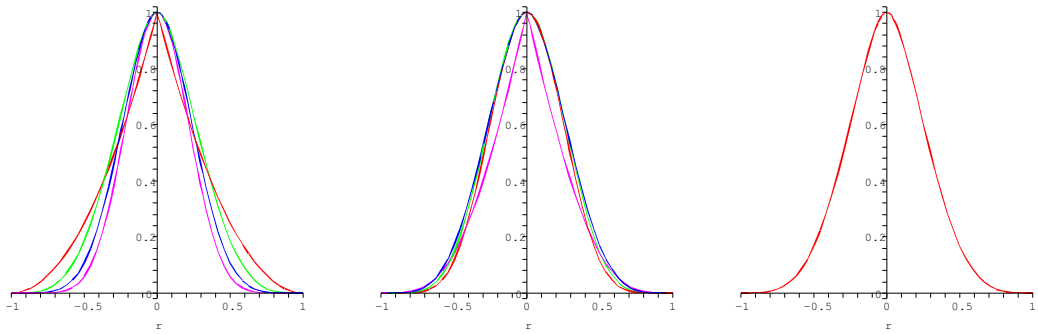

Figure 4.1: Plot of Wendland's functions (left), Wu's functions (center), and Buhmann's function (right) listed as examples.

2. While both families of strictly positive definite compactly supported functions are constructed via dimension walk, Wendland uses integration (and thus obtains a family of increasingly smoother functions), whereas Wu needs to start with a function of sufficient smoothness, and then obtains successively less smooth functions (via differentiation).

4.4 Buhmann's Compactly Supported Functions

A third family of compactly supported strictly positive definite radial functions that has appeared in the literature is due to Buhmann (see [84]). Buhmann's functions contain a logarithmic term in addition to a polynomial. His functions have the general form

$$\phi(r) = \int_0^\infty (1 - r^2/t)_+^\lambda t^\alpha (1 - t^\delta)_+^\rho dt.$$

Here $0 < \delta \leq \frac{1}{2}$, $\rho \geq 1$, and in order to obtain functions that are strictly positive definite and radial on $\mathbb{R}^s$ for $s \leq 3$ the constraints for the remaining parameters are $\lambda \geq 0$, and $-1 < \alpha \leq \frac{\lambda-1}{2}$.

Example: An example with $\alpha = \delta = \frac{1}{2}$, $\rho = 1$ and $\lambda = 2$ is listed in [85]:

$$\phi(r) \doteq 12r^4 \log r - 21r^4 + 32r^3 - 12r^2 + 1, \qquad 0 \leq r \leq 1, \ \in C^2 \cap SPD(\mathbb{R}^3).$$

Remarks:

1. While Buhmann [85] claims that his construction encompasses both Wendland's and Wu's functions, Wendland [634] gives an even more general theorem that shows that integration of a positive function $f \in L_1[0, \infty)$ against a strictly positive definite (compactly supported) kernel K results in a (compactly supported) strictly positive definite function, i.e.,

$$\varphi(r) = \int_0^\infty K(t, r) f(t) dt$$

is strictly positive definite. Buhmann's construction then corresponds to choosing $f(t) = t^{\alpha}(1-t^{\delta})_{+}^{\rho}$ and $K(t,r) = (1-r^2/t)_{+}^{\lambda}$.

2. Multiply monotone functions are covered by this general theorem by taking $K(t,r) = (1-rt)_{+}^{k-1}$ and f an arbitrary positive function in L_1 so that $d\mu(t) = f(t)dt$ in Williamson's characterization Theorem 2.6.2. Also, functions that are strictly positive definite and radial in $\mathbb{R}^s$ for all s (or equivalently completely monotone functions) are covered by choosing $K(t,r) = e^{-rt}$.

Chapter 5

Error Bounds and the Variational Approach

In order to estimate the approximation properties of the functions studied thus far we will introduce the variational approach to scattered data interpolation. This approach was used first for radial basis function interpolation by Madych and Nelson [417], and later adopted by many others (see, e.g., [393, 394], [518], [545], [628, 629], [658]). We will see that for every strictly positive definite radial function there is an associated Hilbert space in which the radial basis function interpolant provides the best approximation to a given function. This optimality of interpolants in Hilbert space is the subject of the theory of optimal recovery described in the late 1950s by Golomb and Weinberger in their paper [264]. The following discussion follows mostly the presentation in Wendland's book [634].

5.1 Reproducing Kernel Hilbert Spaces

We begin with

Definition 5.1.1 *Let $\mathcal{H}$ be a real Hilbert space of functions $f : \Omega \to \mathbb{R}$. A function $K : \Omega \times \Omega \to \mathbb{R}$ is called* reproducing kernel for $\mathcal{H}$ *if*

1. *$K(\boldsymbol{x}, \cdot) \in \mathcal{H}$ for all $\boldsymbol{x} \in \Omega$,*
2. *$f(\boldsymbol{x}) = \langle f, K(\cdot, \boldsymbol{x}) \rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$ and all $\boldsymbol{x} \in \Omega$.*

It is known that the reproducing kernel of a Hilbert space is unique, and that existence of a reproducing kernel is equivalent to the fact that the point evaluation functionals $\delta_{\boldsymbol{x}}$ are bounded linear functionals, i.e., there exists a positive constant $M = M_{\boldsymbol{x}}$ such that

$$|\delta_{\boldsymbol{x}} f| = |f(\boldsymbol{x})| \leq M \|f\|_{\mathcal{H}}$$

for all $f \in \mathcal{H}$. This latter fact is due to the Riesz Representation Theorem.

Other properties of reproducing kernels are

Theorem 5.1.2 *Suppose $\mathcal{H}$ is a Hilbert space of functions $f : \Omega \to \mathbb{R}$ with reproducing kernel K and $\mathcal{H}^*$ its dual space, i.e., the space of linear functionals on $\mathcal{H}$. Then we have*

1. $K(\boldsymbol{x}, \boldsymbol{y}) = \langle K(\boldsymbol{x}, \cdot), K(\cdot, \boldsymbol{y}) \rangle_{\mathcal{H}}$ *for* $\boldsymbol{x}, \boldsymbol{y} \in \Omega$.

2. $K(\boldsymbol{x}, \boldsymbol{y}) = K(\boldsymbol{y}, \boldsymbol{x})$ *for* $\boldsymbol{x}, \boldsymbol{y} \in \Omega$.

3. *Convergence in Hilbert space norm implies pointwise convergence.*

Proof: Since by (1) of Definition 5.1.1 $K(\boldsymbol{x}, \cdot) \in \mathcal{H}$ for every $\boldsymbol{x} \in \Omega$, the reproducing property (2) of the definition gives us

$$K(\boldsymbol{x}, \boldsymbol{y}) = \langle K(\boldsymbol{x}, \cdot), K(\cdot, \boldsymbol{y}) \rangle_{\mathcal{H}}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$. (2) follows from (1) by the symmetry of the Hilbert space inner product. For (3) we use the reproducing property of K along with the Cauchy-Schwarz inequality:

$$|f(\boldsymbol{x}) - f_n(\boldsymbol{x})| = |\langle f - f_n, K(\cdot, \boldsymbol{x}) \rangle_{\mathcal{H}}| \leq \|f - f_n\|_{\mathcal{H}} \|K(\cdot, \boldsymbol{x})\|_{\mathcal{H}}.$$

□

Moreover, the reproducing kernel K is known to be positive definite. In the following we use a slight generalization of the notion of a positive definite function to a positive definite kernel. Essentially, we replace $\Phi(\boldsymbol{x}_j - \boldsymbol{x}_k)$ in Definition 1.2.5 by $K(\boldsymbol{x}_j, \boldsymbol{x}_k)$.

Theorem 5.1.3 *Suppose $\mathcal{H}$ is a reproducing kernel Hilbert function space with reproducing kernel $K : \Omega \times \Omega \to \mathbb{R}$. Then K is positive definite. Moreover, K is strictly positive definite if and only if the point evaluation functionals are linearly independent in $\mathcal{H}^*$.*

Proof: Since the kernel is real-valued we can restrict ourselves to a quadratic form with real coefficients. For distinct points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ and nonzero $\boldsymbol{c} \in \mathbb{R}^N$ we have

$$\begin{aligned}
\sum_{j=1}^{N} \sum_{k=1}^{N} c_j c_k K(\boldsymbol{x}_j, \boldsymbol{x}_k) &= \sum_{j=1}^{N} \sum_{k=1}^{N} c_j c_k \langle K(\boldsymbol{x}_j, \cdot), K(\cdot, \boldsymbol{x}_k) \rangle_{\mathcal{H}} \\
&= \langle \sum_{j=1}^{N} c_j K(\boldsymbol{x}_j, \cdot), \sum_{k=1}^{n} c_k K(\cdot, \boldsymbol{x}_k) \rangle_{\mathcal{H}} \\
&= \| \sum_{j=1}^{N} c_j K(\boldsymbol{x}_j, \cdot) \|_{\mathcal{H}}^2 \geq 0.
\end{aligned}$$

To establish the second claim we assume K is not strictly positive definite and show that the point evaluation functionals must be linearly dependent. If K is not strictly positive definite then there exist distinct points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ and nonzero coefficients such that

$$\sum_{j=1}^{N} \sum_{k=1}^{N} c_j c_k K(\boldsymbol{x}_j, \boldsymbol{x}_k) = 0.$$

The first part of the proof therefore implies

$$\sum_{j=1}^{N} c_j K(\boldsymbol{x}_j, \cdot) = \sum_{j=1}^{N} c_j K(\cdot, \boldsymbol{x}_j) = 0.$$

Taking the Hilbert space inner product with an arbitrary $f \in \mathcal{H}$ and using the reproducing property of K we get that

$$\begin{aligned} 0 &= \langle f, \sum_{j=1}^{N} c_j K(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{H}} \\ &= \sum_{j=1}^{N} c_j \langle f, K(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{H}} \\ &= \sum_{j=1}^{N} c_j f(\boldsymbol{x}_j) \\ &= \sum_{j=1}^{N} c_j \delta_{\boldsymbol{x}_j}(f). \end{aligned}$$

This, however, shows the linear dependence of the point evaluation functionals $\delta_{\boldsymbol{x}}(f) = f(\boldsymbol{x})$. An analogous argument can be used to show the converse. □

This theorem provides a connection between strictly positive definite functions and reproducing kernels. Our interest, however, lies in the other direction. Since we are starting with strictly positive definite functions, we need to show how to construct an associated reproducing kernel Hilbert space.

5.2 Native Spaces for Strictly Positive Definite Functions

First, we note that Definition 5.1.1 tells us that $\mathcal{H}$ contains all functions of the form

$$f = \sum_{j=1}^{N} c_j K(\boldsymbol{x}_j, \cdot)$$

provided $\boldsymbol{x}_j \in \Omega$. In Theorem 5.1.2 we showed that

$$\begin{aligned} \|f\|_{\mathcal{H}}^2 &= \langle f, f \rangle_{\mathcal{H}} = \langle \sum_{j=1}^{N} c_j K(\boldsymbol{x}_j, \cdot), \sum_{k=1}^{N} c_k K(\cdot, \boldsymbol{x}_k) \rangle_{\mathcal{H}} \\ &= \sum_{j=1}^{N} \sum_{k=1}^{N} c_j c_k \langle K(\boldsymbol{x}_j, \cdot), K(\cdot, \boldsymbol{x}_k) \rangle_{\mathcal{H}} \\ &= \sum_{j=1}^{N} \sum_{k=1}^{N} c_j c_k K(\boldsymbol{x}_j, \boldsymbol{x}_k). \end{aligned}$$

Therefore, we *define* the space

$$H_K(\Omega) = \text{span}\{K(\cdot, \boldsymbol{y}) : \boldsymbol{y} \in \Omega\}$$

with an associated bilinear form

$$\langle \sum_{j=1}^{N} c_j K(\cdot, \boldsymbol{x}_j), \sum_{k=1}^{N} d_k K(\cdot, \boldsymbol{y}_k) \rangle_K = \sum_{j=1}^{N} \sum_{k=1}^{N} c_j d_k K(\boldsymbol{x}_j, \boldsymbol{y}_k).$$

Theorem 5.2.1 *If $K : \Omega \times \Omega \to \mathbb{R}$ is a symmetric strictly positive definite kernel, then the bilinear form $\langle \cdot, \cdot \rangle_K$ defines an inner product on $H_K(\Omega)$. Furthermore, $H_K(\Omega)$ is a pre-Hilbert space with reproducing kernel K.*

Proof: $\langle \cdot, \cdot \rangle_K$ is obviously bilinear and symmetric. We just need to show that $\langle f, f \rangle_K > 0$ for nonzero $f \in H_K(\Omega)$. Any such f can be written in the form

$$f = \sum_{j=1}^{N} c_j K(\cdot, \boldsymbol{x}_j), \qquad \boldsymbol{x}_j \in \Omega.$$

Then

$$\langle f, f \rangle_K = \sum_{j=1}^{N} \sum_{k=1}^{N} c_j c_k K(\boldsymbol{x}_j, \boldsymbol{x}_k) > 0$$

since K is strictly positive definite. The reproducing property follows from

$$\langle f, K(\cdot, \boldsymbol{x}) \rangle_K = \sum_{j=1}^{N} c_j K(\boldsymbol{x}_j, \boldsymbol{x}) = f(\boldsymbol{x}).$$

□

The *native space* $\mathcal{N}_K(\Omega)$ of K is now defined to be the completion of $H_K(\Omega)$ with respect to the K-norm $\| \cdot \|_K$ so that $\|f\|_K = \|f\|_{\mathcal{N}_K(\Omega)}$ for all $f \in H_K(\Omega)$. The technical details concerned with this construction are discussed in [634].

In the special case when we are dealing with strictly positive definite (translation invariant) functions $\Phi(\boldsymbol{x} - \boldsymbol{y}) = K(\boldsymbol{x}, \boldsymbol{y})$ and when $\Omega = \mathbb{R}^s$ we get a characterization of native spaces in terms of Fourier transforms.

Theorem 5.2.2 *Suppose $\Phi \in C(\mathbb{R}^s) \cap L_1(\mathbb{R}^s)$ is a real-valued strictly positive definite function. Define*

$$\mathcal{G} = \{f \in L_2(\mathbb{R}^s) \cap C(\mathbb{R}^s) : \frac{\hat{f}}{\sqrt{\hat{\Phi}}} \in L_2(\mathbb{R}^s)\}$$

and equip this space with the bilinear form

$$\langle f, g \rangle_{\mathcal{G}} = \frac{1}{\sqrt{(2\pi)^s}} \langle \frac{\hat{f}}{\sqrt{\hat{\Phi}}}, \frac{\hat{g}}{\sqrt{\hat{\Phi}}} \rangle_{L_2(\mathbb{R}^s)} = \frac{1}{\sqrt{(2\pi)^s}} \int_{\mathbb{R}^s} \frac{\hat{f}(\boldsymbol{\omega}) \overline{\hat{g}(\boldsymbol{\omega})}}{\hat{\Phi}(\boldsymbol{\omega})} d\boldsymbol{\omega}.$$

Then $\mathcal{G}$ is a real Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{G}}$ and reproducing kernel $\Phi(\cdot - \cdot)$. Hence, $\mathcal{G}$ is the native space of Φ on $\mathbb{R}^s$, i.e., $\mathcal{G} = \mathcal{N}_\Phi(\mathbb{R}^s)$ and both inner product coincide. In particular, every $f \in \mathcal{N}_\Phi(\mathbb{R}^s)$ can be recovered from its Fourier transform $\hat{f} \in L_1(\mathbb{R}^s) \cap L_2(\mathbb{R}^s)$.

Remarks:

1. This theorem shows that the native spaces can be viewed as a generalization of the standard Sobolev spaces. Indeed, for $m > s/2$ the Sobolev space W_2^m can be defined as

$$W_2^m(\mathbb{R}^s) = \{f \in L_2(\mathbb{R}^s) \cap C(\mathbb{R}^s) : \hat{f}(\cdot)(1 + \| \cdot \|_2^2)^{m/2} \in L_2(\mathbb{R}^s)\}.$$

Therefore, any strictly positive definite function Φ whose Fourier transform decays only algebraically has a Sobolev space as its native space. In particular, the compactly supported Wendland functions $\Phi_{s,k} = \varphi_{s,k}(\|\cdot\|_2)$ of Chapter 4 can be shown to have native spaces $\mathcal{N}_{\Phi_{s,k}}(\mathbb{R}^s) = W_2^{s/2+k+1/2}(\mathbb{R}^s)$ (where the restriction $s \geq 3$ is required for the special case $k = 0$).

2. Native spaces for strictly conditionally positive definite functions can also be constructed. However, since this is more technical, we limit the discussion here to strictly positive definite functions, and refer the interested reader to the book by Wendland [634] or the papers [554, 555] by Schaback.

3. The native spaces of the powers and thin plate (or surface) splines of Examples 2 and 3 of Sections 3.3 and 3.4 can be shown to be the so-called Beppo-Levi spaces of order k

$$BL_k(\mathbb{R}^s) = \{f \in C(\mathbb{R}^s) : D^{\boldsymbol{\alpha}} f \in L_2(\mathbb{R}^s) \text{ for all } |\boldsymbol{\alpha}| = k,\ \boldsymbol{\alpha} \in \mathbb{N}^s\},$$

where $D^{\boldsymbol{\alpha}}$ denotes a *generalized derivative* of order $\boldsymbol{\alpha}$ (defined in the same spirit as the generalized Fourier transform). In fact, the intersection of all Beppo-Levi spaces $BL_k(\mathbb{R}^s)$ of order $k \leq m$ yields the Sobolev space $W_2^m(\mathbb{R}^s)$. For more details see [634]. These spaces were already studied in the early papers by Duchon [168, 169, 170, 171].

4. The native spaces for Gaussians and (inverse) multiquadrics are rather small. For example, according to Theorem 5.2.2, for Gaussians the Fourier transform of $f \in \mathcal{N}_\Phi(\Omega)$ must decay faster than the Fourier transform of the Gaussian (which is itself a Gaussian). It is known that, however, even though the native space of Gaussians is small, it does contain the so-called *band-limited functions*, i.e., functions whose Fourier transform is compactly supported. These functions play an important role in *sampling theory* where Shannon's famous Sampling Theorem [575] states that any band-limited function can be completely recovered from its discrete samples provided the function is sampled at a sampling rate at least twice its bandwidth. The content of this theorem was already known to Whitaker [640] in 1915.

> **Theorem 5.2.3** *Suppose $f \in C(\mathbb{R}^s) \cap L_1(\mathbb{R}^s)$ such that its Fourier transform vanishes outside the cube $Q = \left[-\frac{1}{2}, \frac{1}{2}\right]^s$. Then f can be uniquely reconstructed from its values on $\mathbb{Z}^s$, i.e.,*
>
> $$f(\boldsymbol{x}) = \sum_{\boldsymbol{\xi} \in \mathbb{Z}^s} f(\boldsymbol{\xi}) \mathrm{sinc}(\boldsymbol{x} - \boldsymbol{\xi}), \qquad \boldsymbol{x} \in \mathbb{R}^s .$$

Here the sinc function is defined for any $\boldsymbol{x} = (x_1, \ldots, x_s) \in \mathbb{R}^s$ as $\mathrm{sinc}\, \boldsymbol{x} = \prod_{i=1}^s \frac{\sin(\pi x_i)}{\pi x_i}$. For more details on Shannon's Sampling Theorem see, e.g., Chapter 29 in the book [132] by Cheney and Light or the paper [610] by Unser.

5.3 The Power Function and Error Estimates for Functions in $\mathcal{N}_\Phi(\Omega)$

Our goal in this section is to provide error estimates for scattered data interpolation with strictly (conditionally) positive definite functions. In their final form these estimates will need to involve some kind of measure of the data distribution. The measure that is usually used is the so-called *fill distance*

$$h = h_{\mathcal{X},\Omega} = \sup_{\boldsymbol{x}\in\Omega} \min_{\boldsymbol{x}_j\in\mathcal{X}} \|\boldsymbol{x} - \boldsymbol{x}_j\|_2$$

which indicates how well the data fill out the domain Ω. The fill distance denotes the radius of the largest possible empty ball that can be placed among the data locations. We will be interested in whether the error

$$\|f - \mathcal{P}_h f\|_\infty$$

tends to zero as $h \to 0$, and if so, how fast. Here $\{\mathcal{P}_h\}_h$ presents a sequence of interpolation (or, more generally, projection) operators that vary with the fill distance h. For example, $\mathcal{P}_h$ could denote interpolation to data given at $(2^n+1)^s$, $n = 1, 2, \ldots$, equally spaced points in the unit cube in $\mathbb{R}^s$ (with $h = 2^{-n}$). Of course, the definition of the fill distance allows for scattered data as well.

Since we want to use the machinery of reproducing kernel Hilbert spaces we will concentrate on error estimates for functions $f \in \mathcal{N}_\Phi$. In the next section we will also mention some more general estimates.

The term that is often used to measure the speed of convergence to zero is *approximation order*. We say that the approximation operator $\mathcal{P}_h$ has L_p-approximation order k if

$$\|f - \mathcal{P}_h f\|_p = \mathcal{O}(h^k) \quad \text{for } h \to 0.$$

Moreover, if we can also show that $\|f - \mathcal{P}_h f\|_p \neq o(h^k)$, then $\mathcal{P}_h$ has *exact* L_p-approximation order k. We will concentrate mostly on the case $p = \infty$, but approximation order in other norms can also be studied.

In order to keep the following discussion as transparent as possible we will restrict ourselves to strictly positive definite functions. With (considerably) more technical details the following can also be formulated for strictly conditionally positive definite functions (see [634] for details).

The key idea for the following discussion is to express the interpolant in Lagrange form, i.e., using *cardinal basis functions*. This idea is due to Schaback and Wu [658]. In the previous chapters we have established that, for any strictly positive definite function Φ, the linear system

$$A\boldsymbol{c} = \boldsymbol{y}$$

with $A_{ij} = \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j)$, $i, j = 1, \ldots, N$, $\boldsymbol{c} = [c_1, \ldots, c_N]^T$, and $\boldsymbol{y} = [f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)]^T$ has a unique solution. In the following we will consider the more general situation where Φ is a strictly positive definite kernel, i.e., the entries of A are given by $A_{ij} = \Phi(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

In order to obtain the cardinal basis functions u_j^*, $j = 1, \ldots, N$, with the property $u_j^*(\boldsymbol{x}_i) = \delta_{ij}$ we consider the linear system

$$A\boldsymbol{u}^*(\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x}),$$

where the matrix A is as above (and therefore invertible), $\boldsymbol{u}^* = [u_1^*, \ldots, u_N^*]^T$, and $\boldsymbol{b} = [\Phi(\cdot, \boldsymbol{x}_1), \ldots, \Phi(\cdot, \boldsymbol{x}_N)]^T$. Thus,

Theorem 5.3.1 *Suppose Φ is a strictly positive definite kernel on $\mathbb{R}^s$. Then, for any distinct points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, there exist functions $u_j^* \in \text{span}\{\Phi(\cdot, \boldsymbol{x}_j), j = 1, \ldots, N\}$ such that $u_j^*(\boldsymbol{x}_i) = \delta_{ij}$.*

Therefore, we can write $\mathcal{P}f$ in the cardinal form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{N} f(\boldsymbol{x}_j) u_j^*(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s .$$

Another important ingredient in our estimates is the so-called *power function*. To this end, we consider a domain $\Omega \subseteq \mathbb{R}^s$. Then for any strictly positive definite kernel $\Phi \in C(\Omega \times \Omega)$, any set of distinct points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega$, and any vector $\boldsymbol{u} \in \mathbb{R}^N$, we define the quadratic form

$$\begin{aligned}
Q(\boldsymbol{u}) &= \Phi(\boldsymbol{x}, \boldsymbol{x}) - 2\sum_{j=1}^{N} u_j \Phi(\boldsymbol{x}, \boldsymbol{x}_j) + \sum_{i=1}^{N}\sum_{j=1}^{N} u_i u_j \Phi(\boldsymbol{x}_i, \boldsymbol{x}_j) \\
&= \langle \Phi(\cdot, \boldsymbol{x}), \Phi(\cdot, \boldsymbol{x}) \rangle_{\mathcal{N}_\Phi(\Omega)} - 2\sum_{j=1}^{N} u_j \langle \Phi(\cdot, \boldsymbol{x}), \Phi(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&\quad + \sum_{i=1}^{N}\sum_{j=1}^{N} u_i u_j \langle \Phi(\cdot, \boldsymbol{x}_i), \Phi(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&= \langle \Phi(\cdot, \boldsymbol{x}) - \sum_{j=1}^{N} u_j \Phi(\cdot, \boldsymbol{x}_j), \Phi(\cdot, \boldsymbol{x}) - \sum_{j=1}^{N} u_j \Phi(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&= \left\| \Phi(\cdot, \boldsymbol{x}) - \sum_{j=1}^{N} u_j \Phi(\cdot, \boldsymbol{x}_j) \right\|^2_{\mathcal{N}_\Phi(\Omega)} . \qquad (5.1)
\end{aligned}$$

Here we have used the definition of the native space norm from the previous section.

Then

Definition 5.3.2 *Suppose $\Omega \subseteq \mathbb{R}^s$ and $\Phi \in C(\Omega \times \Omega)$ is strictly positive definite on $\mathbb{R}^s$. For any distinct points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega$ the* power function *is defined by*

$$[P_{\Phi,\mathcal{X}}(\boldsymbol{x})]^2 = Q(\boldsymbol{u}^*(\boldsymbol{x})),$$

where $\boldsymbol{u}^*$ is the vector of cardinal functions from Theorem 5.3.1.

Remarks:

1. The name *power function* was chosen by Schaback based on its connection to the power function of a statistical decision function [622].

2. In the paper [658] by Wu and Schaback the power function was referred to as *kriging function*. This terminology comes from geostatistics (see, e.g., [474]).

Now we can give a first generic error estimate.

Theorem 5.3.3 *Let $\Omega \subseteq \mathbb{R}^s$, $\Phi \in C(\Omega \times \Omega)$ be strictly positive definite on $\mathbb{R}^s$, and suppose that the points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ are distinct. Denote the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on $\mathcal{X}$ by $\mathcal{P}f$. Then for every $\boldsymbol{x} \in \Omega$*

$$|f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| \leq P_{\Phi,\mathcal{X}}(\boldsymbol{x}) \|f\|_{\mathcal{N}_\Phi(\Omega)}.$$

Proof: We express the interpolant in its cardinal form and apply the reproducing property of Φ. This gives us

$$\begin{aligned} \mathcal{P}f(\boldsymbol{x}) &= \sum_{j=1}^{N} f(\boldsymbol{x}_j) u_j^*(\boldsymbol{x}) \\ &= \sum_{j=1}^{N} u_j^*(\boldsymbol{x}) \langle f, \Phi(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \langle f, \sum_{j=1}^{N} u_j^*(\boldsymbol{x}) \Phi(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)}. \end{aligned}$$

For f the reproducing property of Φ yields

$$f(\boldsymbol{x}) = \langle f, \Phi(\cdot, \boldsymbol{x}) \rangle_{\mathcal{N}_\Phi(\Omega)}.$$

Now we combine these two formulas and apply the Cauchy-Schwarz inequality

$$\begin{aligned} |f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| &= \left| \langle f, \Phi(\cdot, \boldsymbol{x}) - \sum_{j=1}^{N} u_j^*(\boldsymbol{x}) \Phi(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \right| \\ &\leq \|f\|_{\mathcal{N}_\Phi(\Omega)} \left\| \Phi(\cdot, \boldsymbol{x}) - \sum_{j=1}^{N} u_j^*(\boldsymbol{x}) \Phi(\cdot, \boldsymbol{x}_j) \right\|_{\mathcal{N}_\Phi(\Omega)} \\ &= \|f\|_{\mathcal{N}_\Phi(\Omega)} P_{\Phi,\mathcal{X}}(\boldsymbol{x}), \end{aligned}$$

where we have applied (5.1) and the definition of the power function. □

Remark: One of the main benefits of Theorem 5.3.3 is that we are now able to estimate the interpolation error by considering two independent phenomena:

- the smoothness of the data (measured in terms of the native space norm of f – which is independent of the data locations),
- and the contribution due to the use of the basic function Φ and the distribution of the data (measured in terms of the power function – independent of the actual data values).

This is analogous to the standard error estimate for polynomial interpolation cited in most numerical analysis texts.

The next step is to refine this error bound by expressing the influence of the data locations in terms of the fill distance. And then, of course, the bound needs to be specialized to various choices of basic functions Φ.

The strategy to obtaining most error bounds in numerical analysis is to take advantage of the polynomial precision of a method (at least locally), and then to apply a Taylor expansion. With this in mind we observe

Theorem 5.3.4 *Let $\Omega \subseteq \mathbb{R}^s$, and suppose $\Phi \in C(\Omega \times \Omega)$ is strictly positive definite on $\mathbb{R}^s$. Let $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ be a set of distinct points in Ω, and define the quadratic form Q as in (5.1). The minimum of Q is given by the vector $\boldsymbol{u}^*(\boldsymbol{x})$ from Theorem 5.3.1, i.e.,*

$$Q(\boldsymbol{u}^*(\boldsymbol{x})) \leq Q(\boldsymbol{u}) \qquad \text{for all } \boldsymbol{u} \in \mathbb{R}^N .$$

Proof: Using the linear system notation employed earlier, we note that

$$Q(\boldsymbol{u}) = \Phi(\boldsymbol{x}, \boldsymbol{x}) - 2\boldsymbol{u}^T \boldsymbol{b}(\boldsymbol{x}) + \boldsymbol{u}^T A \boldsymbol{u}.$$

The minimum of this quadratic form is given by the solution of the linear system

$$A\boldsymbol{u} = \boldsymbol{b}(\boldsymbol{x}).$$

This, however, yields the cardinal functions $\boldsymbol{u} = \boldsymbol{u}^*(\boldsymbol{x})$. □

Remark: The arguments used in the previous proof suggest two alternative representations of the power function. Using the matrix-vector notation, the power function is given as

$$P_{\Phi,\mathcal{X}}(\boldsymbol{x}) = \sqrt{Q(\boldsymbol{u}^*(\boldsymbol{x}))} = \sqrt{\Phi(\boldsymbol{x}, \boldsymbol{x}) - 2(\boldsymbol{u}^*(\boldsymbol{x}))^T \boldsymbol{b}(\boldsymbol{x}) + (\boldsymbol{u}^*(\boldsymbol{x}))^T A \boldsymbol{u}^*(\boldsymbol{x})}.$$

However, by the definition of the cardinal functions $A\boldsymbol{u}^*(\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x})$, and therefore we have the two new variants

$$\begin{aligned} P_{\Phi,\mathcal{X}}(\boldsymbol{x}) &= \sqrt{\Phi(\boldsymbol{x}, \boldsymbol{x}) - (\boldsymbol{u}^*(\boldsymbol{x}))^T \boldsymbol{b}(\boldsymbol{x})} \\ &= \sqrt{\Phi(\boldsymbol{x}, \boldsymbol{x}) - (\boldsymbol{u}^*(\boldsymbol{x}))^T A \boldsymbol{u}^*(\boldsymbol{x})}. \end{aligned}$$

In the proof below we will use a special coefficient vector $\tilde{\boldsymbol{u}}$ which provides the polynomial precision desired for the proof of the refined error estimate. Its existence is guaranteed by the following theorem on local polynomial reproduction proved by Wendland in [634]. This theorem requires the notion of a domain which satisfies an interior cone condition.

Definition 5.3.5 *A region $\Omega \subseteq \mathbb{R}^s$ satisfies an* interior cone condition *if there exists an angle $\theta \in (0, \pi/2)$ and a radius $r > 0$ such that for every $\boldsymbol{x} \in \Omega$ there exists a unit vector $\boldsymbol{\xi}(\boldsymbol{x})$ such that the cone*

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

is contained in Ω.

Remark: A consequence of the interior cone condition is the fact that a domain that satisfies this condition contains balls of a controllable radius. In particular, this will be important when bounding the remainder of the Taylor expansions below. For more details see [634].

Existence of an approximation scheme with local polynomial precision is guaranteed by

Theorem 5.3.6 *Suppose $\Omega \subseteq \mathbb{R}^s$ is bounded and satisfies an interior cone condition, and let ℓ be a non-negative integer. Then there exist positive constants h_0, sc_1, and c_2 such that for all $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega$ with $h_{\mathcal{X},\Omega} \leq h_0$ and every $\boldsymbol{x} \in \Omega$ there exist numbers $\tilde{u}_1(\boldsymbol{x}), \ldots \tilde{u}_N(\boldsymbol{x})$ with*

1. $\displaystyle\sum_{j=1}^{N} \tilde{u}_j(\boldsymbol{x}) p(\boldsymbol{x}_j) = p(\boldsymbol{x})$ *for all* $p \in \Pi_\ell^s$,
2. $\displaystyle\sum_{j=1}^{N} |\tilde{u}_j(\boldsymbol{x})| \leq c_1$,
3. $\tilde{u}_j(\boldsymbol{x}) = 0$ *if* $\|\boldsymbol{x} - \boldsymbol{x}_j\|_2 \geq c_2 h_{\mathcal{X},\Omega}$.

Remark: Property (1) yields the polynomial precision, and property (3) shows that the scheme is local. The bound in property (2) is essential for controlling the growth of error estimates and the quantity on the left-hand side of (2) is known as the *Lebesgue constant* at $\boldsymbol{x}$.

The error estimate can now be formulated in terms of the fill distance.

Theorem 5.3.7 *Suppose $\Omega \subseteq \mathbb{R}^s$ is bounded and satisfies an interior cone condition. Suppose $\Phi \in C^{2k}(\Omega \times \Omega)$ is symmetric and strictly positive definite. Denote the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on the set $\mathcal{X}$ by $\mathcal{P}f$. Then there exist positive constant h_0 and C (independent of $\boldsymbol{x}$, f and Φ) such that*

$$|f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| \leq C C_\Phi(\boldsymbol{x})^{1/2} h_{\mathcal{X},\Omega}^k \|f\|_{\mathcal{N}_\Phi(\Omega)},$$

provided $h_{\mathcal{X},\Omega} \leq h_0$. Here

$$C_\Phi(\boldsymbol{x}) = \max_{\boldsymbol{w},\boldsymbol{z} \in \Omega \cap B(\boldsymbol{x}, c_2 h_{\mathcal{X},\Omega})} |\Phi(\boldsymbol{w}, \boldsymbol{z})|.$$

Proof: By Theorem 5.3.3 we know

$$|f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| \leq P_{\Phi,\mathcal{X}}(\boldsymbol{x}) \|f\|_{\mathcal{N}_\Phi(\Omega)}.$$

Therefore, we now bound the power function in terms of the fill distance. We know that the power function is defined by

$$[P_{\Phi,\mathcal{X}}(\boldsymbol{x})]^2 = Q(\boldsymbol{u}^*(\boldsymbol{x})).$$

Moreover, we know by Theorem 5.3.4 that the quadratic form $Q(\boldsymbol{u})$ is minimized by $\boldsymbol{u} = \boldsymbol{u}^*(\boldsymbol{x})$. Therefore, any other coefficient vector $\boldsymbol{u}$ will yield an upper bound on the power function. We take $\boldsymbol{u} = \tilde{\boldsymbol{u}}(\boldsymbol{x})$ from Theorem 5.3.6 with polynomial precision of degree $\ell \geq 2k-1$.

We will make repeated use of the multivariate Taylor expansion

$$\Phi(\boldsymbol{w}, \boldsymbol{z}) = \sum_{|\boldsymbol{\beta}|<2k} \frac{D_2^{\boldsymbol{\beta}}\Phi(\boldsymbol{w}, \boldsymbol{w})}{\boldsymbol{\beta}!}(\boldsymbol{z}-\boldsymbol{w})^{\boldsymbol{\beta}} + R(\boldsymbol{w}, \boldsymbol{z})$$

with remainder

$$R(\boldsymbol{w}, \boldsymbol{z}) = \sum_{|\boldsymbol{\beta}|=2k} \frac{D_2^{\boldsymbol{\beta}}\Phi(\boldsymbol{w}, \boldsymbol{\xi}_{\boldsymbol{w},\boldsymbol{z}})}{\boldsymbol{\beta}!}(\boldsymbol{z}-\boldsymbol{w})^{\boldsymbol{\beta}},$$

where $\boldsymbol{\xi}_{\boldsymbol{w},\boldsymbol{z}}$ lies somewhere on the line segment connecting $\boldsymbol{w}$ and $\boldsymbol{z}$.

Following the argumentation above we have

$$[P_{\Phi,\mathcal{X}}(\boldsymbol{x})]^2 \leq Q(\boldsymbol{u}) = \Phi(\boldsymbol{x}, \boldsymbol{x}) - 2\sum_j u_j \Phi(\boldsymbol{x}, \boldsymbol{x}_j) + \sum_i \sum_j u_i u_j \Phi(\boldsymbol{x}_i, \boldsymbol{x}_j),$$

where the sums are over those indices j with $u_j \neq 0$. Now we apply the Taylor expansion to both $\Phi(\boldsymbol{x}, \boldsymbol{x}_j)$ and $\Phi(\boldsymbol{x}_i, \boldsymbol{x}_j)$. This yields

$$\begin{aligned} Q(\boldsymbol{u}) &= \Phi(\boldsymbol{x}, \boldsymbol{x}) - 2\sum_j u_j \left[\sum_{|\boldsymbol{\beta}|<2k} \frac{D_2^{\boldsymbol{\beta}}\Phi(\boldsymbol{x}, \boldsymbol{x})}{\boldsymbol{\beta}!}(\boldsymbol{x}_j - \boldsymbol{x})^{\boldsymbol{\beta}} + R(\boldsymbol{x}, \boldsymbol{x}_j) \right] \\ &\quad + \sum_i \sum_j u_i u_j \left[\sum_{|\boldsymbol{\beta}|<2k} \frac{D_2^{\boldsymbol{\beta}}\Phi(\boldsymbol{x}_i, \boldsymbol{x}_i)}{\boldsymbol{\beta}!}(\boldsymbol{x}_j - \boldsymbol{x}_i)^{\boldsymbol{\beta}} + R(\boldsymbol{x}_i, \boldsymbol{x}_j) \right]. \end{aligned}$$

Next, the polynomial precision property of the coefficient vector $\boldsymbol{u}$ simplifies this expression to

$$\begin{aligned} Q(\boldsymbol{u}) &= \Phi(\boldsymbol{x}, \boldsymbol{x}) - 2\Phi(\boldsymbol{x}, \boldsymbol{x}) - 2\sum_j u_j R(\boldsymbol{x}, \boldsymbol{x}_j) \\ &\quad + \sum_i u_i \underbrace{\sum_{|\boldsymbol{\beta}|<2k} \frac{D_2^{\boldsymbol{\beta}}\Phi(\boldsymbol{x}_i, \boldsymbol{x}_i)}{\boldsymbol{\beta}!}(\boldsymbol{x} - \boldsymbol{x}_i)^{\boldsymbol{\beta}}}_{=\Phi(\boldsymbol{x}_i, \boldsymbol{x}) - R(\boldsymbol{x}_i, \boldsymbol{x})} + \sum_i \sum_j u_i u_j R(\boldsymbol{x}_i, \boldsymbol{x}_j). \end{aligned}$$

Rearranging the terms and another application of the Taylor expansion results in

$$\begin{aligned} Q(\boldsymbol{u}) &= -\Phi(\boldsymbol{x}, \boldsymbol{x}) - \sum_j u_j \left[2R(\boldsymbol{x}, \boldsymbol{x}_j) - \sum_i u_i R(\boldsymbol{x}_i, \boldsymbol{x}_j) \right] \\ &\quad + \sum_i u_i \left[\Phi(\boldsymbol{x}_i, \boldsymbol{x}) - R(\boldsymbol{x}_i, \boldsymbol{x}) \right]. \end{aligned}$$

Theorem 5.3.6 allows us to bound $\sum_j |u_j| \leq c_1$. Moreover, since $\|\boldsymbol{x} - \boldsymbol{x}_j\|_2 \leq c_2 h_{\mathcal{X},\Omega}$ and $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \leq 2c_2 h_{\mathcal{X},\Omega}$ the remainder terms can be bounded as stated. □

Remarks:

1. For infinitely smooth strictly positive definite functions such as the Gaussians and the inverse multiquadrics we see that the approximation order k is arbitrarily high.

2. For strictly positive definite functions with limited smoothness such as the Wendland functions $\varphi_{s,k}$ the approximation order is limited by the smoothness of the basic function.

3. The estimate in Theorem 5.3.7 is still generic, since it does not account for the particular basic function Φ being used for the interpolation.

4. We point out that the factor C_Φ may still depend on $h_{\mathcal{X},\Omega}$. For most basic functions it will be possible to use C_Φ to "squeeze out" additional powers of h. This is the reason for splitting the constant in front of the h-power into a generic C and a C_Φ.

The statement of Theorem 5.3.7 can be generalized for strictly conditionally positive definite functions and also to cover the error for approximating the derivatives of f by derivatives of $\mathcal{P}f$. We state this general theorem without comment.

Theorem 5.3.8 *Suppose $\Omega \subseteq \mathbb{R}^s$ is open and bounded and satisfies an interior cone condition. Suppose $\Phi \in C^{2k}(\Omega \times \Omega)$ is symmetric and strictly conditionally positive definite of order m on $\mathbb{R}^s$. Denote the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on the $(m-1)$-unisolvent set $\mathcal{X}$ by $\mathcal{P}f$. Fix $\boldsymbol{\alpha} \in \mathbb{N}_0^s$ with $|\boldsymbol{\alpha}| \le k$. Then there exist positive constant h_0 and C (independent of $\boldsymbol{x}$, f and Φ) such that*

$$|D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} \mathcal{P}f(\boldsymbol{x})| \le C C_\Phi(\boldsymbol{x})^{1/2} h_{\mathcal{X},\Omega}^{k-|\boldsymbol{\alpha}|} |f|_{\mathcal{N}_\Phi(\Omega)},$$

provided $h_{\mathcal{X},\Omega} \le h_0$. Here

$$C_\Phi(\boldsymbol{x}) = \max_{\substack{\boldsymbol{\beta},\boldsymbol{\gamma} \in \mathbb{N}_0^s \\ |\boldsymbol{\beta}|+|\boldsymbol{\gamma}|=2k}} \max_{\boldsymbol{w},\boldsymbol{z} \in \Omega \cap B(\boldsymbol{x}, c_2 h_{\mathcal{X},\Omega})} |D_1^{\boldsymbol{\beta}} D_2^{\boldsymbol{\gamma}} \Phi(\boldsymbol{w},\boldsymbol{z})|.$$

5.4 More on Error Estimates

The additional refinement of the error estimate of Theorem 5.3.8 for specific functions Φ is rather technical (for details see, e.g., the book by Wendland [634]). We only list the final bounds for various functions Φ.

Application of Theorem 5.3.8 to infinitely smooth functions such as Gaussians or (inverse) multiquadrics immediately yields arbitrarily high algebraic convergence rates, i.e., for every $\ell \in \mathbb{N}$ and $|\boldsymbol{\alpha}| \le \ell$ we have

$$|D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} \mathcal{P}f(\boldsymbol{x})| \le C_\ell h^{\ell-|\boldsymbol{\alpha}|} |f|_{\mathcal{N}_\Phi(\Omega)}.$$

whenever $f \in \mathcal{N}_\Phi(\Omega)$. Considerable amount of work has gone into investigating the dependence of the constant C_ℓ on ℓ. Using different proof techniques it is possible to

show that for Gaussians $\Phi(\boldsymbol{x}) = e^{-\alpha\|\boldsymbol{x}\|^2}$, $\alpha > 0$, we get for some positive constant c that

$$\|f - \mathcal{P}f\|_{L_\infty(\Omega)} \leq e^{\frac{-c|\log h_{\mathcal{X},\Omega}|}{h_{\mathcal{X},\Omega}}} \|f\|_{\mathcal{N}_\Phi(\Omega)} \tag{5.2}$$

provided $h_{\mathcal{X},\Omega}$ is sufficiently small and $f \in \mathcal{N}_\Phi(\Omega)$. The corresponding result for (inverse) multiquadrics $\Phi(\boldsymbol{x}) = (\|\boldsymbol{x}\|^2 + \alpha^2)^\beta$, $\alpha > 0$, $\beta < 0$, or $\beta > 0$ and $\beta \notin \mathbb{N}$, is

$$\|f - \mathcal{P}f\|_{L_\infty(\Omega)} \leq e^{\frac{-c}{h_{\mathcal{X},\Omega}}} |f|_{\mathcal{N}_\Phi(\Omega)} \tag{5.3}$$

For functions with finite smoothness (such as powers, thin plate splines, and Wendland's compactly supported functions) it is possible to bound the constant $C_\Phi(\boldsymbol{x})$ and thereby to improve the order predicted by Theorem 5.3.8 by some additional powers of h. This results in the following error estimates.

For the powers $\Phi(\boldsymbol{x}) = (-1)^{\lceil \beta/2 \rceil} \|\boldsymbol{x}\|^\beta$, $\beta > 0$, $\beta \notin 2\,\mathbb{N}$, we get

$$|D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} \mathcal{P}f(\boldsymbol{x})| \leq C h^{\frac{\beta}{2} - |\boldsymbol{\alpha}|} |f|_{\mathcal{N}_\Phi(\Omega)}. \tag{5.4}$$

provided $|\boldsymbol{\alpha}| \leq \frac{\lceil \beta \rceil - 1}{2}$ and $f \in \mathcal{N}_\Phi(\Omega)$.

For thin plate splines $\Phi(\boldsymbol{x}) = (-1)^{k+1} \|\boldsymbol{x}\|^{2k} \log \|\boldsymbol{x}\|$, we get

$$|D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} \mathcal{P}f(\boldsymbol{x})| \leq C h^{k - |\boldsymbol{\alpha}|} |f|_{\mathcal{N}_\Phi(\Omega)}. \tag{5.5}$$

provided $|\boldsymbol{\alpha}| \leq k - 1$ and $f \in \mathcal{N}_\Phi(\Omega)$.

For Wendland's compactly supported functions $\Phi_{s,k}(\boldsymbol{x}) = \varphi_{s,k}(\|\boldsymbol{x}\|)$ this first refinement leads to

$$|D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} \mathcal{P}f(\boldsymbol{x})| \leq C h^{k + \frac{1}{2} - |\boldsymbol{\alpha}|} \|f\|_{\mathcal{N}_\Phi(\Omega)}. \tag{5.6}$$

provided $|\boldsymbol{\alpha}| \leq k$ and $f \in \mathcal{N}_\Phi(\Omega)$.

Remark: The convergence result for the compactly supported functions assumes that the support radius is kept fixed, and that only the domain Ω is filled out by adding more points to $\mathcal{X}$, and thus decreasing the fill distance $h_{\mathcal{X},\Omega}$. However, this means that for small fill distances (with fixed support radius) the system matrices of the interpolation problem become more and more dense – and thus the advantage of the compact support is lost. This point of view is referred to in the literature as the *non-stationary approach*. We are guaranteed convergence, but at the cost of increased computational complexity. Another possibility is presented by the *stationary approach*, for which we scale the support radius proportional to the fill distance. In this case the sparsity of the interpolation matrix remains fixed, however, convergence is lost. We will revisit this phenomenon later.

The powers and thin plate splines can be interpreted as a generalization of univariate natural splines. Therefore, one can see that the approximation order estimates obtained via the native space approach are not optimal. For example, for interpolation with thin plate splines $\Phi(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 \log \|\boldsymbol{x}\|$ one would expect order $\mathcal{O}(h^2)$, but the above estimate yields only $\mathcal{O}(h)$.

One can improve the estimates for functions with finite smoothness (i.e., powers, thin plate splines, Wendland's functions) by either (or both) of the following two ideas:

- by requiring the data function f to be even smoother than what the native space prescribes, i.e., by building certain boundary conditions into the native space;
- by using weaker norms to measure the error.

The idea to localize the data by adding boundary conditions was introduced in a paper by Light and Wayne [394]. This "trick" allows us to double the approximation order. The second idea can already be found in the early work by Duchon [170]. After applying both improvements the final approximation order estimate for interpolation with the compactly supported functions $\Phi_{s,k}$ is (see [629])

$$\|f - \mathcal{P}f\|_{L_2(\Omega)} \leq Ch^{2k+1+s}\|f\|_{W_2^{2k+1+s}(\mathbb{R}^s)}, \tag{5.7}$$

where f is assumed to lie in the subspace $W_2^{2k+1+s}(\mathbb{R}^s)$ of $\mathcal{N}_\Phi(\mathbb{R}^s)$. For powers and thin plate splines one obtains L_2-error estimates of order $\mathcal{O}(h^{\beta+s})$ and $\mathcal{O}(h^{2k+s})$, respectively. These estimates are optimal, i.e., exact approximation orders, as shown by Bejancu [48].

Work on improved error bounds is also due to others such as Bejancu, Johnson, Powell, Ron, Schaback, and Yoon. In particular, recent work by Yoon provides L_p error estimates for so-called *shifted surface splines* for functions f is standard Sobolev spaces. These functions include all of the (inverse) multiquadrics, powers and thin plate splines. They are of the form

$$\Phi(\boldsymbol{x}) = \begin{cases} (-1)^{\lceil \beta - s/2 \rceil}(\|\boldsymbol{x}\|^2 + \alpha^2)^{\beta - s/2}, & s \text{ odd}, \\ (-1)^{\beta - s/2 + 1}(\|\boldsymbol{x}\|^2 + \alpha^2)^{\beta - s/2}\log(\|\boldsymbol{x}\|^2 + \alpha^2)^{1/2}, & s \text{ even}, \end{cases}$$

where $\beta \in \mathbb{N}$, $\beta > s/2$.

Yoon [668] has the following theorem that is formulated in the stationary setting.

Theorem 5.4.1 *Let Φ be a shifted surface spline with parameter α proportional to the fill distance $h_{\mathcal{X},\Omega}$. Then there exists a positive constant C (independent of $\mathcal{X}$) such that for every $f \in W_2^m(\Omega) \cap W_\infty^m(\Omega)$ we have*

$$\|f - \mathcal{P}f\|_{L_p(\Omega)} \leq Ch^{\gamma_p}|f|_{W_2^m(\mathbb{R}^s)}, \qquad 1 \leq p \leq \infty,$$

with

$$\gamma_p = \min\{m, m - s/2 + s/p\}.$$

Furthermore, if $f \in W_2^k(\Omega) \cap W_\infty^k(\Omega)$ with $\max\{0, s/2 - s/p\} < k < m$, then

$$\|f - \mathcal{P}f\|_{L_p(\Omega)} = o(h^{\gamma_p - m + k}).$$

Remarks:

1. Using the localization idea mentioned above Yoon's estimates can be "doubled" to $\mathcal{O}(h^{m+\gamma_p})$.

2. Yoon's estimates address the question of how well the infinitely smooth (inverse) multiquadrics approximate functions that are less smooth than those in their native space. For example, Theorem 5.4.1 states that approximation to functions in $W_2^2(\Omega)$, $\Omega \subseteq \mathbb{R}^3$, by multiquadrics $\Phi(\boldsymbol{x}) = \sqrt{\|\boldsymbol{x}\|^2 + \alpha^2}$ is of the order $\mathcal{O}(h^2)$. However, it needs to be emphasized that this refers to stationary approximation, i.e., α is scaled proportional to the fill distance, whereas the spectral order given in (5.3) corresponds to the non-stationary case with fixed α. Similar numerical evidence was also provided much earlier by Schaback [545].

3. Moreover, the second part of Yoon's result is a step toward *exact* approximation orders.

4. In order to obtain the estimates for the infinitely smooth functions, Yoon localizes the data function f by preconditioning it via convolution with a Hörmander smoothing kernel.

5.5 The Connection to Optimal Recovery

In the paper [264] by Michael Golomb and Hans Weinberger the following general problem is studied: Given the values $f_1 = \lambda_1(f), \ldots, f_N = \lambda_N(f) \in \mathbb{R}$, where $\{\lambda_1, \ldots, \lambda_N\}$ is a linearly independent set of linear functionals (called *information functionals* yielding the *information* about f), how does one "best" approximate the value $\lambda(f)$ where λ is a given linear functional and f is unknown? The value $\lambda(f)$ is also referred to as a *feature* of f. Moreover, what is the total range of values for $\lambda(f)$?

Remarks:

1. This is a very general problem formulation that allows not only for interpolation of function values, but also for other types of data (such as values of derivatives, integrals of f, moments of f, etc.), as well as other types of approximation.

2. Optimal recovery was also studied in detail by Micchelli, Rivlin and Winograd [457, 458, 459, 460].

In a Hilbert space setting the solution to this "optimal recovery problem" is shown to be the *minimum-norm interpolant*. More precisely, given $f_1 = \lambda_1(f), \ldots, f_N = \lambda_N(f) \in \mathbb{R}$ with $\{\lambda_1, \ldots, \lambda_N\} \subseteq \mathcal{H}^*$, the minimum-norm interpolant is that function $s^* \in \mathcal{H}$ that satisfies

$$\lambda_j(s^*) = f_j, \qquad j = 1, \ldots, N,$$

and for which

$$\|s^*\|_{\mathcal{H}} = \min_{\substack{s \in \mathcal{H} \\ \lambda_j(s) = f_j, j = 1, \ldots, N}} \|s\|_{\mathcal{H}}.$$

It turns out that the radial basis function interpolant satisfies these criteria if $\mathcal{H}$ is taken as the associated native space $\mathcal{N}_\Phi(\Omega)$. The proofs of the two "optimality theorems" below require the following two lemmas.

Lemma 5.5.1 *Assume Φ is a symmetric strictly positive definite kernel on $\mathbb{R}^s$ and let $\mathcal{P}f$ be the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega$. Then*

$$\langle \mathcal{P}f, \mathcal{P}f - s \rangle_{\mathcal{N}_\Phi(\Omega)} = 0$$

for all interpolants $s \in \mathcal{N}_\Phi(\mathcal{X})$, i.e., with $s(\boldsymbol{x}_j) = f(\boldsymbol{x}_j)$, $j = 1, \ldots, N$.

Proof: The interpolant $\mathcal{P}f$ is of the form

$$\mathcal{P}f = \sum_{j=1}^{N} c_j \Phi(\cdot, \boldsymbol{x}_j).$$

Using this representation, the symmetry of the kernel Φ and its reproducing property we have

$$\begin{aligned}
\langle \mathcal{P}f, \mathcal{P}f - s \rangle_{\mathcal{N}_\Phi(\Omega)} &= \langle \sum_{j=1}^{N} c_j \Phi(\cdot, \boldsymbol{x}_j), \mathcal{P}f - s \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&= \sum_{j=1}^{N} c_j \langle \Phi(\cdot, \boldsymbol{x}_j), \mathcal{P}f - s \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&= \sum_{j=1}^{N} c_j \langle \mathcal{P}f - s, \Phi(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&= \sum_{j=1}^{N} c_j (\mathcal{P}f - s)(\boldsymbol{x}_j) \\
&= 0
\end{aligned}$$

since both $\mathcal{P}f$ and s interpolate f on $\mathcal{X}$. □

For the next result, remember the definition of the space $H_\Phi(\mathcal{X})$ as the linear span

$$H_\Phi(\mathcal{X}) = \{ s = \sum_{j=1}^{N} c_j \Phi(\cdot, \boldsymbol{x}_j), \boldsymbol{x}_j \in \mathcal{X} \}$$

given at the beginning of this chapter.

Lemma 5.5.2 *Assume Φ is a strictly positive definite kernel on $\mathbb{R}^s$ and let $\mathcal{P}f$ be the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega$. Then*

$$\langle f - \mathcal{P}f, s \rangle_{\mathcal{N}_\Phi(\Omega)} = 0$$

for all $s \in H_\Phi(\mathcal{X})$.

Proof: Any $s \in H_\Phi(\mathcal{X})$ can be written in the form

$$s = \sum_{j=1}^{N} c_j \Phi(\cdot, \boldsymbol{x}_j).$$

Using this representation of s as well as the reproducing property of Φ we have

$$\begin{aligned}\langle f - \mathcal{P}f, s\rangle_{\mathcal{N}_\Phi(\Omega)} &= \langle f - \mathcal{P}f, \sum_{j=1}^{N} c_j \Phi(\cdot, \boldsymbol{x}_j)\rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \sum_{j=1}^{N} c_j \langle f - \mathcal{P}f, \Phi(\cdot, \boldsymbol{x}_j)\rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \sum_{j=1}^{N} c_j (f - \mathcal{P}f)(\boldsymbol{x}_j).\end{aligned}$$

This last expression, however, is zero since $\mathcal{P}f$ interpolates f on $\mathcal{X}$, i.e., $(f-\mathcal{P}f)(\boldsymbol{x}_j) = 0$, $j = 1, \ldots, N$. □

The following "energy splitting" theorem is an immediate consequence of Lemma 5.5.2. It says that the native space energy of f can be split into the energy of the interpolant $\mathcal{P}f$ and that of the residual $f - \mathcal{P}f$.

Corollary 5.5.3 *The orthogonality property of Lemma 5.5.2 implies the* energy split

$$\|f\|^2_{\mathcal{N}_\Phi(\Omega)} = \|f - \mathcal{P}f\|^2_{\mathcal{N}_\Phi(\Omega)} + \|\mathcal{P}f\|^2_{\mathcal{N}_\Phi(\Omega)}.$$

Proof: The statement follows from

$$\begin{aligned}\|f\|^2_{\mathcal{N}_\Phi(\Omega)} &= \|f - \mathcal{P}f + \mathcal{P}f\|^2_{\mathcal{N}_\Phi(\Omega)} \\ &= \langle (f - \mathcal{P}f) + \mathcal{P}f, (f - \mathcal{P}f) + \mathcal{P}f\rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \|f - \mathcal{P}f\|^2_{\mathcal{N}_\Phi(\Omega)} + 2\langle f - \mathcal{P}f, \mathcal{P}f\rangle_{\mathcal{N}_\Phi(\Omega)} + \|\mathcal{P}f\|^2_{\mathcal{N}_\Phi(\Omega)}\end{aligned}$$

and the fact that $\langle f - \mathcal{P}f, \mathcal{P}f\rangle_{\mathcal{N}_\Phi(\Omega)} = 0$ by Lemma 5.5.2. □

Remark: The above energy split is the fundamental idea behind a number of Krylov-type iterative algorithms for approximately solving the interpolation problem when very large data sets are involved (see, e.g., the papers [212] and [213] by Faul and Powell or [562] by Schaback and Wendland).

The following theorem shows the first optimality property of strictly conditionally positive definite kernels. It is taken from [634].

Theorem 5.5.4 *Suppose $\Phi \in C(\Omega \times \Omega)$ is a strictly conditionally positive definite kernel with respect to the finite-dimensional space $P \subseteq C(\Omega)$ and that $\mathcal{X}$ is P-unisolvent. If the values $f_1, \ldots, f_N$ are given, then the interpolant $\mathcal{P}f$ is the minimum-norm interpolant to $\{f_j\}_{j=1}^N$, i.e.,*

$$|\mathcal{P}f|_{\mathcal{N}_\Phi(\Omega)} = \min_{\substack{s \in \mathcal{N}_\Phi(\Omega) \\ s(\boldsymbol{x}_j) = f_j, j = 1, \ldots, N}} |s|_{\mathcal{N}_\Phi(\Omega)}.$$

Proof: We consider only the strictly positive definite case. Consider an arbitrary interpolant $s \in \mathcal{N}_\Phi(\Omega)$ to $f_1, \ldots, f_N$. Since $\mathcal{P}f \in \mathcal{N}_\Phi(\Omega)$ we can apply Lemma 5.5.1 and get

$$\langle \mathcal{P}f, \mathcal{P}f - s\rangle_{\mathcal{N}_\Phi(\Omega)} = 0.$$

Now

$$\begin{aligned} |\mathcal{P}f|^2_{\mathcal{N}_\Phi(\Omega)} &= \langle \mathcal{P}f, \mathcal{P}f - s + s\rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \langle \mathcal{P}f, \mathcal{P}f - s\rangle_{\mathcal{N}_\Phi(\Omega)} + \langle \mathcal{P}f, s\rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \langle \mathcal{P}f, s\rangle_{\mathcal{N}_\Phi(\Omega)} \\ &\leq |\mathcal{P}f|_{\mathcal{N}_\Phi(\Omega)} |s|_{\mathcal{N}_\Phi(\Omega)} \end{aligned}$$

so that the statement follows. □

Remarks:

1. The space P mentioned in Theorem 5.5.4 is usually taken as a space of multivariate polynomials.

2. For thin plate splines $\phi(r) = r^2 \log r$, $r = \|\boldsymbol{x}\|_2$ with $\boldsymbol{x} = (x, y) \in \mathbb{R}^2$, the corresponding semi-norm in the Beppo-Levi space $BL_2(\mathbb{R}^2)$ is

$$|f|^2_{BL_2(\mathbb{R}^2)} = \int_{\mathbb{R}^2} \left| \frac{\partial^2 f}{\partial x^2}(\boldsymbol{x}) \right|^2 + 2 \left| \frac{\partial^2 f}{\partial x \partial y}(\boldsymbol{x}) \right|^2 + \left| \frac{\partial^2 f}{\partial y^2}(\boldsymbol{x}) \right|^2 d\boldsymbol{x},$$

   which is the bending energy of a thin plate, and thus explains the name of these functions.

Another nice property of the radial basis function interpolant is that it is at the same time the best Hilbert-space approximation to the given data, and thus not just any projection of f but the *orthogonal projection*. More precisely,

Theorem 5.5.5 *Let*

$$H_\Phi(\mathcal{X}) = \{s = \sum_{j=1}^N c_j \Phi(\cdot, \boldsymbol{x}_j) + p \mid p \in P \text{ and } \sum_{j=1}^N c_j q(\boldsymbol{x}_j) = 0 \text{ for all } q \in P \text{ and } \boldsymbol{x}_j \in \mathcal{X}\},$$

where $\Phi \in C(\Omega \times \Omega)$ is a strictly conditionally positive definite kernel with respect to the finite-dimensional space $P \subseteq C(\Omega)$ and $\mathcal{X}$ is P-unisolvent. If only the values $f_1 = f(\boldsymbol{x}_1), \ldots, f_N = f(\boldsymbol{x}_N)$ are given, then the interpolant $\mathcal{P}f$ is the best approximation to f from $H_\Phi(\mathcal{X})$ in $\mathcal{N}_\Phi(\Omega)$, i.e.,

$$|f - \mathcal{P}f|_{\mathcal{N}_\Phi(\Omega)} \leq |f - s|_{\mathcal{N}_\Phi(\Omega)}$$

for all $s \in H_\Phi(\mathcal{X})$.

Proof: We consider only the strictly positive definite case. As explained in Section 5.2, the native space $\mathcal{N}_\Phi(\Omega)$ is the completion of $H_\Phi(\Omega)$ with respect to the $\|\cdot\|_\Phi$-norm so that $\|f\|_\Phi = \|f\|_{\mathcal{N}_\Phi(\Omega)}$ for all $f \in H_\Phi(\Omega)$. Also, $\mathcal{X} \subseteq \Omega$. Therefore, we can express best approximation from $H_\Phi(\mathcal{X})$ by

$$\langle f - s^*, s\rangle_{\mathcal{N}_\Phi(\Omega)} = 0 \qquad \text{for all } s \in H_\Phi(\mathcal{X}).$$

However, Lemma 5.5.2 shows that $s^* = \mathcal{P}f$ satisfies this relation. □

Remarks:

1. The connection between radial basis function interpolation and the optimal recovery theory by Golomb and Weinberger was pointed out by various people (e.g., Schaback [545, 551], or Light and Wayne [394]).

2. These optimality properties of radial basis function interpolants play an important role in applications such as in the design of support vector machines in artificial intelligence or the numerical solutions of partial differential equations.

3. The optimality results above imply that one could also start with some Hilbert space $\mathcal{H}$ with norm $\|\cdot\|_{\mathcal{H}}$ and ask to find the minimum norm interpolant (i.e., Hilbert space best approximation) to some given data. In this way any given space defines a set of *optimal basis functions*, generated by the reproducing kernel of $\mathcal{H}$. This is how Duchon approached the subject in his papers [168, 169, 170, 171]. More recently, Kybic, Blu and Unser [356, 357] take this point of view and explain very nicely from a sampling theory point of view how the thin plate splines can be interpreted a fundamental solutions of the differential operator defining the semi-norm in the Beppo-Levi space $BL_2(\mathbb{R}^2)$, and thus radial basis functions can be viewed as *Green's functions*.

A third optimality result is in the context of quasi-interpolation, i.e.,

Theorem 5.5.6 *Suppose $\Phi \in C(\Omega \times \Omega)$ is a strictly conditionally positive definite kernel with respect to the finite-dimensional space $P \subseteq C(\Omega)$. Suppose $\mathcal{X}$ is P-unisolvent and $\boldsymbol{x} \in \Omega$ is fixed. Let $u_j^*(\boldsymbol{x})$, $j = 1, \ldots, N$, be the cardinal basis functions for interpolation with Φ. Then*

$$\left| f(\boldsymbol{x}) - \sum_{j=1}^{N} f(\boldsymbol{x}_j) u_j^*(\boldsymbol{x}) \right| \leq \left| f(\boldsymbol{x}) - \sum_{j=1}^{N} f(\boldsymbol{x}_j) u_j \right|$$

for all choices of $u_1, \ldots, u_N \in \mathbb{R}$ with $\sum_{j=1}^{N} u_j p(\boldsymbol{x}_j) = p(\boldsymbol{x})$ for any $p \in P$.

Theorem 5.5.6 is proved in [634]. It says in particular that the minimum norm interpolant $\mathcal{P}f$ is also more accurate (in the pointwise sense) than any linear combination of the given data values.

Chapter 6

Least Squares Approximation

As we saw in Chapter 5 we can interpret radial basis function interpolation as a constrained optimization problem. We now take this point of view again, but start with a more general formulation. Let's assume we are seeking a function $\mathcal{P}f$ of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \Phi(\boldsymbol{x}, \boldsymbol{x}_j), \qquad \boldsymbol{x} \in \mathbb{R}^s,$$

such that the quadratic form

$$\frac{1}{2}\boldsymbol{c}^T Q \boldsymbol{c} \tag{6.1}$$

with $\boldsymbol{c} = [c_1, \ldots, c_M]^T$ and some symmetric positive definite matrix Q is minimized subject to the linear constraints

$$A\boldsymbol{c} = \boldsymbol{f} \tag{6.2}$$

where A is an $N \times M$ matrix, and the right-hand side $\boldsymbol{f} = [f_1, \ldots, f_N]^T$ is given. Such a constrained quadratic minimization problem can be converted to a system of linear equations by introducing *Lagrange multipliers*, i.e., we consider finding the minimum of

$$\frac{1}{2}\boldsymbol{c}^T Q \boldsymbol{c} - \boldsymbol{\lambda}^T \left[A\boldsymbol{c} - \boldsymbol{f}\right] \tag{6.3}$$

with respect to $\boldsymbol{c}$ and $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_N]^T$. Since Q is a positive definite matrix, it is well known that the functional to be minimized is convex, and thus has a unique minimum. Therefore, the standard necessary condition for such a minimum (which is obtained by differentiating with respect to $\boldsymbol{c}$ and $\boldsymbol{\lambda}$ and finding the zeros of those derivatives) is also sufficient. This leads to

$$\begin{aligned} Q\boldsymbol{c} - A^T\boldsymbol{\lambda} &= \boldsymbol{0} \\ A\boldsymbol{c} - \boldsymbol{f} &= \boldsymbol{0} \end{aligned}$$

or, in matrix form,

$$\begin{bmatrix} Q & -A^T \\ A & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{f} \end{bmatrix}.$$

By applying Gaussian elimination to this block matrix (Q is invertible since it is assumed to be positive definite) we get

$$\boldsymbol{\lambda} = \left(AQ^{-1}A^T\right)^{-1}\boldsymbol{f} \tag{6.4}$$

$$\boldsymbol{c} = Q^{-1}A^T\left(AQ^{-1}A^T\right)^{-1}\boldsymbol{f}. \tag{6.5}$$

In particular, if the quadratic form represents the native space norm of the interpolant $\mathcal{P}f = \sum_{j=1}^{M} c_j\Phi(\cdot,\boldsymbol{x}_j)$, i.e.,

$$\|\mathcal{P}f\|^2_{\mathcal{N}_\Phi(\Omega)} = \sum_{i=1}^{M}\sum_{j=1}^{M} c_i c_j \Phi(\boldsymbol{x}_i,\boldsymbol{x}_j) = \boldsymbol{c}^T Q\boldsymbol{c}$$

with $Q_{ij} = \Phi(\boldsymbol{x}_i,\boldsymbol{x}_j)$ and $\boldsymbol{c} = [c_1,\ldots,c_M]^T$, and the linear side conditions are the interpolation conditions

$$A\boldsymbol{c} = \boldsymbol{f} \qquad \Longleftrightarrow \qquad \mathcal{P}f(\boldsymbol{x}_i) = f_i, \quad i = 1,\ldots,M,$$

with $A = Q$ (symmetric) and the same $\boldsymbol{c}$ as above and data vector $\boldsymbol{f} = [f_1,\ldots,f_M]^T$, then we see that the Lagrange multipliers (6.4) become

$$\boldsymbol{\lambda} = A^{-T}\boldsymbol{f} = A^{-1}\boldsymbol{f}$$

and the coefficients are given by

$$\boldsymbol{c} = \boldsymbol{\lambda}$$

via (6.5). Therefore, as we saw earlier, the minimum norm interpolant is obtained by solving the interpolation equations alone.

Since we took the more general point of view that $\mathcal{P}$ is generated by M basis functions, and N linear constraints are specified, the above formulation also covers both over- and under-determined least squares fitting where the quadratic form $\boldsymbol{c}^T Q\boldsymbol{c}$ represents an added *smoothing* (or *regularization*) term. This term is not required to obtain a unique solution of the system $A\boldsymbol{c} = \boldsymbol{f}$ in the over-determined case ($M \le N$), but in the under-determined case such a constraint is needed (cf. the solution of under-determined linear systems via singular value decomposition in the numerical linear algebra literature (e.g., [608])).

Usually the regularized least squares approximation problem is formulated as minimization of

$$\frac{1}{2}\boldsymbol{c}^T Q\boldsymbol{c} + \omega\sum_{j=1}^{N}\left(\mathcal{P}f(\boldsymbol{x}_j) - f_j\right)^2. \tag{6.6}$$

The quadratic form controls the smoothness of the fitting function and the least squares term measures the closeness to the data. The parameter ω controls the tradeoff between these two terms. The formulation (6.6) is used in *regularization theory* (see, e.g., [185, 252]). The same formulation is also used in *penalized least squares* fitting (see, e.g., [263]), the literature on smoothing splines [528, 572], and in papers by Wahba on thin plate splines (e.g., [615, 621]). In fact, the idea of smoothing a data fitting process by this kind of formulation seems to go back to at least Whittaker [641] in 1923. In practice a penalized least squares formulation is especially useful if the data f_i cannot

be completely trusted, i.e., it is contaminated by noise. In this case, a (penalized) least squares fit is advisable. The problem of minimizing (6.6) is known as *ridge regression* in the statistics literature.

The equivalence with our formulation (6.3) above follows from

$$\begin{aligned}\frac{1}{2}\boldsymbol{c}^T Q\boldsymbol{c} + \omega \sum_{j=1}^{N} (\mathcal{P}f(\boldsymbol{x}_j) - f_j)^2 &= \frac{1}{2}\boldsymbol{c}^T Q\boldsymbol{c} + \omega[A\boldsymbol{c} - \boldsymbol{f}]^T[A\boldsymbol{c} - \boldsymbol{f}] \\ &= \frac{1}{2}\boldsymbol{c}^T Q\boldsymbol{c} - \boldsymbol{\lambda}^T[A\boldsymbol{c} - \boldsymbol{f}],\end{aligned}$$

where

$$\boldsymbol{\lambda} = -\omega[A\boldsymbol{c} - \boldsymbol{f}].$$

We are now interested in the more general setting where we still sample the given function f on the set $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, but now introduce a second set $\Xi = \{\boldsymbol{\xi}_i\}_{i=1}^M$ at which we center the basis functions. Usually we will have $M \leq N$, and the case $M = N$ with $\Xi = \mathcal{X}$ recovers the traditional interpolation setting discussed thus far. Therefore,

$$\mathcal{Q}f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \Phi(\boldsymbol{x}, \boldsymbol{\xi}_j), \quad \boldsymbol{x} \in \mathbb{R}^s, \tag{6.7}$$

and the coefficients c_j can be found by minimizing $\|Qf - f\|_2^2$, where the ℓ_2-norm

$$\|f\|_2^2 = \sum_{i=1}^{N} [f(\boldsymbol{x}_i)]^2$$

is induced by the discrete inner product

$$\langle f, g \rangle = \sum_{i=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_i). \tag{6.8}$$

This approximation problem has a unique solution if the (rectangular) *collocation matrix* A with entries

$$A_{jk} = \Phi(\boldsymbol{x}_j, \boldsymbol{\xi}_k), \quad j = 1, \ldots, N, \ k = 1, \ldots, M,$$

has full rank.

Remarks:

1. If the centers in Ξ are chosen to form a subset of the data locations $\mathcal{X}$ then A has full rank provided the radial basis function is selected according to our previous chapters on interpolation. This is true, since in this case A will have an $M \times M$ square submatrix which is non-singular (by virtue of being an *interpolation matrix*).

2. The over-determined linear system $A\boldsymbol{c} = \boldsymbol{y}$ which arises in the solution of the least squares problem can be solved using standard algorithms from numerical linear algebra such as QR or singular value decomposition.

In the following section we give a brief account of theoretical results known for the general problem in which the centers and data sites differ.

6.1 Theoretical Results

The results mentioned here are due to Sivakumar and Ward [583], and Quak, Sivakumar and Ward [521]. The first paper deals with discrete least squares, the second with continuous least squares approximation. In both papers the authors do not discuss the collocation matrix A above, but rather base their results on the non-singularity of the coefficient matrix obtained from a system of normal equations. In the discrete setting they use the inner product (6.8) which induces the ℓ_2 norm, and then discuss non-singularity of the *Gramian* which occurs in the following system of *normal equations*

$$G\boldsymbol{c} = \boldsymbol{w}, \tag{6.9}$$

where the entries of G are the ℓ_2 inner products of the radial basis functions, i.e.,

$$G_{jk} = \langle \Phi(\cdot, \boldsymbol{\xi}_j), \Phi(\cdot, \boldsymbol{\xi}_k) \rangle = \sum_{i=1}^{N} \Phi(\boldsymbol{x}_i, \boldsymbol{\xi}_j)\Phi(\boldsymbol{x}_i, \boldsymbol{\xi}_k), \quad j,k = 1, \ldots, M,$$

and the right-hand side vector $\boldsymbol{w}$ in (6.9) is given by

$$\boldsymbol{w}_j = \langle \Phi(\cdot, \boldsymbol{\xi}_j), \boldsymbol{f} \rangle = \sum_{i=1}^{N} \Phi(\boldsymbol{x}_i, \boldsymbol{\xi}_j) f(\boldsymbol{x}_i), \quad j = 1, \ldots, M.$$

Remarks:

1. Note that in the interpolation case with $M = N$ and $\Xi = \mathcal{X}$ we have

   $$\langle \Phi(\cdot, \boldsymbol{x}_j), \Phi(\cdot, \boldsymbol{x}_k) \rangle = \langle \Phi(\cdot, \boldsymbol{x}_j), \Phi(\cdot, \boldsymbol{x}_k) \rangle_{\mathcal{N}_\Phi(\Omega)} = \Phi(\boldsymbol{x}_j, \boldsymbol{x}_k)$$

   so that G is just the interpolation matrix A.

2. Of course, this also presents an interpretation of the interpolation matrix A as the system matrix for the normal equations in the case of best approximation with respect to the native space norm – a fact already mentioned earlier in the section on optimal recovery.

In both papers, [583] as well as [521], even the formulation of the main theorems is very technical. We therefore just try to give a feel for their results.

Essentially, the authors show that the Gramian for certain radial basis functions (norm, (inverse) multiquadrics, and Gaussians) is non-singular if the centers $\boldsymbol{\xi}_k$, $k = 1, \ldots, M$, are sufficiently well distributed and the data points $\boldsymbol{x}_j$, $j = 1, \ldots, N$, are fairly evenly clustered about the centers with the diameter of the clusters being relatively small compared to the separation distance of the data points. Figure 6.1 illustrates the clustering idea.

One of the key ingredients in the proof of the non-singularity of G is to set up an interpolation matrix B for which the basis functions are centered at certain representatives of the clusters of knots about the data sites. One then splits the matrix B (which is non-symmetric in general) into a part which is symmetric and one which is

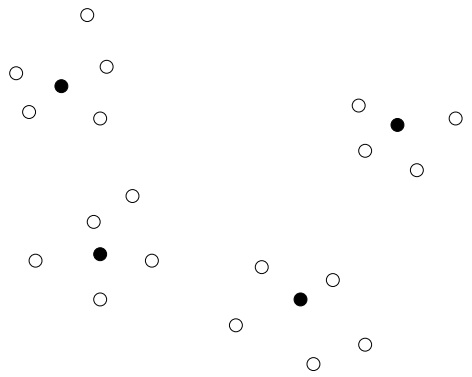

Figure 6.1: Clusters of data points ∘ around well separated centers •.

anti-symmetric, a standard strategy in linear algebra, i.e., $B = B_1 + B_2$ where B_1 and B_2 are defined by

$$\begin{aligned} B_1 &= \frac{B + B^T}{2}, \quad \text{(symmetric)}, \\ B_2 &= \frac{B - B^T}{2}, \quad \text{(anti-symmetric)}. \end{aligned}$$

Then, lower estimates for the norm of these two parts are found and used to conclude that, under certain restrictions, G is non-singular.

Remarks:

1. As a by-product of this argumentation the authors obtain a proof for the non-singularity of *interpolation* matrices for the case in which the centers of the basis functions are chosen different from the data sites, namely as small perturbations thereof.

2. The discussion of the continuous case is very similar to that of the discrete one.

6.2 Adaptive Least Squares using Knot Insertion

In this and in the following section we mention some strategies for an algorithm for solving the least squares problem in an adaptive fashion. When fitting data by linear combinations of certain basis functions, it is a classical technique to improve the quality of a given initial approximation by increasing the number of basis functions used for the fit, i.e., by refining the space from which we are approximating. Since every radial basis function is associated with one particular center (or *knot*), this can be achieved by adding new knots to the existing ones. This idea was explored for multiquadrics on $\mathbb{R}^2$ in [237, 238], and for radial basis functions on spheres in [192].

We will now describe an algorithm which adaptively adds knots to a radial basis function approximant in order to improve the ℓ_2 error.

Let us assume we are given a large number, N, of data and we want to fit them with a radial basis expansion to within a given tolerance. The idea is to start with very

few initial knots, and then to repeatedly insert a knot at that data location whose ℓ_2 error component is largest. This is done as long as the least squares error exceeds a given tolerance. The following algorithm may be used.

Algorithm: Knot insertion

(1) Let data sites $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, data f_i, $i = 1, \ldots, N$, and a tolerance ε be given.

(2) Choose M initial knots $\Xi = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M\}$.

(3) Calculate the least squares fit

$$\mathcal{Q}f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \Phi(\boldsymbol{x}, \boldsymbol{\xi}_j)$$

with its associated error

$$e = \sum_{i=1}^{N} [f_i - \mathcal{Q}f(\boldsymbol{x}_i)]^2.$$

While $e > \varepsilon$ **do**

(4) "Weight" each data point $\boldsymbol{x}_i$, $i = 1, \ldots, N$, according to its error component, i.e., let

$$w_i = |f_i - \mathcal{Q}f(\boldsymbol{x}_i)|, \quad i = 1, \ldots, N.$$

(5) Find the data point $\boldsymbol{x}_\nu \notin \Xi$ with maximum weight w_ν and insert it as a knot, i.e.,

$$\Xi = \Xi \cup \{\boldsymbol{x}_\nu\} \quad \text{and} \quad M = M + 1.$$

(6) Recalculate fit and associated error.

Remarks:

1. We note that we have to solve one linear least squares problem for every knot we add. This can be done employing standard techniques such as QR or SVD factorization. The size of these problems increases by one at each step. In order to improve the runtime of this algorithms an updating QR factorization (see e.g., [265]) could be used. However, neither [237, 238] nor [192] have found this necessary since the problems they considered begin with very small systems (and therefore fast solutions), and the desired accuracy was achieved with fairly few additional knots.

2. If the initial knots are chosen to lie at data sites, the process described in the above algorithm will always have a full rank collocation matrices A. This is guaranteed, since we only add data sites as new knots, and we make sure in step (5) that no multiple knots are created (which would obviously lead to a rank deficiency).

6.3 Adaptive Least Squares using Knot Removal

The idea of knot removal was primarily motivated by the need for data reduction, but it can also be used for the purpose of adaptive approximation (for a survey of knot removal see, e.g., [412]). The basic idea is to start with a good fit (e.g., an interpolation to the data), and then successively reduce the number of knots (and therefore basis functions) used until a certain given tolerance is reached.

Specifically, this means we will start with an initial fit and then use some kind of weighting strategy for the knots, so that we can repeatedly remove those contributing least to the accuracy of the fit. The following algorithm performs this task.

Algorithm: Knot removal

(1) Let data points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, data f_i, $i = 1, \ldots, N$, and a tolerance ε be given.

(2) Choose M initial knots $\Xi = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M\}$.

(3) Calculate an initial fit

$$\mathcal{Q}f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \Phi(\boldsymbol{x}, \boldsymbol{\xi}_j)$$

with its associated least squares error

$$e = \sum_{i=1}^{N} [f_i - \mathcal{Q}f(\boldsymbol{x}_i)]^2.$$

While $e < \varepsilon$ **do**

(4) "Weight" each knot $\boldsymbol{\xi}_j$, $j = 1, \ldots, M$, according to its least squares error, i.e., form

$$\Xi^* = \Xi \setminus \{\boldsymbol{\xi}_j\},$$

and calculate the weights

$$w_j = \sum_{i=1}^{N} \left[f_i - \mathcal{Q}^* f(\boldsymbol{x}_i)\right]^2,$$

where

$$\mathcal{Q}^* f(\boldsymbol{x}) = \sum_{j=1}^{M-1} c_j \Phi(\boldsymbol{x}, \boldsymbol{\xi}_j^*)$$

is the approximation based on the reduced set of knots Ξ^*.

(5) Find the knot $\boldsymbol{\xi}_\mu$ with lowest weight $w_\mu < \varepsilon$ and permanently remove it, i.e.

$$\Xi = \Xi \setminus \{\boldsymbol{\xi}_\mu\} \quad \text{and} \quad M = M - 1.$$

(6) Recalculate fit and associated error.

Again we would like to comment on the algorithm.

Remarks:

1. As far as computational times are concerned, this algorithm is *much* slower than the knot insertion algorithm, since the weight for every knot is determined by the solution of a least squares problem, i.e., in every iteration we solve M least squares problems. These problems, however, become increasingly smaller.

2. This approach should be especially beneficial when the number of evaluations of the constructed approximant is much larger than its actual computation, since, for the same tolerance, one would expect knot removal to result in fewer knots than knot insertion.

3. If the initial knots are chosen at the data sites, then again there will be no problems with the collocation matrix becoming rank deficient.

4. In [192, 193] some alternatives to this knot removal strategy were also considered. One of them is the removal of certain groups of knots at one time in order to speed up the process. We used this version of knot removal algorithm in our examples in the next subsection. Another is based on choosing the weights based on the size of the coefficients c_j in the expansion of $\mathcal{Q}f$, i.e., to remove that knot whose associated coefficient is smallest.

5. A further variation of the adaptive algorithms was considered in both [237] and in [192]. Instead of treating only the coefficients of the expansion of $\mathcal{Q}f$ as parameters in the minimization process, one can also include the knot locations themselves and possibly the parameters which are inherent in the definition of some of the radial functions used in practice. This however, leads to *nonlinear* least squares problems. We will not discuss this topic further here.

6. Buhmann, Derrien, and Le Méhauté [88], and Le Méhauté [366] also discuss knot removal. Their approach is based on an *a priori estimate* for the error made when removing a certain knot. These estimates depend on the specific choice of radial basis function, and so far they only cover the inverse multiquadric type, i.e.,

$$\varphi(r) = (r^2 + \alpha^2)^{\beta/2}, \quad -s \leq \beta < 0, \ \alpha > 0.$$

7. Iske [317] suggests an alternative knot removal strategy for least squares approximation. His removal heuristics are based on so-called *thinning algorithms*. In particular, in each iteration a point is removed if it belongs to a pair of points in Ξ with minimal separation distance. The thinning phase of the algorithm is then enhance by an exchange phase in which points can be "swapped back in" if this process reduces the fill-distance of Ξ.

	μ	ρ	# knots used	time
KI	8.074767e-02	1.773359e-04	30	4 sec
KR	6.948009e-02	1.488779e-04	36	153 sec

Table 6.1: Comparison of adaptive algorithms for Franke's function ($\varepsilon = 0.01$).

6.4 Some Numerical Examples

For the following tests we consider Franke's test function

$$\begin{aligned} f(x,y) &= \frac{3}{4}e^{-1/4((9x-2)^2+(9y-2)^2)} + \frac{3}{4}e^{-(1/49)(9x+1)^2-(1/10)(9y+1)^2} \\ &\quad + \frac{1}{2}e^{-1/4((9x-7)^2+(9y-3)^2)} - \frac{1}{5}e^{-(9x-4)^2-(9y-7)^2}. \end{aligned} \tag{6.10}$$

The range of this function on the grid $\mathcal{G}$ described below is

$$\text{range } f = [0.003280, 1.220000].$$

The graph of Franke's function is shown in Figure 6.2. We choose the set $\mathcal{X}$ of data sites as the grid G_{64} of 8×8 equally spaced points in the unit square $[0,1] \times [0,1]$ of $\mathbb{R}^2$.

For the evaluation and rendering of our test examples we use a grid $\mathcal{G}$ of 30×30 equally spaced points in the unit square. On this grid we compute maximum errors

$$\mu := \|f - \mathcal{Q}f\|_\infty = \max_{\boldsymbol{x}\in\mathcal{G}} |f(\boldsymbol{x}) - \mathcal{Q}f(\boldsymbol{x})|,$$

and mean-square errors

$$\rho := \frac{1}{900}\sum_{\boldsymbol{x}\in\mathcal{G}} |f(\boldsymbol{x}) - \mathcal{Q}f(\boldsymbol{x})|^2,$$

where f is the known test function, and $\mathcal{Q}f$ is a radial basis function approximant to it.

For the results shown in Table 6.1 we have used the multiquadric without any constant added to its expansion. The value of $\alpha = 0.3$, and the tolerance ε was chosen to be 0.01. We compare the knot insertion algorithm (KI) with a version of the knot removal algorithm (KR) which removes certain groups of knots all at once before a new weighting step is performed, rather than re-weighting after the removal of each individual knot. The effect of this is a considerable speedup in execution time with some loss of accuracy (we point out that the numerical experiments were performed in 1996, and therefore the times are only to be used as relative measures of performance). We do not consider this loss of accuracy to be so crucial, since there does not exist a strategy which would prescribe the optimal order in which to remove knots one at a time, either. The knot insertion algorithm is initialized with a single random knot in the unit square, the knot removal algorithm begins with an interpolant to all 64 data values.

Figure 6.2 shows the original function, and Figures 6.3 and 6.4 the approximation via knot insertion and the fit obtained by performing knot removal along with the grid

	μ	ρ	# knots used	time
KI	4.191418e-02	4.630199e-05	52	18 sec
KR	4.1404723-02	4.618084e-05	49	399 sec

Table 6.2: Comparison of adaptive algorithms for Franke's function ($\varepsilon = 0.0001$).

of data sites and the knots used for the fit. The color shading of the graphs of the approximations is such that a dark color reflects small maximum errors, and a light color indicates large errors.

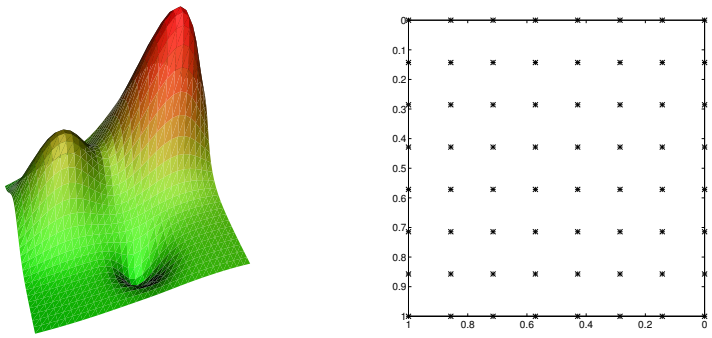

Figure 6.2: (a) Franke's function, and (b) 8×8 grid.

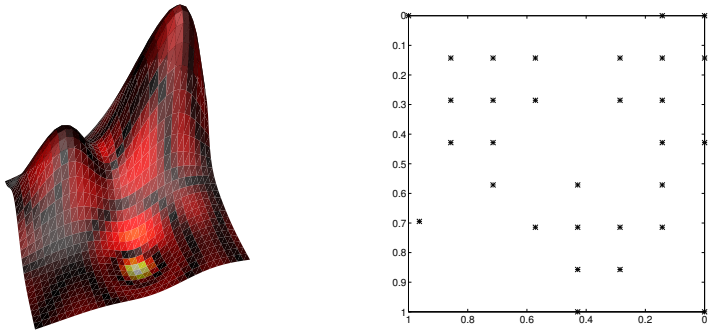

Figure 6.3: (a) Knot insertion fit, and (b) selected knots for $\varepsilon = 0.01$.

Table 6.2 along with Figures 6.5 and 6.6 show the approximation via knot insertion and the fit obtained by performing knot removal using a tighter tolerance of $\varepsilon = 0.0001$ along with the respective knot sets.

It is clearly apparent that the knot insertion algorithm is *much* faster than the one for knot removal. However, we should remark that the implementation of the method used for the solution of the least squares systems uses SVD, which emphasizes this discrepancy even more. The knot removal algorithm for the tighter tolerance performed relatively more exact than that with tolerance $\varepsilon = 0.01$. This can be seen when one observes the size of the groups of knots which were removed. For tolerance

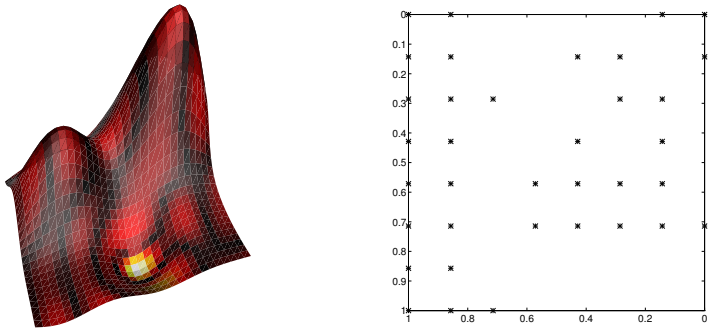

Figure 6.4: (a) Knot removal fit, and (b) selected knots for $\varepsilon = 0.01$.

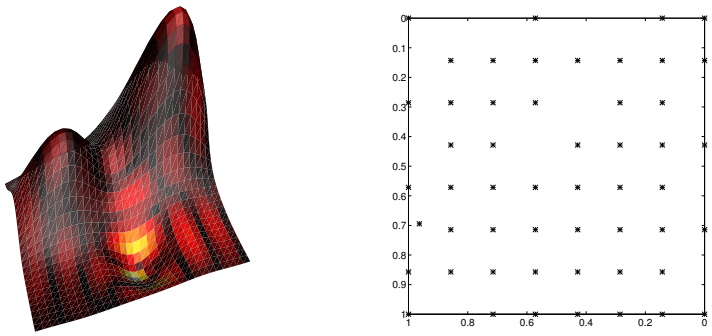

Figure 6.5: (a) Knot insertion fit, and (b) selected knots for $\varepsilon = 0.0001$.

$\varepsilon = 0.01$ first 16 and then 12 knots were removed, whereas for $\varepsilon = 0.0001$ the algorithm was not quite as radical and removed knots in groups of size 3, 4, 4, 4, and 2. This is also evident from the fact that for $\varepsilon = 0.0001$ the KR algorithm ended up using fewer knots than the KI algorithm, which was not the case for $\varepsilon = 0.01$. In general, one should expect the KR algorithm to be able to produce fits with a comparable accuracy to that of a KI fit, but using fewer knots. This should be so since the KR algorithm removes redundancies from an ideal initial fit, whereas the KI algorithm starts with almost no information and has to find a good fit from there.

It is also interesting to note the actual sets of knots chosen by the two methods. They certainly are not the same, but quite a few features are in common.

Remark: Since we developed the native space theory in Chapter 5 it is clear that the above algorithms can also be performed in an analogous way with respect to best approximation in the native space norm. Now, the power function can be used as an error indicator. This idea is pursued in recent papers by Schaback and Wendland [562, 563] to design so-called *greedy algorithms*.

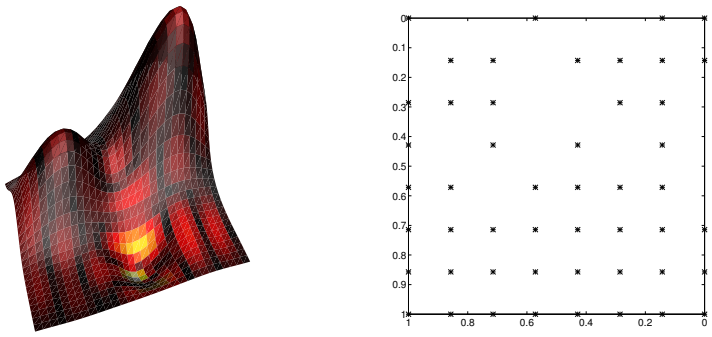

Figure 6.6: (a) Knot removal fit, and (b) selected knots for $\varepsilon = 0.0001$.

Chapter 7

Moving Least Squares Approximation

An alternative to radial basis function interpolation and approximation is the so-called *moving least squares* method. As we will see below, in this method the approximation $\mathcal{P}f$ to f is obtained by solving many (small) linear systems, instead of via solution of a single – but large – linear system as we did in the previous chapters.

To make a connection with the previous chapters we start with the Backus-Gilbert formulation of the moving least squares method since this corresponds to a linearly constrained quadratic minimization problem.

7.1 Moving Least Squares Approximation: The Backus-Gilbert Approach

The connection between the standard moving least squares formulation (to be explained in the next section) and Backus-Gilbert theory was pointed out by Bos and Šalkauskas in [67]. Mathematically, in the Backus-Gilbert approach one considers a quasi-interpolant of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}), \tag{7.1}$$

where $\boldsymbol{f} = [f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)]^T$ represents the given data. From Theorem 5.5.6 we know that the quasi-interpolant that minimizes the point-wise error is given if the *generating functions* Ψ_i are cardinal functions, i.e., $\Psi_i(\boldsymbol{x}_j) = \delta_{ij}$, $i, j = 1, \ldots, N$.

In the moving least squares method one does not attempt to minimize the pointwise error, but instead seeks to find the values of the generating functions $\Psi_i(\boldsymbol{x}) = \Psi(\boldsymbol{x}, \boldsymbol{x}_i)$ by minimizing

$$\frac{1}{2}\sum_{i=1}^{N} \Psi_i^2(\boldsymbol{x})\frac{1}{W(\boldsymbol{x}, \boldsymbol{x}_i)} \tag{7.2}$$

subject to the polynomial reproduction constraints

$$\sum_{i=1}^{N} p(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}) = p(\boldsymbol{x}), \quad \text{for all } p \in \Pi_d^s, \tag{7.3}$$

where Π_d^s is the space of s-variate polynomials of total degree at most d which has dimension $m = \binom{s+d}{d}$.

Remarks:

1. In the above formulation there is no explicit emphasis on nearness of fit as this is implicitly obtained by the quasi-interpolation "ansatz" and the closeness of the generating functions to the pointwise optimal delta functions. This is achieved by the above problem formulation if the $W(\cdot, \boldsymbol{x}_i)$ are weight functions that decrease with distance from the origin. Many of the radial functions used earlier are candidates for the weight functions. However, strict positive definiteness is not required, so that, e.g., (radial or tensor product) B-splines can also be used. The polynomial reproduction constraint is added so that the quasi-interpolant will achieve a desired approximation order. This will become clear in Section 7.6 below.

2. The smoothness functional (7.2) used here is also motivated by practical applications. In the Backus-Gilbert theory which was developed in the context of geophysics (see [17]) it is desired that the generating functions Ψ_i are as close as possible to the ideal cardinal functions (i.e., delta functions). Therefore, one needs to minimize their "spread". The polynomial reproduction constraints correspond to *discrete moment conditions* for the function $\Psi = \Psi(\boldsymbol{x}, \cdot)$.

If we think of $\boldsymbol{x}$ as a fixed (evaluation) point, then we have another constrained quadratic minimization problem of the form discussed in previous chapters. The unknowns are collected in the "coefficient vector" $\Psi(\boldsymbol{x}) = [\Psi(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, \Psi(\boldsymbol{x}, \boldsymbol{x}_N)]^T$. The smoothness functional (7.2)

$$\frac{1}{2}\Psi(\boldsymbol{x})^T Q(\boldsymbol{x})\Psi(\boldsymbol{x})$$

is given via the diagonal matrix

$$Q(\boldsymbol{x}) = \operatorname{diag}\left(\frac{1}{W(\boldsymbol{x}, \boldsymbol{x}_1)}, \ldots, \frac{1}{W(\boldsymbol{x}, \boldsymbol{x}_N)}\right), \tag{7.4}$$

where $W(\cdot, \boldsymbol{x}_i)$ are positive *weight functions* (and thus for any $\boldsymbol{x}$ the matrix $Q(\boldsymbol{x})$ is positive definite).

The linear polynomial reproduction constraint (7.3) can be written in matrix form as

$$A\Psi(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x}),$$

where A is the $m \times N$ matrix with entries $A_{ji} = p_j(\boldsymbol{x}_i)$, $i = 1, \ldots, N$, $j = 1, \ldots, m$, and $\boldsymbol{p} = [p_1, \ldots, p_m]^T$ is a vector that contains a basis for the space Π_d^s of polynomials of degree d.

According to our earlier work we use Lagrange multipliers and then know that (cf. (6.4) and (6.5))

$$\boldsymbol{\lambda}(\boldsymbol{x}) = \left(AQ^{-1}(\boldsymbol{x})A^T\right)^{-1}\boldsymbol{p}(\boldsymbol{x}) \tag{7.5}$$

$$\Psi(\boldsymbol{x}) = Q^{-1}(\boldsymbol{x})A^T\boldsymbol{\lambda}(\boldsymbol{x}). \tag{7.6}$$

Equation (7.5) implies that the Lagrange multipliers are obtained as the solution of a Gram system

$$G(\boldsymbol{x})\boldsymbol{\lambda}(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x}),$$

where the entries of G are the weighted ℓ_2 inner products

$$G_{jk}(\boldsymbol{x}) = \langle p_j, p_k \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} p_j(\boldsymbol{x}_i) p_k(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i), \quad j,k = 1, \ldots, m. \tag{7.7}$$

The special feature here is that the weight varies with the evaluation point $\boldsymbol{x}$.

Two short comments are called for. First, the Gram matrix is symmetric and positive definite since the polynomial basis is linearly independent and the weights are positive. Second, in practice, the polynomials will be represented in shifted form, i.e., centered at the point of evaluation $\boldsymbol{x}$, so that only $p_1(\boldsymbol{x}) \equiv 1 \neq 0$.

Equation (7.6) can be written componentwise, i.e., the generating functions in (7.1) are given by

$$\Psi_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i) \sum_{j=1}^{m} \lambda_j(\boldsymbol{x}) p_j(\boldsymbol{x}_i), \qquad i = 1, \ldots, N.$$

Therefore, once the values of the Lagrange multipliers $\lambda_j(\boldsymbol{x})$, $j = 1, \ldots, N$, have been determined we have explicit formulas for the values of the generating functions. In general, however, finding the Lagrange multipliers involves solving a (small) linear system that changes as soon as $\boldsymbol{x}$ changes.

7.2 Standard Interpretation of MLS Approximation

We now consider the following approximation problem. Assume we are given data values $f(\boldsymbol{x}_i)$, $i = 1, \ldots, N$, on some set $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{R}^s$ of distinct data sites, where f is some (smooth) function, as well as an approximation space $\mathcal{U} = \text{span}\{u_1, \ldots, u_m\}$ (with $m < N$), along with the same weighted ℓ_2 inner product

$$\langle f, g \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_i) W_i(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s \text{ fixed}, \tag{7.8}$$

as introduced above in (7.7). Again, the positive weights W_i, $i = 1, \ldots, N$, depend on the evaluation point $\boldsymbol{x}$. We will interpret the weight functions in a way similar to radial basis functions, i.e., $W_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i)$, with the points $\boldsymbol{x}_i$ coming from the set $\mathcal{X}$.

We now wish to find the best approximation from $\mathcal{U}$ to f at the point $\boldsymbol{x}$ with respect to the norm induced by (7.8). This means we will obtain the approximation (at the point $\boldsymbol{x}$) as

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x}) u_j(\boldsymbol{x}), \tag{7.9}$$

where the coefficients $c_j(\boldsymbol{x})$ are such that

$$\sum_{i=1}^{N} \left[\mathcal{P}f(\boldsymbol{x}_i) - f(\boldsymbol{x}_i)\right]^2 W_i(\boldsymbol{x}) \tag{7.10}$$

is minimized. Due to the definition of the inner product (7.8) whose weight function "moves" with the evaluation point $\boldsymbol{x}$, the coefficients c_j in (7.9) depend also on $\boldsymbol{x}$. As a consequence one has to solve the normal equations

$$\sum_{j=1}^{m} c_j(\boldsymbol{x})\langle u_j, u_k\rangle_{W(\boldsymbol{x})} = \langle f, u_k\rangle_{W(\boldsymbol{x})}, \qquad k = 1, \ldots, m, \tag{7.11}$$

anew each time the evaluation point $\boldsymbol{x}$ is changed. In matrix notation (7.11) becomes

$$G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_u(\boldsymbol{x}), \tag{7.12}$$

with the positive definite Gram matrix $G(\boldsymbol{x}) = \left(\langle u_j, u_k\rangle_{W(\boldsymbol{x})}\right)_{j,k=1}^{m}$, coefficient vector $\boldsymbol{c}(\boldsymbol{x})$ and right-hand side vector $\boldsymbol{f}_u(\boldsymbol{x})$ as in (7.11) all depending on $\boldsymbol{x}$.

In the moving least squares method one usually takes $\mathcal{U}$ to be a space of (multivariate) polynomials, i.e.,

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x})p_j(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s, \tag{7.13}$$

where the $\{p_1, \ldots, p_m\}$ is a basis for the space Π_d^s of s-variate polynomials of degree d.

The Gram system (7.12) now becomes

$$G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x}), \tag{7.14}$$

where the matrix $G(\boldsymbol{x})$ has entries

$$G_{jk}(\boldsymbol{x}) = \langle p_j, p_k\rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} p_j(\boldsymbol{x}_i)p_k(\boldsymbol{x}_i)W(\boldsymbol{x}, \boldsymbol{x}_i), \quad j, k = 1, \ldots, m, \tag{7.15}$$

and the right-hand side vector consists of the projections of the data f onto the basis functions, i.e.,

$$\boldsymbol{f}_p(\boldsymbol{x}) = \left[\langle f, p_1\rangle_{W(\boldsymbol{x})}, \ldots, \langle f, p_m\rangle_{W(\boldsymbol{x})}\right]^T.$$

Remarks:

1. The fact that the coefficients depend on the evaluation point $\boldsymbol{x}$, and thus for every evaluation of $\mathcal{P}f$ a Gram system (with different matrix $G(\boldsymbol{x})$) needs to be solved, initially scared people away from the moving least squares approach. However, one can either choose compactly supported weight functions so that only a few terms are "active" in the sum in (7.15), or even completely avoid the solution of linear systems (see, e.g., [202]).

2. We point out that since we are working with a polynomial basis, the matrix G can also be interpreted as a moment matrix for the weight W. This interpretation is used in the engineering literature (see, e.g., [381]), and also plays an essential role when connecting moving least squares approximation to the more efficient concept of *approximate approximation* [434]. For a discussion of approximate moving least squares approximation see [203, 204, 205, 206].

The connection to the constrained quadratic minimization problems discussed earlier can be seen as follows. We are now minimizing (for fixed $\boldsymbol{x}$)

$$\frac{1}{2}\boldsymbol{c}^T(\boldsymbol{x})G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) - \boldsymbol{\mu}^T(\boldsymbol{x})\left[G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) - AQ^{-1}(\boldsymbol{x})\boldsymbol{f}\right], \tag{7.16}$$

where $G(\boldsymbol{x})$ is the Gram matrix (7.7), $Q(\boldsymbol{x})$ the diagonal matrix of weight functions (7.4) and A the matrix of polynomials used earlier. The term $AQ^{-1}(\boldsymbol{x})\boldsymbol{f}$ corresponds to the right-hand side vector $\boldsymbol{f}_p(\boldsymbol{x})$ of (7.14). The solution of the linear system resulting from the minimization problem (7.16) gives us

$$\begin{aligned}\boldsymbol{\mu}(\boldsymbol{x}) &= \left(G(\boldsymbol{x})G^{-1}(\boldsymbol{x})G^T(\boldsymbol{x})\right)^{-1} AQ^{-1}(\boldsymbol{x})\boldsymbol{f} = G^{-T}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f}\\ \boldsymbol{c}(\boldsymbol{x}) &= G^{-1}(\boldsymbol{x})G^T(\boldsymbol{x})\boldsymbol{\mu}(\boldsymbol{x}) = \boldsymbol{\mu}(\boldsymbol{x})\end{aligned}$$

so that – as in the case of radial basis function interpolation – by solving only the Gram system $G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x})$ we automatically minimize the functional

$$\begin{aligned}\boldsymbol{c}^T(\boldsymbol{x})G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) &= \sum_{j=1}^{m}\sum_{k=1}^{m} c_j(\boldsymbol{x})c_k(\boldsymbol{x})G_{jk}(\boldsymbol{x})\\ &= \sum_{j=1}^{m}\sum_{k=1}^{m} c_j(\boldsymbol{x})c_k(\boldsymbol{x})\langle p_j, p_k\rangle_{W(\boldsymbol{x})}\end{aligned}$$

which should be interpreted as the native space norm of the approximant $\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x})p_j(\boldsymbol{x})$.

7.3 A Dual Representation for the Standard Approach

We now know that on the one hand (from the Backus-Gilbert formulation)

$$G(\boldsymbol{x})\boldsymbol{\lambda}(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x}) \quad \Longleftrightarrow \quad \boldsymbol{\lambda}(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})\boldsymbol{p}(\boldsymbol{x}). \tag{7.17}$$

By taking multiple right-hand sides $\boldsymbol{p}(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathcal{X}$ we get

$$\Lambda = G^{-1}(\boldsymbol{x})A, \tag{7.18}$$

where the $m \times N$ matrix Λ has entries $\Lambda_{ji} = \lambda_j(\boldsymbol{x}_i)$.

The standard formulation, on the other hand, gives us

$$G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x}) \quad \Longleftrightarrow \quad \boldsymbol{c}(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})\boldsymbol{f}_p(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f} \tag{7.19}$$

with

$$\boldsymbol{f}_p(\boldsymbol{x}) = \left[\langle f, p_1\rangle_{W(\boldsymbol{x})}, \ldots, \langle f, p_m\rangle_{W(\boldsymbol{x})}\right]^T = AQ^{-1}(\boldsymbol{x})\boldsymbol{f}$$

as above. By combining (7.18) with (7.19) we get

$$\boldsymbol{c}(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f} = \Lambda Q^{-1}(\boldsymbol{x})\boldsymbol{f} = \boldsymbol{f}_\lambda(\boldsymbol{x}),$$

where $\boldsymbol{f}_\lambda(\boldsymbol{x})$ is defined analogously to $\boldsymbol{f}_p(\boldsymbol{x})$. Componentwise this gives us

$$c_j(\boldsymbol{x}) = \langle f, \lambda_j \rangle_{W(\boldsymbol{x})}, \qquad j = 1, \ldots, m,$$

and therefore,

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} \langle f, \lambda_j \rangle_{W(\boldsymbol{x})} p_j(\boldsymbol{x}). \tag{7.20}$$

It is also possible to formulate the moving least squares method by using the Lagrange multipliers of the Backus-Gilbert approach as basis functions for the approximation space $\mathcal{U}$. Then, using the same argumentation as above, we end up with

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} d_j(\boldsymbol{x}) \lambda_j(\boldsymbol{x}) \tag{7.21}$$

with

$$d_j(\boldsymbol{x}) = \langle f, p_j \rangle_{W(\boldsymbol{x})}, \qquad j = 1, \ldots, m.$$

Remarks:

1. The Lagrange multipliers form a basis that is *dual* to the polynomials. In particular one can show that for any $\boldsymbol{x} \in \mathcal{X}$

   $$\langle \lambda_j, p_k \rangle_{W(\boldsymbol{x})} = \delta_{jk}, \qquad j, k = 1, \ldots, m.$$

   This shows that we have two sets of basis functions that are *bi-orthogonal* on the set $\mathcal{X}$.

2. Note that the expansions (7.20) and (7.21) are generalizations of (finite) eigenfunction or Fourier series expansions.

7.4 Equivalence of Our Approaches to Moving Least Squares Approximation

We now show that the two main approaches to the moving least squares method described above are equivalent, i.e., we show that $\mathcal{P}f(\boldsymbol{x})$ computed via (7.1) and (7.13) are the same. The approximant (7.1) in the Backus-Gilbert "ansatz" is of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{i=1}^{N} f(\boldsymbol{x}_i) \Psi_i(\boldsymbol{x}) = \Psi^T(\boldsymbol{x}) \boldsymbol{f},$$

where as before $\Psi(\boldsymbol{x}) = [\Psi(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, \Psi(\boldsymbol{x}, \boldsymbol{x}_N)]^T$ and $\boldsymbol{f} = [f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)]^T$. The standard moving least squares formulation (7.13), on the other hand, establishes $\mathcal{P}f(\boldsymbol{x})$ in the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x}) p_j(\boldsymbol{x}) = \boldsymbol{p}^T(\boldsymbol{x}) \boldsymbol{c}(\boldsymbol{x}),$$

where $\boldsymbol{p} = [p_1, \ldots, p_m]^T$ and $\boldsymbol{c}(\boldsymbol{x}) = [c_1(\boldsymbol{x}), \ldots, c_m(\boldsymbol{x})]^T$.

In (7.16) we wrote the normal equations for the standard approach as

$$G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = AQ^{-1}(\boldsymbol{x})\boldsymbol{f}$$

which implies

$$\boldsymbol{c}(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f}.$$

Thus, using the standard approach, we get

$$\mathcal{P}f(\boldsymbol{x}) = \boldsymbol{p}^T(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{p}^T(\boldsymbol{x})G^{-1}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f}. \tag{7.22}$$

For the Backus-Gilbert approach we derived (see (7.5) and (7.6))

$$\begin{aligned} \boldsymbol{\lambda}(\boldsymbol{x}) &= G^{-1}(\boldsymbol{x})\boldsymbol{p}(\boldsymbol{x}) \\ \Psi(\boldsymbol{x}) &= Q^{-1}(\boldsymbol{x})A^T\boldsymbol{\lambda}(\boldsymbol{x}), \end{aligned}$$

where $G(\boldsymbol{x}) = AQ^{-1}(\boldsymbol{x})A^T$ (see (7.7) or (7.15)). Therefore, we now obtain

$$\mathcal{P}f(\boldsymbol{x}) = \Psi^T(\boldsymbol{x})\boldsymbol{f} = \left[Q^{-1}(\boldsymbol{x})A^TG^{-1}(\boldsymbol{x})\boldsymbol{p}(\boldsymbol{x})\right]^T\boldsymbol{f}$$

which, by the symmetry of $Q(\boldsymbol{x})$ and $G(\boldsymbol{x})$, is the same as (7.22).

Remarks:

1. The equivalence of the two approaches shows that the moving least squares approximant has all of the following properties:

   - It reproduces any polynomial of degree at most d in s variables exactly.
   - It produces the best locally weighted least squares fit.
   - Viewed as a quasi-interpolant, the generating functions Ψ_i are as close as possible to the optimal cardinal basis functions in the sense that (7.2) is minimized.
   - Since polynomials are infinitely smooth, either of the representations of $\mathcal{P}f$ shows that its smoothness is determined by the smoothness of the weight function(s) $W_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i)$.

2. In particular, the standard moving least squares method will reproduce the polynomial basis functions $p_1, \ldots, p_m$ even though this is not explicitly enforced by the minimization (solution of the normal equations). Moreover, the more general "ansatz" with approximation space $\mathcal{U}$ allows us to build moving least squares approximations that also reproduce any other function that is included in $\mathcal{U}$. This can be very beneficial for the solution of partial differential equations with known singularities (see, e.g., the papers [16] by Babuška and Melenk, and [49] by Belytschko and co-authors).

By also considering the dual expansion (7.21) we have three alternative representations for the moving least squares quasi-interpolant. This is summarized in the following theorem.

Theorem 7.4.1 *Let $f : \Omega \to \mathbb{R}$ be some function whose values on the set of points $\mathcal{X} = \{\boldsymbol{x}_i\}_{i=1}^N \subset \mathbb{R}^s$ are given as data. Let $p_1, \ldots, p_m$ be a basis for Π_d^s, let $\{W(\cdot, \boldsymbol{x}_i)\}_{i=1}^N$ be a set of positive weight functions centered at the points of $\mathcal{X}$, and let λ_j, $j = 1, \ldots, m$, be the Lagrange multipliers defined by (7.5). Furthermore, consider the generating functions*

$$\Psi_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i) \sum_{j=1}^m \lambda_j(\boldsymbol{x}) p_j(\boldsymbol{x}_i), \qquad i = 1, \ldots, N.$$

The best local least squares approximation to f on $\mathcal{X}$ in the sense of (7.10) is given by

$$\begin{aligned} \mathcal{P}f(\boldsymbol{x}) &= \sum_{j=1}^m \langle f, \lambda_j \rangle_{W(\boldsymbol{x})} p_j(\boldsymbol{x}) \\ &= \sum_{j=1}^m \langle f, p_j \rangle_{W(\boldsymbol{x})} \lambda_j(\boldsymbol{x}) \\ &= \sum_{i=1}^N f(\boldsymbol{x}_i) \Psi_i(\boldsymbol{x}). \end{aligned}$$

7.5 Examples

7.5.1 Shepard's Method

The moving least squares method in the case $m = 1$ with $p_1(\boldsymbol{x}) \equiv 1$ is known to yield Shepard's method [578]. In the statistics literature Shepard's method is known as a *kernel method* (see, e.g., the papers from the 1950s and 60s [534, 501, 476, 623]). Using our notation we have

$$\mathcal{P}f(\boldsymbol{x}) = c_1(\boldsymbol{x}).$$

The Gram "matrix" consists of only one element

$$G(\boldsymbol{x}) = \langle p_1, p_1 \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^N W(\boldsymbol{x}, \boldsymbol{x}_i)$$

so that

$$G(\boldsymbol{x}) \boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x})$$

implies

$$c_1(\boldsymbol{x}) = \frac{\displaystyle\sum_{i=1}^N f(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i)}{\displaystyle\sum_{i=1}^N W(\boldsymbol{x}, \boldsymbol{x}_i)}.$$

The dual basis is defined by

$$G(\boldsymbol{x}) \boldsymbol{\lambda}(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x})$$

so that

$$\lambda_1(\boldsymbol{x}) = \frac{1}{\sum_{i=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_i)},$$

and

$$\mathcal{P}f(\boldsymbol{x}) = d_1(\boldsymbol{x})\lambda_1(\boldsymbol{x}) \tag{7.23}$$

with

$$d_1(\boldsymbol{x}) = \langle f, p_1 \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} f(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i).$$

The generating functions are defined as

$$\Psi_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i)\lambda_1(\boldsymbol{x})p_1(\boldsymbol{x}_i) = \frac{W(\boldsymbol{x}, \boldsymbol{x}_i)}{\sum_{i=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_i)}.$$

This gives rise to the well-known quasi-interpolation formula for Shepard's method

$$\begin{aligned} \mathcal{P}f(\boldsymbol{x}) &= \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}) \\ &= \sum_{i=1}^{N} f(\boldsymbol{x}_i) \frac{W(\boldsymbol{x}, \boldsymbol{x}_i)}{\sum_{k=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_k)}. \end{aligned}$$

Of course this is the same as the basis expansion $c_1(\boldsymbol{x})$ and the dual expansion (7.23).

We should now have bi-orthogonality of the basis and dual basis, i.e.,

$$\langle \lambda_1, p_1 \rangle_{W(\boldsymbol{x})} = 1.$$

Indeed

$$\begin{aligned} \langle \lambda_1, p_1 \rangle_{W(\boldsymbol{x})} &= \sum_{i=1}^{m} \lambda_1(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i) \\ &= \sum_{i=1}^{N} \frac{W(\boldsymbol{x}, \boldsymbol{x}_i)}{\sum_{k=1}^{N} W(\boldsymbol{x}_i, \boldsymbol{x}_k)}, \end{aligned}$$

and this equals 1 if we restrict $\boldsymbol{x}$ to be an element of the set $\mathcal{X}$.

7.5.2 Plots of Basis-Dual Basis Pairs

We also illustrate the moving least squares basis functions, dual basis functions and generating functions for a one-dimensional example with $\mathcal{X}$ being the set of 13 equally spaced points in $[-5, 5]$. We take $m = 2$, i.e., we consider the case that ensures

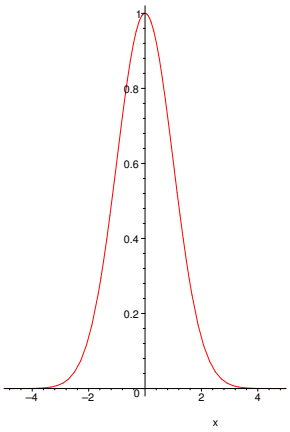

Figure 7.1: Plot of Gaussian weight function centered at $\boldsymbol{x}_7 = 0$.

reproduction of quadratic polynomials. The weight function is taken to be a standard Gaussian as depicted in Figure 7.1.

The three basis polynomials $p_1(x) = 1$, $p_2(x) = x$, and $p_3(x) = x^2$ are shown in Figure 7.2, whereas the dual basis functions λ_1, λ_2, and λ_3 are displayed in Figure 7.3. The figure shows that, except for the boundary effects caused by the finite interval, these functions resemble a quadratic, linear and constant polynomial.

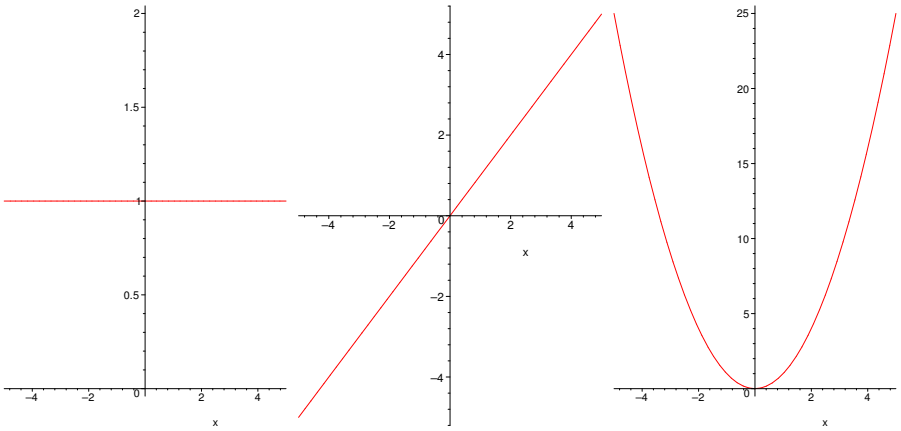

Figure 7.2: Plot of three polynomial basis functions for moving least squares approximation.

In Figure 7.4 we plot one of the generating functions (centered at $\boldsymbol{x}_7 = 0$) along with an approximate moving least squares generating function of the form

$$\Psi(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{\sqrt{\sigma\pi}} \left(\frac{3}{2} - \frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{\sigma} \right) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{\sigma}}$$

with scale parameter σ as derived in [205].

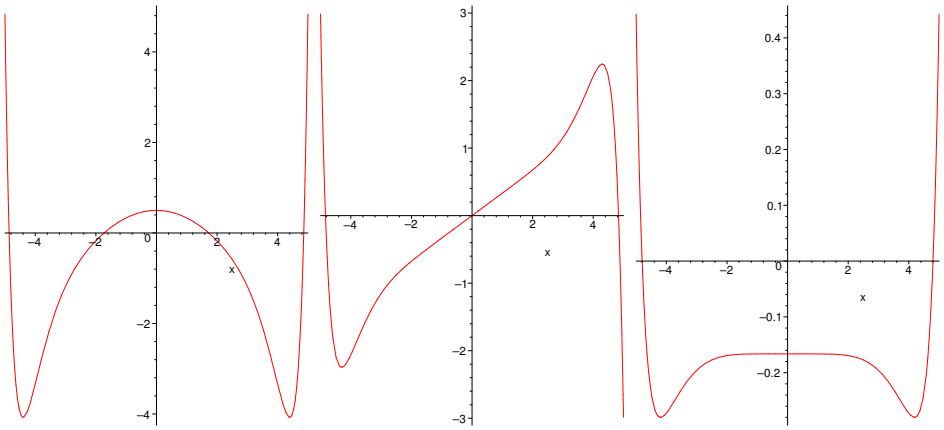

Figure 7.3: Plot of three dual basis functions for moving least squares approximation.

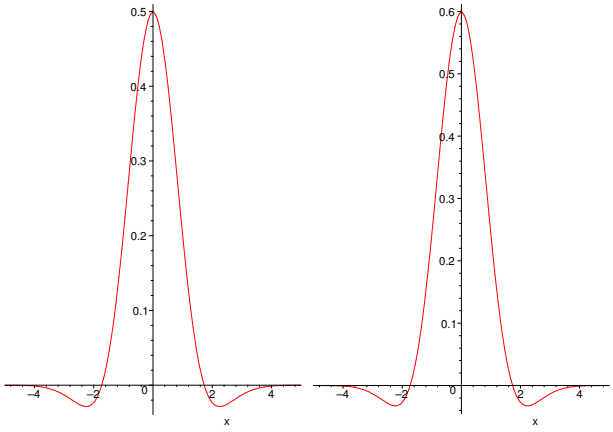

Figure 7.4: Plot of moving least squares generating function (left) and approximate generating function (right) centered at $\boldsymbol{x}_7 = 0$.

7.6 Approximation Order of Moving Least Squares

Since the moving least squares approximants can be written as quasi-interpolants, we can use standard techniques to derive their point-wise error estimates. The standard argument proceeds as follows. Let f be a given (smooth) function that generates the data, i.e., $f_1 = f(\boldsymbol{x}_1), \ldots, f_N = f(\boldsymbol{x}_N)$, and let p be an arbitrary polynomial. Moreover, assume that the moving least squares approximant is given in the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x})$$

with the generating functions Ψ_i satisfying the polynomial reproduction property

$$\sum_{i=1}^{N} p(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}) = p(\boldsymbol{x}), \quad \text{for all } p \in \Pi_d^s,$$

as described at the beginning of this chapter. Then, due to the polynomial reproduction property of the generating functions,

$$\begin{aligned} |f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| &\leq |f(\boldsymbol{x}) - p(\boldsymbol{x})| + |p(\boldsymbol{x}) - \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x})| \\ &= |f(\boldsymbol{x}) - p(\boldsymbol{x})| + |\sum_{i=1}^{N} p(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}) - \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x})| \\ &\leq |f(\boldsymbol{x}) - p(\boldsymbol{x})| + \sum_{i=1}^{N} |p(\boldsymbol{x}_i) - f(\boldsymbol{x}_i)||\Psi_i(\boldsymbol{x})| \\ &\leq \|f - p\|_\infty \left[1 + \sum_{i=1}^{N} |\Psi_i(\boldsymbol{x})|\right]. \end{aligned} \tag{7.24}$$

We see that in order to refine the error estimate we now have to answer two questions:

- How well do polynomials approximate f? This will be done with standard Taylor expansions.
- Are the generating functions bounded? The expression $\sum_{i=1}^{N} |\Psi_i(\boldsymbol{x})|$ is known as the *Lebesgue function*, and finding a bound for the Lebesgue function is the main task that remains.

By taking the polynomial p above to be the Taylor polynomial for f at $\boldsymbol{x}$ of total degree d, the remainder term immediately yields an estimate of the form

$$\begin{aligned} \|f - p\|_\infty &\leq C_1 h^{d+1} \max_{\boldsymbol{x}\in\Omega} |D^{\boldsymbol{\alpha}} f(\boldsymbol{x})|, \qquad |\boldsymbol{\alpha}| = d+1, \\ &= C_1 h^{d+1} |f|_{d+1}, \end{aligned} \tag{7.25}$$

where we have used the abbreviation

$$|f|_{d+1} = \max_{\boldsymbol{x}\in\Omega} |D^{\boldsymbol{\alpha}} f(\boldsymbol{x})|, \qquad |\boldsymbol{\alpha}| = d+1.$$

Thus, if we can establish a uniform bound for the Lebesgue function, then (7.24) and (7.25) will result in

$$|f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| \leq C h^{d+1} |f|_{d+1}$$

which shows that moving least squares approximation with polynomial reproduction of degree d has approximation order $\mathcal{O}(h^{d+1})$.

For Shepard's method, i.e., moving least squares approximation with constant reproduction (i.e., $m = 1$ or $d = 0$), we saw above that the generating functions are of the form

$$\Psi_i(\boldsymbol{x}) = \frac{W(\boldsymbol{x}, \boldsymbol{x}_i)}{\sum_{j=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_j)}$$

and therefore the Lebesgue function admits the uniform bound

$$\sum_{i=1}^{N} |\Psi_i(\boldsymbol{x})| = 1,$$

This shows that Shepard's method has approximation order $\mathcal{O}(h)$.

Bounding the Lebesgue function in the general case is more involved and is the subject of the papers [378] by Levin and [632] by Wendland. This results in approximation order $\mathcal{O}(h^{d+1})$ for a moving least squares method that reproduces polynomials of degree d. In both papers the authors assumed that the weight function is compactly supported, and that the support size is scaled proportional to the fill distance. However, similar estimates should be possible if the weight function only decays fast enough (see, e.g., the survey by de Boor [60]).

Aside from this consideration, the choice of weight function W does not play a role in determining the approximation order of the moving least squares method. As noted earlier, it only determines the smoothness of $\mathcal{P}f$. For example, in the paper [146] from the statistics literature on local regression the authors state that often "the choice [of weight function] is not too critical", and the use of the so-called *tri-cube*

$$W(\boldsymbol{x}, \boldsymbol{x}_i) = (1 - \|\boldsymbol{x} - \boldsymbol{x}_i\|^3)_+^3, \qquad \boldsymbol{x} \in \mathbb{R}^s,$$

is suggested. Of course, many other weight functions such as (radial) B-splines or any of the (bell-shaped) radial basis functions studied earlier can also be used. If the weight function is compactly supported, then the generating functions Ψ_i will be so, too. This leads to computationally efficient methods since the Gram matrix $G(\boldsymbol{x})$ will be sparse.

An interesting question is also the size of the support of the different local weight functions. Obviously, a fixed support size for all weight functions is possible. However, this will cause serious problems as soon as the data are not uniformly distributed. Therefore, in the arguments in [378] and [632] the assumption is made that the data are at least quasi-uniformly distributed. Another choice for the support size of the individual weight functions is based on the number of nearest neighbors, i.e., the support size is chosen so that each of the local weight functions contains the same number of centers in its support. A third possibility is suggested by Schaback [556]. He proposes to use another moving least squares approximation based on (equally spaced) auxiliary points to determine a smooth function δ so that at each evaluation point $\boldsymbol{x}$ the radius of the support of the weight function is given by $\delta(\boldsymbol{x})$. However, convergence estimates for these latter two choices do not exist.

Sobolev error estimates are provided for moving least squares approximation with compactly supported weight functions in [7]. The rates obtained in that paper are not in terms of the fill distance but instead in terms of the support size R of the weight function. Moreover, it is assumed that for general s and $m = \binom{s+d}{d}$ the local Lagrange functions are bounded. As mentioned above, this is the hard part, and such bounds are only provided in the case $s = 2$ with $d = 1$ and $d = 2$ in [7]. However, if combined with the general bounds for the Lebesgue function provided by Wendland the paper [7] yields the following estimates:

$$|f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| \leq CR^{d+1}|f|_{d+1}$$

but also

$$|\nabla(f - \mathcal{P}f)(\boldsymbol{x})| \leq CR^d |f|_{d+1}.$$

In the weaker L_2-norm we have

$$\|f - \mathcal{P}f\|_{L_2(B_j \cap \Omega)} \leq CR^{d+1} |f|_{W_2^{d+1}(B_j \cap \Omega)}$$

and

$$\|\nabla(f - \mathcal{P}f)\|_{L_2(B_j \cap \Omega)} \leq CR^d |f|_{W_2^{d+1}(B_j \cap \Omega)},$$

where the balls B_j provide a finite cover of the domain Ω, i.e., $\Omega \subseteq \bigcup_j B_j$, and the number of overlapping balls is bounded.

Remarks:

1. In the statistics literature the moving least squares idea is known as local (polynomial) regression. There is a book by Fan and Gijbels [186] and a review article by Cleveland and Loader [146] according to which the basic ideas of local regression can be traced back at least to work of Gram [267], Woolhouse [648], and De Forest [148, 149] from the 1870s and 1880s.

2. In particular, in the statistics literature one learns that the use of least squares approximation is justified when the data $f_1, \ldots, f_N$ are normally distributed, whereas, if the noise in the data is not Gaussian, then other criteria should be used. See, e.g., the survey article [146] for more details.

3. The general moving least squares method first appeared in the approximation theory literature in a paper by Lancaster and Šalkauskas [358] who also pointed out the connection to earlier (more specialized) work by Shepard [578] and McLain [436].

4. Early error estimates for some special cases were provided by Farwig in [188, 189].

Chapter 8

Some Issues Related to Practical Implementations

In this chapter we will collect some information about issues that are important for the practical use of radial basis function and moving least squares methods. These issues include stability and conditioning of radial basis function interpolants, the *trade-off principle* which explains the trade-off between achievable convergence rates and numerical stability or efficiency, as well as algorithms for fast solution and evaluation of radial basis interpolants and moving least squares approximants.

8.1 Stability and Conditioning of Radial Basis Function Interpolants

A standard criterion to measure the numerical stability of an approximation method is its condition number. In particular, for radial basis function interpolation we need to look at the condition number of the interpolation matrix A with entries $A_{ij} = \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j)$. For any matrix A the ℓ_2-condition number of A is given by

$$\mathrm{cond}(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_{\max}}{\sigma_{\min}},$$

where $\sigma_{\max}$ and $\sigma_{\min}$ are the largest and smallest singular values of A. If we concentrate on positive definite matrices A, then we can also take the ratio

$$\frac{\lambda_{\max}}{\lambda_{\min}}$$

of largest and smallest eigenvalues as an indicator for the condition number of A.

What do we know about these eigenvalues? First, Gershgorin's Theorem says that

$$|\lambda_{\max} - A_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{N} |A_{ij}|.$$

Therefore,

$$\lambda_{\max} \leq N \max_{i,j=1,\ldots,N} |A_{ij}| = N \max_{\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{X}} \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j),$$

which, since Φ is strictly positive definite, becomes

$$\lambda_{\max} \leq N\Phi(\mathbf{0})$$

by the properties of positive definite functions listed in Theorem 1.2.6. Now, as long as the data are not too wildly distributed, N will grow as $h_{\mathcal{X},\Omega}^{-s}$ which is acceptable. Therefore, the main work in establishing a bound for the condition number of A lies in finding lower bounds for $\lambda_{\min}$ (or correspondingly upper bounds for the norm of the inverse $\|A^{-1}\|_2$). This is the subject of several papers by Ball, Narcowich, Sivakumar and Ward [19, 479, 481, 482, 483] who make use of a result by Ball [18] on eigenvalues of distance matrices. Ball's result follows from the Rayleigh quotient, which gives the smallest eigenvalue of a positive definite matrix as

$$\lambda_{\min} = \inf_{\boldsymbol{c}\in\mathbb{R}^N\setminus\mathbf{0}} \frac{\boldsymbol{c}^T A\boldsymbol{c}}{\boldsymbol{c}^T\boldsymbol{c}}.$$

This leads to the following bound for the norm of the inverse of A.

Lemma 8.1.1 *Let $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N$, be distinct points in $\mathbb{R}^s$ and let $\Phi:\mathbb{R}^s\to\mathbb{R}$ be either strictly positive definite or strictly conditionally negative definite of order one with $\Phi(\mathbf{0})\leq 0$. Also, let A be the interpolation matrix with entries $A_{ij}=\Phi(\boldsymbol{x}_i-\boldsymbol{x}_j)$. If the inequality*

$$\sum_{i=1}^N\sum_{j=1}^N c_i c_j A_{ij} \geq \theta\|\boldsymbol{c}\|_2^2$$

is satisfied whenever the components of $\boldsymbol{c}$ satisfy $\sum_{j=1}^N c_j = 0$, then

$$\|A^{-1}\|_2 \leq \theta^{-1}.$$

Note that for positive definite matrices the Rayleigh quotient implies $\theta=\lambda_{\min}$ which shows why lower bounds on the smallest eigenvalue correspond to to upper bounds on the norm of the inverse of A. In order to obtain the bound for conditionally negative matrices the Courant-Fischer Theorem 3.1.2 needs to be employed.

Narcowich and Ward establish bounds on the norm of the inverse of A in terms of the *separation distance* of the data sites

$$q_{\mathcal{X}} = \frac{1}{2}\min_{i\neq j}\|\boldsymbol{x}_i-\boldsymbol{x}_j\|_2.$$

We can picture $q_{\mathcal{X}}$ as the radius of the largest ball that can be placed around every point in $\mathcal{X}$ such that no two balls overlap (see Figure 8.1).

The derivation of these bounds is rather technical, and for details we refer to either the original papers by Narcowich, Ward and co-workers, the more recent paper [557] by Schaback (who uses a slightly simpler strategy), or Wendland's book [634]. We now list several bounds as derived in [634].

Examples: In the examples below the explicit constants

$$M_s = 12\left(\frac{\pi\Gamma^2(\frac{s+2}{2})}{9}\right)^{1/(s+1)} \leq 6.38s \quad\text{and}\quad C_s = \frac{1}{2\Gamma(\frac{s+2}{2})}\left(\frac{M_s}{\sqrt{8}}\right)^s$$

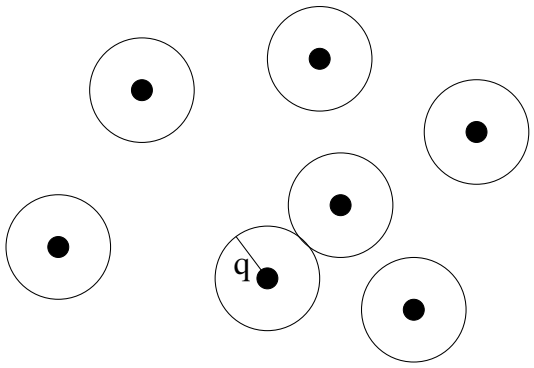

Figure 8.1: The separation distance $q_{\mathcal{X}}$ for a set of data sites in $\mathbb{R}^2$.

are used. The upper bound for M_s can be obtained using Stirling's formula (see, e.g., [634]).

1. For Gaussians $\Phi(\boldsymbol{x}) = e^{-\alpha\|\boldsymbol{x}\|^2}$ one obtains

$$\lambda_{\min} \geq C_s(2\alpha)^{-s/2} e^{-40.71s^2/(q_{\mathcal{X}}^2\alpha)} q_{\mathcal{X}}^{-s}.$$

2. For (inverse) multiquadrics $\Phi(\boldsymbol{x}) = \left(\|\boldsymbol{x}\|^2 + \alpha^2\right)^{\beta}$, $\beta \in \mathbb{R} \setminus \mathbb{N}_0$ one obtains

$$\lambda_{\min} \geq C(\alpha, \beta, s) q_{\mathcal{X}}^{\beta - \frac{s}{2} + \frac{1}{2}} e^{-2\alpha M_s/q_{\mathcal{X}}}$$

   with another explicitly known constant $C(\alpha, \beta, s)$.

3. For thin plate splines $\Phi(\boldsymbol{x}) = (-1)^{k+1}\|\boldsymbol{x}\|^{2k}\log\|\boldsymbol{x}\|$, $k \in \mathbb{N}$, one obtains

$$\lambda_{\min} \geq C_s c_k (2M_s)^{-s-2k} q_{\mathcal{X}}^{2k}$$

   with another explicitly known constant c_k.

4. For the powers $\Phi(\boldsymbol{x}) = (-1)^{\lceil \beta/2 \rceil}\|\boldsymbol{x}\|^{\beta}$, $\beta > 0$, $\beta \notin 2\,\mathbb{N}$, one obtains

$$\lambda_{\min} \geq C_s c_\beta (2M_s)^{-s-\beta} q_{\mathcal{X}}^{\beta}$$

   with another explicitly known constant c_β.

5. For the compactly supported functions $\Phi_{s,k}(\boldsymbol{x}) = \varphi_{s,k}(\|\boldsymbol{x}\|)$ of Section 4 one obtains

$$\lambda_{\min} \geq C(s,k) q_{\mathcal{X}}^{2k+1}$$

   with a constant $C(s,k)$ depending on s and k.

By providing matching *lower* bounds for $\|A^{-1}\|_2$ Schaback [547] showed that the upper bounds on the norm of the inverse obtained earlier by Narcowich, Ward and others are near optimal.

For the infinitely smooth functions of Examples 1 and 2 we see that, for a fixed shape parameter α, the lower bound for $\lambda_{\min}$ goes exponentially to zero, and therefore the

condition number of the interpolation matrix A grows exponentially, as the separation distance $q_{\mathcal{X}}$ decreases. This shows that, if one adds more interpolation points in order to improve the accuracy of the interpolant (within the same domain Ω), then the problem becomes increasingly ill-conditioned. Of course one would always expect this to happen, but here the ill-conditioning grows primarily due to the decrease in the separation distance $q_{\mathcal{X}}$, and not to the increase in the number N of data points. We will come back to this observation when we discuss a possible change of basis in Section 8.4.

On the other hand, if one keeps the number of points (or at least the separation distance) fixed and instead increases (reduces) the value of α for Gaussians (multiquadrics), then the condition number of A is improved. This corresponds to the *stationary* approximation setting (which we did not discuss in detail earlier). In this case it is possible to show that the upper bound for the error estimate increases, i.e., the accuracy of the interpolant deteriorates. Conversely, one can attempt to improve the accuracy of a radial basis function interpolant by decreasing (increasing) α for Gaussians (multiquadrics). However, this is only possible at the cost of numerical instability (ill-conditioning of A). This is to be expected since for small (large) values of α the Gaussians (multiquadrics) more and more resemble a constant function, and therefore the rows (as well as columns) of the matrix A become more and more alike, so that the matrix becomes almost singular – even for well separated data sites. In the literature this phenomenon has been referred to as *trade-off* or (*uncertainty*) *principle* (see, e.g., the paper [549] by Schaback). The relation between the power function (as part of the upper bound on the approximation error) and minimal eigenvalue (as part of the measure of the condition number) is derived below.

Remarks:

1. This trade-off has led a number of people to search for an "optimal" value of the shape parameter, i.e., a value that yields maximal accuracy, while still maintaining numerical stability. For example, in his original work on (inverse) multiquadric interpolation in $\mathbb{R}^2$ Hardy [286] suggested using $\alpha = 0.815d$, where $d = \frac{1}{N}\sum_{i=1}^{N} d_i$, and d_i is the distance from $\boldsymbol{x}_i$ to its nearest neighbor. Later Franke [231] suggested using $\alpha = 1.25\frac{D}{\sqrt{N}}$, where D is the diameter of the smallest circle containing all data points. Foley [221] based his strategy for finding a good value for α on the observation that that good value was similar for multiquadrics and inverse multiquadrics. Other studies were reported in [102] and [103]. A more recent algorithm was proposed by Rippa in [531]. He suggests a variant of cross validation known as "leave-one-out" cross validation. This method is rather popular in the statistics literature where it is also known as PRESS (Predictive REsidual Sum of Squares). In this algorithm an "optimal" value of α is selected by minimizing the least squares error for a fit based on the data for which one of the centers was "left out". A similar strategy was proposed earlier in [262] for the solution of elliptic partial differential equations via the dual reciprocity method based on multiquadric interpolation.

2. More recently, Fornberg and co-workers have investigated the dependence of the stability on the values of the shape parameter α in a series of papers (e.g., [159, 228, 361]). On the one hand, they suggest a way of stably computing very accurate (inverse) multiquadric and Gaussian interpolants (with extreme values

of α) by using a complex Contour-Padé integration algorithm. This algorithm is rather expensive, and so far only applicable for problems involving no more than 100 centers. On the other hand, Fornberg and co-workers as well as Schaback [558] have shown that in the limiting case of the shape parameter α, i.e., with very "flat" basis functions, the infinitely smooth radial basis function interpolants approach multivariate polynomial interpolants. Therefore, Fornberg and his co-workers suggest using radial basis functions as a generalization of spectral methods (applicable also in the case of scattered data) for the numerical solution of partial differential equations. This approach was also taken recently by Sarra [543].

We can observe that for the functions with finite smoothness (as in Examples 3–5) the lower bounds for $\|A^{-1}\|_2$ are of the same order as the upper bounds for the power function. The following theorem therefore shows that the order of both of these bounds is optimal. The theorem also provides the foundation for the trade-off principle referred to above.

Theorem 8.1.2 *Let $u_j^*(\boldsymbol{x})$, $j = 1, \ldots, N$, denote the cardinal functions for interpolation with the strictly positive definite function Φ as explained in Chapter 5, and let $\lambda_{\boldsymbol{x}}$ be the minimal eigenvalue of the $(N+1) \times (N+1)$ matrix $A_{\boldsymbol{x}}$ with entries $A_{\boldsymbol{x},ij} = \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j)$, $i, j = 0, 1, \ldots, N$, where we have added the evaluation point to the set of centers, i.e., $\boldsymbol{x}_0 = \boldsymbol{x}$. Then*

$$\lambda_{\boldsymbol{x}}^{-1} P_{\Phi,\mathcal{X}}^2(\boldsymbol{x}) \geq 1 + \sum_{j=1}^{N} \left[u_j^*(\boldsymbol{x})\right]^2.$$

Proof: The definition of the power function (with standard interpolation matrix A, and right-hand side vector $\boldsymbol{b}(\boldsymbol{x}) = [\Phi(\boldsymbol{x} - \boldsymbol{x}_1), \ldots, \Phi(\boldsymbol{x} - \boldsymbol{x}_N)]^T$, see Section 5.3) yields

$$\begin{aligned} P_{\Phi,\mathcal{X}}^2(\boldsymbol{x}) &= (\boldsymbol{u}^*(\boldsymbol{x}))^T A \boldsymbol{u}^*(\boldsymbol{x}) - 2(\boldsymbol{u}^*(\boldsymbol{x}))^T \boldsymbol{b}(\boldsymbol{x}) + \Phi(\boldsymbol{0}) \\ &= \sum_{j=1}^{N}\sum_{k=1}^{N} u_j^*(\boldsymbol{x}) u_k^*(\boldsymbol{x}) \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) - 2\sum_{j=1}^{N} u_j^*(\boldsymbol{x}) \Phi(\boldsymbol{x} - \boldsymbol{x}_j) + \Phi(\boldsymbol{x} - \boldsymbol{x}). \end{aligned}$$

If we define $u_0^*(\boldsymbol{x}) = -1$ and $\boldsymbol{x}_0 = \boldsymbol{x}$, then

$$P_{\Phi,\mathcal{X}}^2(\boldsymbol{x}) = \sum_{j=0}^{N}\sum_{k=0}^{N} u_j^*(\boldsymbol{x}) u_k^*(\boldsymbol{x}) \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k).$$

Finally, by using the Rayleigh quotient to get

$$\lambda_{\boldsymbol{x}} \leq \frac{\boldsymbol{c}^T A_{\boldsymbol{x}} \boldsymbol{c}}{\boldsymbol{c}^T \boldsymbol{c}}$$

for the (augmented) coefficient vector $\boldsymbol{c} = (\boldsymbol{u}_{\boldsymbol{x}}^*(\boldsymbol{x})) \in \mathbb{R}^{N+1}$ and $(N+1) \times (N+1)$ matrix $A_{\boldsymbol{x}}$, we have

$$P_{\Phi,\mathcal{X}}^2(\boldsymbol{x}) \geq \lambda_{\boldsymbol{x}} \sum_{j=0}^{N} \left[u_j^*(\boldsymbol{x})\right]^2,$$

and the statement follows by splitting off the $j = 0$ term again. $\square$

Theorem 8.1.2 not only implies the *uncertainty principle* (or *trade-off principle*) [549]

$$\lambda_{\boldsymbol{x}} \leq P^2_{\Phi,\mathcal{X}}(\boldsymbol{x}) \quad \text{or} \quad \lambda_{\min} \leq \min_{j=1,\ldots,N} P^2_{\Phi,\mathcal{X}\setminus\{\boldsymbol{x}_j\}}(\boldsymbol{x}),$$

which gives the power function as an upper bound for the smallest eigenvalue and vice versa. The theorem also provides an upper bound on the Lebesgue function, i.e.,

$$\sum_{j=1}^{N} \left|u_j^*(\boldsymbol{x})\right|^2 \leq \frac{P^2_{\Phi,\mathcal{X}}(\boldsymbol{x})}{\lambda_{\boldsymbol{x}}} - 1.$$

Since the power function can be bounded in terms of the fill distance $h_{\mathcal{X},\Omega}$, and the minimum eigenvalue in terms of the separation distance $q_{\mathcal{X}}$, we see that for quasi-uniformly distributed data, i.e., if $h_{\mathcal{X},\Omega} \simeq q_{\mathcal{X}}$, the Lebesgue function will not grow too rapidly.

There is also a trade-off principle for compactly supported functions. This was explained theoretically as well as illustrated with numerical experiments by Schaback [553]. The consequences are as follows. In the case of stationary interpolation, i.e., if we scale the support size of the basis functions proportional to the fill distance $h_{\mathcal{X},\Omega}$, then the "bandwidth" of the interpolation matrix A is constant. This means we can apply numerical algorithms (e.g., conjugate gradient) that can be performed in $\mathcal{O}(N)$ computational complexity. The method is numerically stable, but there will be essentially no convergence (see Table 8.1). In the non-stationary case, i.e., with fixed support size, the bandwidth of A increases as $h_{\mathcal{X},\Omega}$ decreases. This results in convergence (i.e., the error decreases) as we showed in Chapter 5, but the interpolation matrices will become more and more dense as well as ill-conditioned. Therefore, this approach is not very efficient (see Table 8.2).

In Tables 8.1 and 8.2 we illustrate this behavior. We use the compactly supported function $\varphi_{3,1}(r) = (1-r)_+^4 (4r+1)$ to interpolate Franke's function

$$\begin{aligned} F(x,y) \;=\; & \frac{3}{4}\left[\exp\left(-\frac{(9x-2)^2}{4} - \frac{(9y-2)^2}{4}\right) + \exp\left(-\frac{(9x+1)^2}{49} - \frac{(9y+1)^2}{10}\right)\right] \\ & + \frac{1}{2}\exp\left(-\frac{(9x-7)^2}{4} - (9y-3)^2\right) - \frac{1}{5}\exp\left(-(9x-4)^2 - (9y-7)^2\right) \end{aligned}$$

on a grid of equally spaced points in the unit square $[0,1]^2$. In the stationary case (Table 8.1) the support of the basis function is scaled to contain 25 grid points. Therefore, the "bandwidth" of the interpolation matrix A is kept constant (at 25), so that A is very sparse for finer grids. We can observe convergence for the first few iterations, but once an ℓ_2-error of approximately 2×10^{-3} is reached, there is no further improvement. This behavior is not yet fully understood. However, it is similar to what happens in the *approximate approximation* method of Maz'ya (see, e.g., [434]). The rate listed in the table is the exponent β of the observed ℓ_2-convergence rate $\mathcal{O}(h^\beta)$. The % nonzero column indicates the sparsity of the interpolation matrices, and the time is measured in seconds.

In the non-stationary case (Table 8.2) we used the basis function without adjusting its support size. This is the situation to which the error bounds of Chapter 5 apply.

Mesh	ℓ_2-error	rate	% nonzero	time
3×3	2.367490e-01		100	0
5×5	6.572754e-02	1.849	57.8	0
9×9	1.740723e-02	1.917	23.2	0
17×17	2.362950e-03	2.881	7.47	1
33×33	2.060493e-03	0.198	2.13	1
65×65	2.012010e-03	0.034	0.06	11
129×129	2.007631e-03	0.003	0.01	158

Table 8.1: 2D stationary interpolation with $\varphi(r) = (1-r)_+^4(4r+1)$, 25 points in support.

Mesh	ℓ_2-error	rate
3×3	2.407250e-01	
5×5	7.101748e-02	1.761
9×9	1.833534e-02	1.954
17×17	1.392914e-03	3.718
33×33	3.050789e-04	2.191
65×65	9.314516e-06	5.034

Table 8.2: 2D non-stationary interpolation with $\varphi(r) = (1-r)_+^4(4r+1)$, unit support.

We have convergence – although it is not obvious what the rate might be. However, the matrices become increasingly dense. Therefore, Table 8.2 is missing the entry for the 129×129 case, and even though no times are provided in that table, the time for the 65×65 case is already more than 20 minutes on a standard desktop PC.

8.2 Multilevel Interpolation and Approximation

In order to overcome the problems with both approaches for interpolation with compactly supported radial functions described above, Schaback suggested using a multilevel stationary scheme. This scheme was implemented first by Floater and Iske [217] and later studied by a number of other researchers (see, e.g., [115, 209, 281, 297, 318, 478, 630].

The basic idea of the multilevel interpolation algorithm is to scale the size of the support of the basis function with $h_{\mathcal{X},\Omega}$, but to interpolate to residuals on progressively refined sets of centers. This method has all of the combined benefits of the methods described earlier: it is computationally efficient (can be performed in $\mathcal{O}(N)$ operations), well-conditioned, and convergent.

An algorithm for multilevel interpolation is as follows:

Algorithm: (Multilevel interpolation)

1. Create nested point sets $\mathcal{X}_1 \subset \cdots \subset \mathcal{X}_K = \mathcal{X} \subset \mathbb{R}^s$, and initialize $\mathcal{P}f(\boldsymbol{x}) = 0$.
2. For $k = 1, 2, \ldots, K$ do

Mesh	ℓ_2-error	rate	% nonzero	time
3×3	2.367490e-01		100	0
5×5	6.665899e-02	1.828	57.8	0
9×9	2.087575e-02	1.675	23.2	0
17×17	1.090837e-04	4.258	7.47	0
33×33	1.497227e-04	2.865	2.13	6
65×65	5.313053e-05	1.495	0.06	37
129×129	1.112638e-05	2.256	0.01	212

Table 8.3: 2D (stationary) multilevel interpolation with $\varphi(r) = (1-r)_+^4(4r+1)$.

(a) Solve $u(\boldsymbol{x}) = f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})$ on $\mathcal{X}_k$.

(b) Update $\mathcal{P}f(\boldsymbol{x}) = \mathcal{P}f(\boldsymbol{x}) + u(\boldsymbol{x})$.

The representation of the update u at step k is of the form

$$u(\boldsymbol{x}) = \sum_{\boldsymbol{x}_j \in \mathcal{X}_k} c_j^{(k)} \varphi\left(\frac{\|\boldsymbol{x} - \boldsymbol{x}_j\|}{\rho_k}\right)$$

with $\rho_k \simeq h_{\mathcal{X}_k,\Omega}$. This requires the solution of a linear system whose size is determined by the number of points in $\mathcal{X}_k$.

In the numerical example listed in Table 8.3 we again use the compactly supported function $\varphi_{3,1}(r) = (1-r)_+^4(4r+1)$ and Franke's function.

The initial scale ρ_1 was chosen so that the basis function was supported on $[-2, 2]$. Subsequent scales were successively divided by 2 – just as the fill distance of the computational grids $\mathcal{X}_k$. The rate listed in the table is the exponent β of the observed ℓ_2-convergence rate $\mathcal{O}(h^\beta)$. The % nonzero column indicates the sparsity of the interpolation matrices, and the time is measured in seconds.

So far there are only limited theoretical results concerning the convergence of this multilevel algorithm. Narcowich, Schaback and Ward [478] show that a related algorithm (in which additional boundary conditions are imposed) converges at least linearly, and Hartmann analyzed the multilevel algorithm in his Ph.D. thesis [297]. He showed at least linear convergence for multilevel interpolation on a regular lattice for various radial basis functions. Similar results are obtained by Hales and Levesley [281] for polyharmonic splines, i.e., thin plate splines and powers. The main difficulty in proving the convergence of the multilevel algorithm is the fact that the approximation space changes from one level to the next. The approximation spaces are not nested (as they usually are for wavelets). This means that the native space norm changes from one level to the next. Hales and Levesley avoid this problem by scaling the (uniformly spaced) data instead of the basis functions. Then the fact that polyharmonic splines are in a certain sense harmonic (see Section 8.4) simplifies the analysis. This fact was also used by Wendland [634] to prove linear convergence for multilevel (scattered data) interpolation based on thin plate splines.

The same basic multilevel algorithm can also be used for other approximation methods. In [203] the idea was applied to moving least squares methods and approximate moving least squares methods. Tables 8.4 and 8.5 illustrate the effect of

Mesh	Shepard			linear precision		
	ℓ_2-error	rate	time	ℓ_2-error	rate	time
3×3	2.737339e-01		7	2.749670e-01		14
5×5	1.100713e-01	1.314	7	1.033060e-01	1.412	13
9×9	5.393041e-02	1.029	5	5.242492e-02	0.979	9
17×17	1.507797e-02	1.839	3	1.502361e-02	1.803	5
33×33	4.124059e-03	1.870	3	4.111092e-03	1.870	4
65×65	1.061904e-03	1.957	2	1.047348e-03	1.973	3
129×129	2.628645e-04	2.014	2	2.628645e-04	1.994	3

Table 8.4: 2D MLS approximation with weight $\varphi(r) = (1-r)_+^4(4r+1)$.

the multilevel algorithm for Shepard's method and a moving least squares approximation with linear precision, both based on the compactly supported weight function $\varphi_{3,1}(r) = (1-r)_+^4(4r+1)$. This experiment was conducted with a mollified Franke function f on the unit square $[0,1]^2$, i.e.,

$$\begin{aligned} F(x,y) &= \frac{3}{4}\left[\exp\left(-\frac{(9x-2)^2}{4} - \frac{(9y-2)^2}{4}\right) + \exp\left(-\frac{(9x+1)^2}{49} - \frac{(9y+1)^2}{10}\right)\right] \\ &\quad + \frac{1}{2}\exp\left(-\frac{(9x-7)^2}{4} - (9y-3)^2\right) - \frac{1}{5}\exp\left(-(9x-4)^2 - (9y-7)^2\right), \\ f(x,y) &= 15\exp\left(\frac{-1}{1-4(x-1/2)^2}\right)\exp\left(\frac{-1}{1-4(y-1/2)^2}\right)F(x,y) \ . \end{aligned}$$

The support scaling was as in the previous multilevel example.

One can observe that the basic Shepard's method actually performs much better than the predicted $\mathcal{O}(h)$ (see Table 8.4). Notice that the multilevel algorithm (illustrated in Table 8.5) improves the accuracy considerably at very little extra cost. It is interesting to note that this effect is much more pronounced for computations in $\mathbb{R}^2$ than in $\mathbb{R}$ (cf. [202]). The times listed in Tables 8.4 and 8.5 are due only to the evaluation on a very fine evaluation mesh since the method was coded so that no linear systems had to be solved. This means that the Lagrange multipliers for the case of linear precision were determined explicitly by solving the 3×3 Gram system analytically (cf. (7.5)). The resulting generating functions 7.6) were directly coded into the program.

There seems to be no theoretical investigation of the convergence properties of the multilevel algorithm for moving least squares approximation.

8.3 Preconditioning

In the first section of this chapter we noted that the system matrices arising in scattered data interpolation with radial basis functions tend to become very ill-conditioned as the minimal separation distance $q_{\mathcal{X}}$ between the data sites $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, is reduced. Therefore it is natural to devise strategies to prevent such instabilities by either preconditioning the system, or by finding a better basis for the approximation space we

Mesh	Shepard			linear precision		
	ℓ_2-error	rate	time	ℓ_2-error	rate	time
3×3	2.737339e-01		7	2.749670e-01		14
5×5	1.076424e-01	1.347	7	1.013114e-01	1.440	12
9×9	3.909725e-02	1.461	5	4.308322e-02	1.234	9
17×17	7.327282e-03	2.416	3	8.549613e-03	2.333	6
33×33	9.545860e-04	2.940	2	8.937409e-04	3.258	4
65×65	1.424136e-04	2.745	2	9.896052e-05	3.175	3
129×129	3.946680e-05	1.851	2	1.361339e-05	2.872	2

Table 8.5: 2D multilevel MLS approximation with $\varphi(r) = (1-r)_+^4(4r+1)$.

are using. The former approach is standard procedure in numerical linear algebra, and in fact we can use any of the well-established methods (such as preconditioned conjugate gradient iteration) to improve the stability and convergence of the interpolation systems that arise for strictly positive definite functions. In particular, the sparse systems that arise in (multilevel) interpolation with compactly supported radial basis functions can be efficiently solved with the preconditioned conjugate gradient method, and in fact the examples reported in the previous section were implemented using the conjugate gradient method with a diagonal (Jacobi) preconditioner.

The idea of using a more stable basis is well known from univariate polynomial and spline interpolation. The Lagrange basis functions for univariate polynomial interpolation are of course the ideal basis if we are interested in stably solving the interpolation equations since the resulting interpolation matrix is the identity matrix (which is certainly much better conditioned than, e.g., the Vandermonde matrix that we get if we use a monomial basis). Similarly, B-splines give rise to diagonally dominant, sparse system matrices which are much easier to deal with than the matrices we would get if we were to represent a spline interpolant using the alternative truncated power basis. Both of these examples are studied in great detail in standard numerical analysis texts (see, e.g., [350]) or in the literature on splines (see, e.g., [574]). We will address an analogous approach for radial basis functions in the next section.

Before we describe any of the specialized preconditioning procedures for radial basis function interpolation matrices we give two examples presented by Jackson [326] to illustrate the effects of and motivation for preconditioning in the context of radial basis functions.

8.3.1 Two Simple Examples

Example 1: (Uniform data) Let $s = 1$ and $\varphi(r) = r$ with no polynomial terms added. As data sites we choose $\mathcal{X} = \{1, 2, \ldots, 10\}$. This leads to the system matrix

$$A = \begin{bmatrix} 0 & 1 & 2 & 3 & \ldots & 9 \\ 1 & 0 & 1 & 2 & \ldots & 8 \\ 2 & 1 & 0 & 1 & \ldots & 7 \\ 3 & 2 & 1 & 0 & \ldots & 6 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 9 & 8 & 7 & 6 & \ldots & 0 \end{bmatrix}$$

with ℓ_2-condition number $\text{cond}(A) \approx 67$. Instead of solving the linear system $A\boldsymbol{c} = \boldsymbol{y}$, where $\boldsymbol{y} = [y_1, \ldots, y_{10}]^T \in \mathbb{R}^{10}$ is a vector of given real numbers (data values), we can find a suitable matrix B to premultiply both sides of the equation such that the system is simpler to solve. Ideally, the new system matrix BA should be the identity matrix, i.e., B should be an approximate inverse of A. Thus, having found an appropriate matrix B, we must now solve the linear system $BA\boldsymbol{c} = B\boldsymbol{y}$. For the matrix A above we can choose a preconditioner B as

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} & 0 & \ldots & 0 & 0 \\ 0 & \frac{1}{2} & -1 & \frac{1}{2} & \ldots & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & -1 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix}.$$

This leads to the following preconditioned system matrix BA in the system

$$\begin{bmatrix} 0 & 1 & 2 & \ldots & 8 & 9 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 9 & 8 & 7 & \ldots & 1 & 0 \end{bmatrix} \boldsymbol{c} = B\boldsymbol{y},$$

which is almost an identity matrix. Now $\text{cond}(BA) \approx 45$.

The motivation for this choice of B is the following. The function $\varphi(r) = r$ or $\Phi(x) = |x|$ is a fundamental solution of the Laplacian Δ, i.e.,

$$\Delta\Phi(x) = \frac{d^2}{dx^2}|x| = \frac{1}{2}\delta_0(x),$$

where δ_0 is the Dirac delta function. Thus, B is chosen as a discretization of the Laplacian with special choices at the endpoints of the data set.

Remark: One can also verify that the function $\Phi(x) = |x|$ minimizes the functional

$$\int_{\mathbb{R}} \left[f'(x)\right]^2 dx$$

over all functions interpolating data sampled from a function in the space

$$F = \{f \in C(\mathbb{R}), f' \in L_2(\mathbb{R})\}.$$

Therefore, the radial basis function $\varphi(r) = r$ is a linear (natural) spline. An analogous argument can be used to show that the function $\varphi(r) = r^3$ in $\mathbb{R}$ is nothing but a cubic natural spline. This is in agreement with the variational theory presented earlier according to which every radial basis function represents the minimum norm interpolant in its native space.

Example 2: (Nonuniform data) For nonuniformly distributed data we can use a different discretization of the Laplacian Δ for each row of B. To see this, let $s = 1$, and $\mathcal{X} = \{1, \frac{3}{2}, \frac{5}{2}, 4, \frac{9}{2}\}$, and again consider the radial function $\varphi(r) = r$. Then

$$A = \begin{bmatrix} 0 & \frac{1}{2} & \frac{3}{2} & 3 & \frac{7}{2} \\ \frac{1}{2} & 0 & 1 & \frac{5}{2} & 3 \\ \frac{3}{2} & 1 & 0 & \frac{3}{2} & 2 \\ 3 & \frac{5}{2} & \frac{3}{2} & 0 & \frac{1}{2} \\ \frac{7}{2} & 3 & 2 & \frac{1}{2} & 0 \end{bmatrix}$$

with $\text{cond}(A) \approx 18.15$, and if we choose

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -\frac{3}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{5}{6} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & -\frac{4}{3} & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

based on second-order backward differences of the points in $\mathcal{X}$, then the preconditioned system to be solved becomes

$$\begin{bmatrix} 0 & \frac{1}{2} & \frac{3}{2} & 3 & \frac{7}{2} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \frac{7}{2} & 3 & 2 & \frac{1}{2} & 0 \end{bmatrix} \boldsymbol{c} = B\boldsymbol{y}.$$

Once more, this system is almost trivial to solve and has an improved condition number of $\text{cond}(BA) \approx 8.94$.

8.3.2 Early Preconditioners

Ill-conditioning of the interpolation matrices was identified as a serious problem very early, and Nira Dyn along with some of her co-workers (see, e.g., [172], [173], [178], or [180]) provided some of the first preconditioning strategies tailored especially to radial basis functions.

For the following discussion we consider the general interpolation problem which includes polynomial reproduction. Therefore, we have to solve the following system of linear equations

$$\begin{bmatrix} A & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix}, \tag{8.1}$$

with the individual pieces given by $A_{jk} = \varphi(\|\boldsymbol{x}_j - \boldsymbol{x}_k\|)$, $j,k = 1,\ldots,N$, $P_{j\ell} = p_\ell(\boldsymbol{x}_j)$, $j = 1,\ldots,N$, $\ell = 1,\ldots,M$, $\boldsymbol{c} = [c_1,\ldots,c_N]^T$, $\boldsymbol{d} = [d_1,\ldots,d_M]^T$, $\boldsymbol{y} = [y_1,\ldots,y_N]^T$, and $\boldsymbol{0}$ a zero vector of length M with $M = \dim\Pi^s_{m-1}$. Here, as discussed earlier, φ should be strictly conditionally positive definite of order m and radial on $\mathbb{R}^s$ and the set $\mathcal{X} = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\}$ should be $(m-1)$-unisolvent.

The preconditioning scheme proposed by Dyn and her co-workers is a generalization of the simple differencing scheme discussed above. It is motivated by the fact that the polyharmonic splines

$$\varphi(r) = \begin{cases} r^{2k-s}\log r, & s \text{ even}, \\ r^{2k-s}, & s \text{ odd}, \end{cases}$$

$2k > s$, are fundamental solutions of the k-th iterated Laplacian in $\mathbb{R}^s$, i.e.,

$$\Delta^k\varphi(\|\boldsymbol{x}\|) = c\delta_{\boldsymbol{0}}(\boldsymbol{x}),$$

where $\delta_{\boldsymbol{0}}$ is the Dirac delta function, and c is an appropriate constant. For the (inverse) multiquadrics $\varphi(r) = (r^2+\alpha^2)^{\pm 1/2}$, which are also discussed in the papers mentioned above, application of the Laplacian yields a similar limiting behavior, i.e.,

$$\lim_{r\to\infty} \Delta^k\varphi(r) = 0,$$

and for $r \to 0$

$$\Delta^k\varphi(r) \gg 1.$$

One now wants to discretize the Laplacian on the (irregular) mesh given by the (scattered) data sites in $\mathcal{X}$. To this end Dyn, Levin, and Rippa [180] suggest the following procedure for the case of scattered data interpolation over $\mathbb{R}^2$.

1. Start with a triangulation of the set $\mathcal{X}$, e.g., the *Delaunay triangulation* will do. This triangulation can be visualized as follows.

   (a) Begin with the points in $\mathcal{X}$ and construct their *Dirichlet tesselation*. The Dirichlet tile of a particular point $\boldsymbol{x}$ is that subset of points in $\mathbb{R}^2$ which are closer to $\boldsymbol{x}$ than to any other point in $\mathcal{X}$. The left part of Figure 8.2 shows the Dirichlet tesselation for a given set of 6 points.

   (b) Construct the Delaunay triangulation, which is the dual of the Dirichlet tesselation, i.e., connect all strong neighbors in the Dirichlet tesselation, i.e., points whose tiles share a common edge. The right part of Figure 8.2 shows the corresponding Delaunay triangulation of the 6 points.

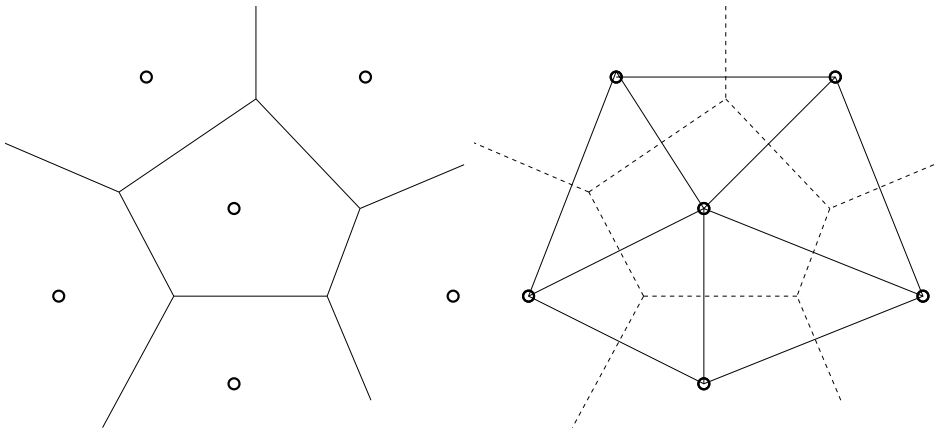

Figure 8.2: Dirichlet tesselation (left) and corresponding Delaunay triangulation (right) of the points ∘.

2. Discretize the Laplacian on this triangulation. In order to also take into account the boundary points Dyn, Levin and Rippa instead use a discretization of an iterated Green's formula which has the space Π^2_{m-1} as its null space. The necessary partial derivatives are then approximated on the triangulation using certain sets of vertices of the triangulation. (3 points for first order partials, 6 for second order).

The discretization described above yields the matrix $B = (b_{ji})_{j,i=1}^N$ as the preconditioning matrix in an analogous to the previous section. We now obtain

$$(BA)_{jk} = \sum_{i=1}^{N} b_{ji}\varphi(\|\boldsymbol{x}_i - \boldsymbol{x}_k\|) \approx \Delta^m \varphi(\|\cdot - \boldsymbol{x}_k\|)(\boldsymbol{x}_j), \quad j,k = 1,\ldots,N, \tag{8.2}$$

which has the property that the entries close to the diagonal are large compared to those away from the diagonal, which decay to zero as the distance between the two points involved goes to infinity. Since the part $BP = 0$ by step 2 of the construction, one must now solve the system

$$\begin{aligned} BA\boldsymbol{c} &= B\boldsymbol{y} \\ P^T\boldsymbol{c} &= 0. \end{aligned}$$

Actually, the system $BA\boldsymbol{c} = B\boldsymbol{y}$ is singular, but it is shown in the paper [180] that the additional constraints $P^T\boldsymbol{c} = 0$ guarantee existence of a unique solution. Furthermore, the coefficients $\boldsymbol{d}$ in the original expansion of the interpolant s can be obtained by solving

$$P\boldsymbol{d} = \boldsymbol{y} - A\boldsymbol{c},$$

i.e., by fitting the polynomial part of the expansion to the residual $\boldsymbol{y} - A\boldsymbol{c}$.

The approach just described leads to localized basis functions ψ which are linear combinations of the original basis functions φ. More precisely,

$$\psi_j(\boldsymbol{x}) = \sum_{i=1}^{N} b_{ji}\varphi(\|\boldsymbol{x} - \boldsymbol{x}_i\|) \approx \Delta^m \varphi(\|\cdot - \boldsymbol{x}_j\|)(\boldsymbol{x}), \tag{8.3}$$

φ	N	Grid I orig.	Grid I precond.	Grid II orig.	Grid II precond.
TPS	49	1181	4.3	1885	3.4
	121	6764	5.1	12633	3.9
MQ	49	7274	69.2	17059	222.8
	121	10556	126.0	107333	576.0

Table 8.6: Condition numbers without and with preconditioning.

where the coefficients b_{ji} are those determined in the discretization above.

Remarks:

1. The localized basis functions ψ_j, $j = 1, \ldots, N$, (see (8.3)) can be viewed as an alternative (better conditioned) basis for the approximation space spanned by the functions $\varphi_j = \varphi(\| \cdot - \boldsymbol{x}_j \|)$.

2. In [180] it is described how the preconditioned matrices can be used efficiently with various iterative schemes such as Chebyshev iteration or a version of the conjugate gradient method. The authors also mention smoothing of noisy data, or low-pass filtering as another application for this preconditioning scheme.

The effectiveness of the above preconditioning strategy was illustrated with some numerical examples in [180]. We list some of their results in Table 8.6. Thin plate splines and multiquadrics were tested on two different data sets (grid I and grid II) in $\mathbb{R}^2$. The shape parameter α for the multiquadrics was chosen to be the average mesh size. A linear term was added for thin plate splines, and a constant for multiquadrics.

One can see that the most dramatic improvement is achieved for the thin plate splines. This is to be expected since the method is tailored to them. As noted earlier, for the multiquadrics an application of the Laplacian does not yield the delta function, but for values of r close to zero gives just relatively large values.

Remarks:

1. Another early preconditioning strategy was suggested by Powell [516]. Powell uses Householder transformations to convert the matrix of the interpolation system (8.1) to a symmetric positive definite matrix, and then uses the conjugate gradient method. However, Powell reports that this method is not particularly effective for large thin plate spline interpolation problems in $\mathbb{R}^2$.

2. Baxter [27, 30] discusses the use of a preconditioned conjugate gradient method for solving the interpolation problem in the case when Gaussians or multiquadrics are used on a regular grid. The resulting matrices are Toeplitz matrices, and a large body of literature exists for dealing with this special case (see, e.g., [110]).

8.3.3 Preconditioned GMRES via Approximate Cardinal Functions

More recently, Beatson, Cherrie and Mouat [34] have proposed a preconditioner for the iterative solution of radial basis function interpolation systems using the GMRES method of Saad and Schultz [538]. The GMRES method is a general purpose iterative solver that can be applied to nonsymmetric (nondefinite) systems. For fast convergence the matrix should be preconditioned such that its eigenvalues are clustered around 1 and away from the origin. Obviously, if the basis functions for the radial basis function space were cardinal functions, then the matrix would be the identity matrix with all its eigenvalues equal to 1. Therefore, the GMRES method would converge in a single iteration. Consequently, the preconditioning strategy for the GMRES method is to obtain a preconditioning matrix B that is close to the inverse of A.

Since it is too expensive to find the true cardinal basis (this would involve at least as much work as solving the interpolation problem), the idea pursued in [34] (and suggested earlier in [36, 46]) is to find *approximate* cardinal functions similar to the functions ψ_j in the previous subsection. Now, however, there is also an emphasis on efficiency, i.e., we are interested in *local* approximate cardinal functions, if possible. Several different strategies were suggested in [34]. We will now explain the basic idea.

Given the centers $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, the j-th approximate cardinal function is given as a linear combination of the basis functions $\varphi_i = \varphi(\| \cdot - \boldsymbol{x}_i\|)$, where i runs over (some subset of) $\{1, \ldots, N\}$, i.e.,

$$\psi_j = \sum_{i=1}^{N} b_{ji}\varphi(\| \cdot - \boldsymbol{x}_i\|) + p_j, \tag{8.4}$$

where (for the conditionally positive definite case) p_j is a polynomial in Π^s_{m-1} and the coefficients b_{ji} satisfy the usual conditions

$$\sum_{i=1}^{N} b_{ji}p_j(\boldsymbol{x}_i) = 0 \qquad \text{for all } p_j \in \Pi^s_{m-1}. \tag{8.5}$$

The key feature in designing the approximate cardinal functions is to have only a few $n \ll N$ coefficients in (8.4) to be nonzero. In that case the functions ψ_j are found by solving small $n \times n$ linear systems, which is much more efficient than dealing with the original $N \times N$ system. For example, in [34] the authors use $n \approx 50$ for problems involving up to 10,000 centers. The resulting preconditioned system is of the same form as the earlier preconditioner (8.2), i.e., we now have to solve the preconditioned problem

$$(BA)\boldsymbol{c} = B\boldsymbol{y},$$

where the entries of the matrix BA are just $\psi_j(\boldsymbol{x}_k)$, $j, k = 1, \ldots, N$.

The simplest strategy for determining the coefficients b_{ji} is to select the n nearest neighbors of $\boldsymbol{x}_j$, and to find b_{ji} by solving the (local) cardinal interpolation problem

$$\psi_j(\boldsymbol{x}_i) = \delta_{ij}, \qquad i = 1, \ldots, n,$$

subject to the moment constraint (8.5) listed above. Here δ_{ij} is the Kronecker-delta, and the points $\boldsymbol{x}_i$ are the nearest neighbors selected above.

φ	N	unprecond.	local precond.	local precond. w/special
TPS	289	4.005e06	1.464e03	5.721e00
	1089	2.753e08	6.359e05	1.818e02
	4225	2.605e09	2.381e06	1.040e06
MQ	289	1.506e08	3.185e03	2.639e02
	1089	2.154e09	8.125e05	5.234e04
	4225	3.734e10	1.390e07	4.071e04

Table 8.7: Condition numbers without and with preconditioning.

This basic strategy is improved by adding so-called *special points* that are distributed (very sparsely) throughout the domain.

A few numerical results for thin plate spline and multiquadric interpolation in $\mathbb{R}^2$ from [34] are listed in Table 8.7. The condition numbers are ℓ_2-condition numbers, and the points were randomly distributed in the unit square. The "local precond." column uses the $n = 50$ nearest neighbors to determine the approximate cardinal functions, whereas the right-most column uses the 41 nearest neighbors plus 9 special points placed uniformly in the unit square. The effect of the preconditiong on the performance of the GMRES algorithm was, e.g., a reduction from 103 to 8 iterations for the 289 point data set for thin plate splines, or from 145 to 11 for multiquadrics.

Remark: An extension of the ideas of Beatson, Cherrie and Mouat [34] to linear systems arising in the collocation solution of partial differential equations (see Chapter 9) was explored in Mouat's Ph.D. thesis [468] and also in the recent paper by Ling and Kansa [395].

8.4 Change of Basis

Another idea that can be used to obtain a "better" basis for conditionally positive definite radial basis functions is closely connected to finding the reproducing kernel of the associated native space. Since we did not elaborate on the construction of the native spaces for conditionally positive definite functions earlier, we will now present the relevant formulas (without going into the details). In particular, for polyharmonic splines we will be able to find a basis that is in a certain sense *homogeneous*, and therefore the condition number of the related interpolation matrix will depend only on the number N of data points, but *not* on their separation distance.

This approach was suggested by Beatson, Light and Billings [42], and has its roots in work by Sibson and Stone [582].

Let Φ be a strictly conditionally positive definite kernel of order m, and $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega \subset \mathbb{R}^s$ be an $(m-1)$-unisolvent set of centers. Then the reproducing kernel for the native space $\mathcal{N}_\Phi(\Omega)$ is given by

$$K(\boldsymbol{x}, \boldsymbol{y}) \;=\; \Phi(\boldsymbol{x}, \boldsymbol{y}) - \sum_{k=1}^{M} p_k(\boldsymbol{x})\Phi(\boldsymbol{x}_k, \boldsymbol{y}) - \sum_{\ell=1}^{M} p_\ell(\boldsymbol{y})\Phi(\boldsymbol{x}, \boldsymbol{x}_\ell)$$

$$+\sum_{k=1}^{M}\sum_{\ell=1}^{M} p_k(\boldsymbol{x})p_\ell(\boldsymbol{y})\Phi(\boldsymbol{x}_k,\boldsymbol{x}_\ell) + \sum_{\ell=1}^{M} p_\ell(\boldsymbol{x})p_\ell(\boldsymbol{y}),$$

where the points $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_M$ are an $(m-1)$-unisolvent subset of $\mathcal{X}$ and the polynomials p_k, $k=1,\ldots,M$, form a *cardinal* basis for Π^s_{m-1} whose dimension is $M=\binom{s+m-1}{m-1}$, i.e.,

$$p_\ell(\boldsymbol{x}_k)=\delta_{k,\ell}, \qquad k,\ell=1,\ldots,M.$$

An immediate consequence is that we can express the radial basis function interpolant to values of some function f given on $\mathcal{X}$ in the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{N} c_j K(\boldsymbol{x},\boldsymbol{x}_j), \qquad \boldsymbol{x}\in\mathbb{R}^s\,.$$

The coefficients c_j are determined by satisfying the interpolation conditions

$$\mathcal{P}f(\boldsymbol{x}_i)=f(\boldsymbol{x}_i), \qquad i=1,\ldots,N.$$

We will see below (in Tables 8.8 and 8.9) that this basis already performs "better" than the standard basis $\{\Phi(\cdot,\boldsymbol{x}_1),\ldots,\Phi(\cdot,\boldsymbol{x}_N)\}$ if we keep the number of centers fixed, and vary only their separation distance.

To obtain the homogeneous basis referred to above we modify K by subtracting the tensor product polynomial, i.e.,

$$\kappa(\boldsymbol{x},\boldsymbol{y}) = K(\boldsymbol{x},\boldsymbol{y}) - \sum_{\ell=1}^{M} p_\ell(\boldsymbol{x})p_\ell(\boldsymbol{y}).$$

Now, if $\boldsymbol{y}$ is one of the points $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_M$ in the $(m-1)$-unisolvent subset of $\mathcal{X}$ mentioned above, then

$$\begin{aligned}\kappa(\cdot,\boldsymbol{y}) &= \Phi(\cdot,\boldsymbol{y}) - \sum_{k=1}^{M} p_k(\cdot)\Phi(\boldsymbol{x}_k,\boldsymbol{y}) - \sum_{\ell=1}^{M} p_\ell(\boldsymbol{y})\Phi(\cdot,\boldsymbol{x}_\ell) + \sum_{k=1}^{M}\sum_{\ell=1}^{M} p_k(\cdot)p_\ell(\boldsymbol{y})\Phi(\boldsymbol{x}_k,\boldsymbol{x}_\ell)\\ &= \Phi(\cdot,\boldsymbol{y}) - \sum_{k=1}^{M} p_k(\cdot)\Phi(\boldsymbol{x}_k,\boldsymbol{y}) - \Phi(\cdot,\boldsymbol{y}) + \sum_{k=1}^{M} p_k(\cdot)\Phi(\boldsymbol{x}_k,\boldsymbol{y}) = 0.\end{aligned}$$

This means that the functions $\kappa(\cdot,\boldsymbol{x}_j)$, $j=1,\ldots,N$, cannot be used as a basis of our approximation space. However, it turns out that the matrix C with entries $C_{i,j}=\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j)$, $i,j=M+1,\ldots,N$, is positive definite, and therefore we obtain the following basis

$$\{p_1,\ldots,p_M\}\cup\{\kappa(\cdot,\boldsymbol{x}_{M+1}),\ldots,\kappa(\cdot,\boldsymbol{x}_N)\},$$

and the interpolant can be represented in the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{M} d_j p_j(\boldsymbol{x}) + \sum_{k=M+1}^{N} c_k\kappa(\boldsymbol{x},\boldsymbol{x}_k), \qquad \boldsymbol{x}\in\mathbb{R}^s\,.$$

Spacing h	Standard matrix	Reproducing kernel	Homogeneous matrix
1/8	3.5158e03	1.8930e04	7.5838e03
1/16	3.8938e04	2.6514e05	1.1086e05
1/32	5.1363e05	4.0007e06	1.6864e06
1/64	7.6183e06	6.2029e07	2.6264e07

Table 8.8: Condition numbers for different thin plate spline bases on $[0,1]^2$ with increasing number of points and varying separation distance.

Since the polynomials p_k are cardinal on $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ the coefficients are determined by solving the linear system

$$\begin{bmatrix} I & 0 \\ P^T & C \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{c} \end{bmatrix} = \boldsymbol{y}, \tag{8.6}$$

with I an $M \times M$ identity matrix, C as above, $P_{ij} = p_j(\boldsymbol{x}_i)$, $j = 1, \ldots, M$, $i = M+1, \ldots, N$, $\boldsymbol{c} = [c_{M+1}, \ldots, c_N]^T$, $\boldsymbol{d} = [d_1, \ldots, d_M]^T$, and the right-hand side $\boldsymbol{y} = [f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_M), f(\boldsymbol{x}_{M+1}), \ldots, f(\boldsymbol{x}_N)]^T$. The identity block (cardinality of the polynomial basis functions) implies that the coefficient vector $\boldsymbol{d}$ is given by

$$d_j = f(\boldsymbol{x}_j), \qquad j = 1, \ldots, M,$$

and therefore the system (8.6) can be solved as

$$C\boldsymbol{c} = \tilde{\boldsymbol{y}} - P^T \boldsymbol{d},$$

where $\tilde{\boldsymbol{y}} = [f(\boldsymbol{x}_{M+1}), \ldots, f(\boldsymbol{x}_N)]^T$ and the matrix C is symmetric and positive definite. Finally, for polyharmonic splines, the ℓ_2-condition number of the matrix C is invariant under a uniform scaling of the centers, i.e., if $C^h = (\kappa(h\boldsymbol{x}_i, h\boldsymbol{x}_j))$, then $\mathrm{cond}(C^h) = \mathrm{cond}(C)$. This is proved to varying degrees in the paper [42] by Beatson, Light and Billings, the book [634] by Wendland, and the paper [320] by Iske.

We close with some numerical experiments from [42]. They use thin plate splines in $\mathbb{R}^2$. In the first experiment (illustrated in Table 8.8) the problem is formulated on the unit square $[0,1]^2$. Here both the number of points and the separation distance vary from one row in the table to the next. The three different columns list the ℓ_2-condition numbers of the interpolation matrix for the three different approaches mentioned above, i.e., using the standard basis consisting of functions $\Phi(\cdot, \boldsymbol{x}_j)$ and monomials, using the reproducing kernels $K(\cdot, \boldsymbol{x}_j)$, and using the matrix C. The three polynomial cardinal functions are based on the three corners $(0,0)$, $(0,1)$, and $(1,0)$. With this setup all three methods perform comparably.

In the second experiment (shown in Table 8.9) the number of points is kept fixed at 5×5 equally spaced points. However, the domain is scaled to the square $[0,a]^2$ with scale parameter a, so that only the separation distance $q_{\mathcal{X}}$ changes from one row to the next. Now, clearly the two new methods show less dependence on the separation distance, with the homogeneous matrix C being completely insensitive as claimed earlier.

Remark: Iske takes advantage of the scale invariance of polyharmonic splines (and thin plate splines in particular) in the construction of a numerical multiscale solver for transport problems (see, e.g., [47]).

Scale parameter	Standard matrix	Reproducing kernel	Homogeneous matrix
0.001	2.4349e08	8.4635e08	5.4938e02
0.01	2.4364e06	8.4640e06	5.4938e02
0.1	2.5179e04	8.5134e04	5.4938e02
1.0	3.6458e02	1.3660e03	5.4938e02
10	1.8742e06	1.2609e03	5.4938e02
100	1.1520e11	1.1396e05	5.4938e02
1000	3.4590e15	1.1386e07	5.4938e02

Table 8.9: Condition numbers for different thin plate spline bases on $[0, a]^2$ with fixed number of points and varying separation distance.

8.5 Special Numerical Algorithms

Since the use of radial basis functions for interpolation of scattered data leads to (large) linear systems that are frequently ill-conditioned it is important to devise algorithms that can

1. efficiently solve the interpolation system (preferably in $\mathcal{O}(N)$ operations), and
2. efficiently evaluate a radial basis function expansion once its coefficients have been determined (preferably in a constant number of operations – independent of N).

The second goal is also important for approximation via the moving least squares method or by quasi-interpolation.

All of the work described below is very recent, and it is quite likely that much more insight can be gained, and many improvements are still possible.

8.5.1 Iterative Algorithms

This section is based on the contents of the papers [562, 563] by Schaback and Wendland, the book [634] by Wendland, and the papers [212, 213] by Faul and Powell.

We concentrate on systems for strictly positive definite functions (variations for strictly conditionally positive definite functions also exist). One of the central ingredients is the native space inner product discussed in Chapter 5. As always we assume that our data sites are $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, but now we also consider a second set $\mathcal{Y} \subseteq \mathcal{X}$.

If we consider $\mathcal{P}_\mathcal{Y} f$ to be the interpolant to f on $\mathcal{Y} \subseteq \mathcal{X}$, then $\langle f - \mathcal{P}_\mathcal{Y} f, \mathcal{P}_\mathcal{Y} f \rangle_{\mathcal{N}_\Phi(\Omega)} = 0$ and we obtain the energy split (see Corollary 5.5.3)

$$\|f\|^2_{\mathcal{N}_\Phi(\Omega)} = \|f - \mathcal{P}_\mathcal{Y} f\|^2_{\mathcal{N}_\Phi(\Omega)} + \|\mathcal{P}_\mathcal{Y} f\|^2_{\mathcal{N}_\Phi(\Omega)}.$$

One possible point of view is now to consider an iteration on residuals. To this end we start with our desired interpolant $r_0 = \mathcal{P}_\mathcal{X} f$ on the entire set $\mathcal{X}$, and an appropriate sequence of sets $\mathcal{Y}_k$, $k = 0, 1, \ldots$ (we will discuss some possible choices later). Then, we iteratively define

$$r_{k+1} = r_k - \mathcal{P}_{\mathcal{Y}_k} r_k, \qquad k = 0, 1, \ldots. \tag{8.7}$$

Now, the energy splitting identity with $f = r_k$ gives us

$$\|r_k\|^2_{\mathcal{N}_\Phi(\Omega)} = \|r_k - \mathcal{P}_{\mathcal{Y}_k} r_k\|^2_{\mathcal{N}_\Phi(\Omega)} + \|\mathcal{P}_{\mathcal{Y}_k} r_k\|^2_{\mathcal{N}_\Phi(\Omega)} \tag{8.8}$$

or, using the iteration formula (8.7),

$$\|r_k\|^2_{\mathcal{N}_\Phi(\Omega)} = \|r_{k+1}\|^2_{\mathcal{N}_\Phi(\Omega)} + \|r_k - r_{k+1}\|^2_{\mathcal{N}_\Phi(\Omega)}. \tag{8.9}$$

Therefore,

$$\begin{aligned}
\sum_{k=0}^{K} \|\mathcal{P}_{\mathcal{Y}_k} r_k\|^2_{\mathcal{N}_\Phi(\Omega)} &= \sum_{k=0}^{K} \|r_k - r_{k+1}\|^2_{\mathcal{N}_\Phi(\Omega)} \\
&= \sum_{k=0}^{K} \left\{ \|r_k\|^2_{\mathcal{N}_\Phi(\Omega)} - \|r_{k+1}\|^2_{\mathcal{N}_\Phi(\Omega)} \right\} \\
&= \|r_0\|^2_{\mathcal{N}_\Phi(\Omega)} - \|r_{K+1}\|^2_{\mathcal{N}_\Phi(\Omega)} \leq \|r_0\|^2_{\mathcal{N}_\Phi(\Omega)},
\end{aligned}$$

which shows that the sequence of partial sums is monotone increasing and bounded, and therefore convergent – even for a poor choice of the sets $\mathcal{Y}_k$. If we can show that the residuals r_k converge to zero, then we would have that the iteratively computed approximation

$$s_{K+1} = \sum_{k=0}^{K} \mathcal{P}_{\mathcal{Y}_k} r_k = \sum_{k=0}^{K} (r_k - r_{k+1}) = r_0 - r_{K+1} \tag{8.10}$$

converges to the original interpolant $r_0 = \mathcal{P}_{\mathcal{X}} f$.

Remark: While this residual iteration algorithm has some structural similarities with the multilevel algorithm of Section 8.2 we now are considering a way to efficiently compute the interpolant $\mathcal{P}_{\mathcal{X}} f$ on some fine set $\mathcal{X}$ based on a single function Φ, whereas earlier, our final interpolant was the result of using the spaces $\bigcup_{k=1}^{K} \mathcal{N}_{\Phi_k}(\Omega)$, where Φ_k was an appropriately scaled version of the basis function Φ. And the goal there was to approximate f.

In order to prove convergence of the residual iteration let us assume that we can find sets of points $\mathcal{Y}_k$ such that at step k at least some fixed percentage of the energy of the residual is picked up by its interpolant, i.e.,

$$\|\mathcal{P}_{\mathcal{Y}_k} r_k\|^2_{\mathcal{N}_\Phi(\Omega)} \geq \gamma \|r_k\|^2_{\mathcal{N}_\Phi(\Omega)} \tag{8.11}$$

with some fixed $\gamma \in (0, 1]$. Then (8.9) and the iteration formula (8.7) imply

$$\|r_{k+1}\|^2_{\mathcal{N}_\Phi(\Omega)} = \|r_k\|^2_{\mathcal{N}_\Phi(\Omega)} - \|\mathcal{P}_{\mathcal{Y}_k} r_k\|^2_{\mathcal{N}_\Phi(\Omega)},$$

and therefore

$$\|r_{k+1}\|^2_{\mathcal{N}_\Phi(\Omega)} \leq \|r_k\|^2_{\mathcal{N}_\Phi(\Omega)} - \gamma \|r_k\|^2_{\mathcal{N}_\Phi(\Omega)} = (1 - \gamma) \|r_k\|^2_{\mathcal{N}_\Phi(\Omega)}.$$

Applying this bound recursively yields (see [562])

Theorem 8.5.1 *If the choice of sets $\mathcal{Y}_k$ satisfies (8.11), then the residual iteration (8.10) converges linearly in the native space norm, and after K steps of iterative refinement there is an error bound*

$$\|r_0 - s_K\|^2_{\mathcal{N}_\Phi(\Omega)} = \|r_K\|^2_{\mathcal{N}_\Phi(\Omega)} \leq (1-\gamma)^K \|r_0\|^2_{\mathcal{N}_\Phi(\Omega)}.$$

This theorem has various limitations. In particular, the norm involves the function Φ which makes it difficult to find sets $\mathcal{Y}_k$ that satisfy (8.11). Moreover, the native space norm of the initial residual r_0 is not known, either. Therefore, using an equivalent discrete norm on the set $\mathcal{X}$, Schaback and Wendland establish an estimate of the form

$$\|r_0 - s_K\|^2_{\mathcal{X}} \leq \frac{C}{c}\left(1 - \delta\frac{c^2}{C^2}\right)^{K/2} \|r_0\|^2_{\mathcal{X}},$$

where c and C are constants denoting the norm equivalence, i.e.,

$$c\|s\|_{\mathcal{X}} \leq \|s\|_{\mathcal{N}_\Phi(\Omega)} \leq C\|s\|_{\mathcal{X}}$$

for any $s \in \mathcal{N}_\Phi(\Omega)$, and where δ is a constant analogous to γ (but based on use of the discrete norm $\|\cdot\|_{\mathcal{X}}$ in (8.11)).

In [563] a basic version of this algorithm – where the sets $\mathcal{Y}_k$ consist of a single point – is described and tested. The resulting approximation yields the *best K-term approximation* to the interpolant. This idea is related to the concept of *greedy approximation algorithms* (see, e.g., [607]) and *sparse approximation* (see, e.g., [252]).

If the set $\mathcal{Y}_k$ consists only of a single point $\boldsymbol{y}_k$, then the partial interpolant $\mathcal{P}_{\mathcal{Y}_k} r_k$ is particularly simple, namely

$$\mathcal{P}_{\mathcal{Y}_k} r_k = \beta\Phi(\cdot, \boldsymbol{y}_k)$$

with

$$\beta = \frac{r_k(\boldsymbol{y}_k)}{\Phi(\boldsymbol{y}_k, \boldsymbol{y}_k)}.$$

The point $\boldsymbol{y}_k$ is picked to be the point in $\mathcal{X}$ where the residual is largest, i.e., $|r_k(\boldsymbol{y}_k)| = \|r_k\|_\infty$. For this choice of "set" $\mathcal{Y}_k$ we certainly satisfy the constraint (8.11). Moreover, the interpolation problem is (approximately) solved without having to invert any linear systems. The algorithm can be summarized as

Algorithm (Greedy one-point algorithm)

Input data locations $\mathcal{X}$, associated values of f, tolerance $\epsilon > 0$

Set initial residual $r_0 = \mathcal{P}_{\mathcal{X}} f$, initialize $s_0 = 0$, $e = \infty$, $k = 0$

Choose starting point $\boldsymbol{y}_k \in \mathcal{X}$

While $e > \epsilon$ do

Set $\beta = \dfrac{r_k(\boldsymbol{y}_k)}{\Phi(\boldsymbol{y}_k, \boldsymbol{y}_k)}$

For $1 \leq i \leq N$ do

$r_{k+1}(\boldsymbol{x}_i) = r_k(\boldsymbol{x}_i) - \beta\Phi(\boldsymbol{x}_i, \boldsymbol{y}_k)$
$s_{k+1}(\boldsymbol{x}_i) = s_k(\boldsymbol{x}_i) + \beta\Phi(\cdot, \boldsymbol{y}_k)$
Find $e = \max_{\mathcal{X}} |r_{k+1}|$ and the point $\boldsymbol{y}_{k+1}$ where it occurs

end

Increment $k = k + 1$

end

Remarks:

1. One advantage of this very simple (but fairly slow) algorithm is that no linear systems need to be solved. Nor are any matrix-vector multiplications required. This can be beneficial for problems that are very large (and possibly ill-conditioned), since in that situation the conjugate gradient method (which does use matrix-vector multiplications) may take very long.

2. For practical situations, e.g., for smooth radial basis functions and densely distributed points in $\mathcal{X}$ the convergence can be rather slow. In order to speed up the algorithm one should couple it with an algorithm that efficiently evaluates the residuals. If the basis functions are compactly supported, then a fast tree code algorithm can be used. Otherwise, fast multipole or fast Fourier transforms for non-uniform data can be used (see below for more details on these methods).

3. Schaback and Wendland [563] extend the simple greedy algorithm described above to a version that adaptively uses basis functions of varying scales.

Another iterative algorithm was suggested by Faul and Powell [212, 213]. From our earlier discussions we know that it is possible to express the radial basis function interpolant in terms of cardinal functions $u_j^*(\boldsymbol{x})$, $j = 1, \ldots, N$, i.e.,

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{N} f(\boldsymbol{x}_j) u_j^*(\boldsymbol{x}).$$

The basic idea of the Faul-Powell algorithm is to use *approximate* cardinal functions instead. Of course, this will only give an approximate value for the interpolant, and therefore an iteration on the residuals is suggested to improve the accuracy of this approximation. As done several times before, the approximate cardinal functions ψ_j, $j = 1, \ldots, N$, are determined as linear combinations of the basis functions $\Phi(\cdot, \boldsymbol{x}_i)$, i.e.,

$$\psi_j = \sum_{i \in \mathcal{L}_j} b_{ji} \Phi(\cdot, \boldsymbol{x}_i), \tag{8.12}$$

where $\mathcal{L}_j$ is an index set consisting of n ($n \approx 50$) indices that are used to determine the approximate cardinal function. For example, the n nearest neighbors of $\boldsymbol{x}_j$ with some additional special points (as in Section 8.3.3) will do. For every $j = 1, \ldots, N$, the coefficients b_{ji} found as solution of the linear system

$$\psi_j(\boldsymbol{x}_k) = \delta_{jk}, \qquad k \in \mathcal{L}_j. \tag{8.13}$$

These approximate cardinal functions are computed in a pre-processing step.

In its simplest form the residual iteration can be formulated as

$$\begin{aligned} s^{(0)}(\boldsymbol{x}) &= \sum_{j=1}^{N} f(\boldsymbol{x}_j)\psi_j(\boldsymbol{x}) \\ s^{(k+1)}(\boldsymbol{x}) &= s^{(k)}(\boldsymbol{x}) + \sum_{j=1}^{N} \left[f(\boldsymbol{x}_j) - s^{(k)}(\boldsymbol{x}_j) \right] \psi_j(\boldsymbol{x}), \qquad k = 0, 1, \ldots. \end{aligned}$$

Instead of adding the contribution of all approximate cardinal functions at the same time, this is done in a three-step process in the Faul-Powell algorithm. To this end index sets $\mathcal{L}_j$, $j = 1, \ldots, N-n$, are chosen so that $\mathcal{L}_j \subseteq \{j, j+1, \ldots, N\}$ making sure that $j \in \mathcal{L}_j$. Also, one needs to ensure that the corresponding centers form an $(m-1)$-unisolvent set.

Now, in the first step we define $s_0^{(k)} = s^{(k)}$, and then iterate

$$s_j^{(k)} = s_{j-1}^{(k)} + \theta_j^{(k)} \psi_j, \qquad j = 1, \ldots, N-n, \tag{8.14}$$

with

$$\theta_j^{(k)} = \frac{\langle \mathcal{P}f - s_{j-1}^{(k)}, \psi_j \rangle_{\mathcal{N}_\Phi(\Omega)}}{\langle \psi_j, \psi_j \rangle_{\mathcal{N}_\Phi(\Omega)}}. \tag{8.15}$$

The stepsize $\theta_j^{(k)}$ is chosen so that the native space best approximation to the residual $\mathcal{P}f - s_{j-1}^{(k)}$ from the space $\operatorname{span}\{\psi_j\}$ is added. Using the representation (8.12) of ψ_j in terms of the basis $\{\Phi(\cdot, \boldsymbol{x}_i) : i = 1, \ldots, N\}$, the reproducing kernel property of Φ, and the (local) cardinality property (8.13) of ψ_j we can calculate

$$\begin{aligned} \langle \psi_j, \psi_j \rangle_{\mathcal{N}_\Phi(\Omega)} &= \langle \psi_j, \sum_{i \in \mathcal{L}_j} b_{ji} \Phi(\cdot, \boldsymbol{x}_i) \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \sum_{i \in \mathcal{L}_j} b_{ji} \langle \psi_j, \Phi(\cdot, \boldsymbol{x}_i) \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \sum_{i \in \mathcal{L}_j} b_{ji} \psi_j(\boldsymbol{x}_i) = b_{jj}. \end{aligned}$$

Similarly, we get

$$\begin{aligned} \langle \mathcal{P}f - s_{j-1}^{(k)}, \psi_j \rangle_{\mathcal{N}_\Phi(\Omega)} &= \langle \mathcal{P}f - s_{j-1}^{(k)}, \sum_{i \in \mathcal{L}_j} b_{ji} \Phi(\cdot, \boldsymbol{x}_i) \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \sum_{i \in \mathcal{L}_j} b_{ji} \langle \mathcal{P}f - s_{j-1}^{(k)}, \Phi(\cdot, \boldsymbol{x}_i) \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \sum_{i \in \mathcal{L}_j} b_{ji} \left(\mathcal{P}f - s_{j-1}^{(k)} \right)(\boldsymbol{x}_i) \\ &= \sum_{i \in \mathcal{L}_j} b_{ji} \left(f(\boldsymbol{x}_i) - s_{j-1}^{(k)}(\boldsymbol{x}_i) \right). \end{aligned}$$

Therefore (8.14) and (8.15) can be written as

$$s_j^{(k)} = s_{j-1}^{(k)} + \frac{\psi_j}{b_{jj}} \sum_{i \in \mathcal{L}_j} b_{ji} \left(f(\boldsymbol{x}_i) - s_{j-1}^{(k)}(\boldsymbol{x}_i) \right), \qquad j = 1, \ldots, N-n.$$

In the second step the residual is interpolated on the remaining n points (collected via the index set $\mathcal{L}^*$). Thus, we find a function $\sigma^{(k)}$ such that

$$\sigma^{(k)}(\boldsymbol{x}_i) = f(\boldsymbol{x}_i) - s^{(k)}_{N-n}(\boldsymbol{x}_i), \qquad i \in \mathcal{L}^*,$$

and the approximation is updated, i.e.,

$$s^{(k+1)} = s^{(k)}_{N-n} + \sigma^{(k)}.$$

In the third step the residuals are updated, i.e.,

$$r^{(k+1)}_i = f(\boldsymbol{x}_i) - s^{(k+1)}(\boldsymbol{x}_i), \qquad i = 1, \ldots, N.$$

The outer iteration (on k) is now repeated unless the largest of these residuals is small enough.

We can summarize this algorithm as

Algorithm (Faul-Powell algorithm)

Input data locations $\mathcal{X}$, associated values of f, tolerance $\epsilon > 0$

Set $k = 0$ and $s^{(k)}_0 = 0$

Compute residuals $r^{(k)}_i = f(\boldsymbol{x}_i) - s^{(k)}(\boldsymbol{x}_i),\ i = 1, \ldots, N$, and set $e = \max\limits_{i=1,\ldots,N} |r^{(k)}_i|$.

While $e > \epsilon$ do

For $1 \le j \le N - n$ do

Update

$$s^{(k)}_j = s^{(k)}_{j-1} + \frac{\psi_j}{b_{jj}} \sum_{i \in \mathcal{L}_j} b_{ji} \left(f(\boldsymbol{x}_i) - s^{(k)}_{j-1}(\boldsymbol{x}_i) \right)$$

end

Solve the interpolation problem

$$\sigma^{(k)}(\boldsymbol{x}_i) = f(\boldsymbol{x}_i) - s^{(k)}_{N-n}(\boldsymbol{x}_i), \qquad i \in \mathcal{L}^*$$

Update the approximation

$$s^{(k+1)}_0 = s^{(k)}_{N-n} + \sigma^{(k)}$$

Compute new residuals and new value for e

Increment $k = k + 1$

end

Faul and Powell prove that this algorithm converges to the solution of the original interpolation problem. Similar to some of the other algorithms (greedy one-point or preconditioned GMRES) one needs to make sure that the residuals are evaluated efficiently by using a fast multipole expansion, fast Fourier transform, or compactly supported functions. Since the approximate cardinal functions can be computed in a preprocessing step this evaluation along with the determination of the sets $\mathcal{L}_j$ is the most expensive operation in the algorithm.

Remark: In its most basic form the Krylov subspace algorithm of Faul and Powell can also be explained via a dual approach to the greedy residual iteration algorithm of Schaback and Wendland. Instead of defining appropriate sets of points $\mathcal{Y}_k$, in the Faul and Powell algorithm one picks certain subspaces S_k of the native space. In particular, if S_k is the one-dimensional space $S_k = \text{span}\{\psi_k\}$ (where ψ_k is a local Lagrange function as in Section 8.3.3) we get the algorithm described above. For more details see [563].

8.5.2 Fast Fourier Transforms

In the recent papers [354, 490, 508] by Kunis, Nieslony, Potts and Steidl use of the fast Fourier transform at nonuniformly spaced points was suggested as an efficient way to solve and evaluate radial basis function problems. The software package NFFT by the authors is available for free download [353]. A discussion of the actual NFFT software would go beyond the scope of this manuscript. Instead, we briefly describe how to use NFFTs and FFTs to evaluate expansions of the form

$$\mathcal{P}f(\boldsymbol{y}_j) = \sum_{k=1}^{N} f(\boldsymbol{x}_k)\Phi(\boldsymbol{y}_j - \boldsymbol{x}_k) \tag{8.16}$$

simultaneously at many evaluation points $\boldsymbol{y}_j$, $j = 1, \ldots, M$. Direct summation requires $\mathcal{O}(MN)$ operations, while it can be shown that use of the NFFT reduces the cost to $\mathcal{O}(M + N)$ operations. Therefore, as is always the case with fast Fourier transforms, use of the algorithm will pay off for sufficiently many evaluations.

In their papers Nieslony, Potts and Steidl distinguish between kernels Φ that are singular and those that are non-singular. Singular kernels are C^∞ everywhere except at the origin and include examples such as

$$\frac{1}{r},\ \frac{1}{r^2},\ \log r,\ r^2 \log r,$$

where $r = \|\cdot\|$. Non-singular kernels are smooth everywhere such as Gaussians and (inverse) multiquadrics. We will restrict our discussion to this latter class.

The basic idea for the following algorithm is remarkably simple. It relies on the fact that the exponential $e^{-\alpha(\boldsymbol{y}_j - \boldsymbol{x}_k)}$ can be written as $e^{-\alpha \boldsymbol{y}_j} e^{\alpha \boldsymbol{x}_k}$. Moreover, the method applies to arbitrary kernels (which is in strong contrast to the fast multipole type methods discussed in the next section). One starts out by approximating the (arbitrary, but smooth) kernel using standard Fourier series, i.e.,

$$\Phi(\boldsymbol{x}) \approx \sum_{\boldsymbol{\ell} \in I_n} b_{\boldsymbol{\ell}} e^{2\pi i \boldsymbol{\ell} \boldsymbol{x}}$$

with index set $I_n = \left[-\frac{n}{2}, \frac{n}{2}\right)^s$. The coefficients $b_{\boldsymbol{\ell}}$ are found by the discrete inverse Fourier transform

$$b_{\boldsymbol{\ell}} = \frac{1}{n^s} \sum_{\boldsymbol{k} \in I_n} \Phi\left(\frac{\boldsymbol{k}}{n}\right) e^{-2\pi i \boldsymbol{k}\boldsymbol{\ell}/n}.$$

Numerically, this task is accomplished with software for the standard (inverse) FFT (e.g., [245]).

Remark: Note that this definition of the Fourier transform (as well as the one below) is different from the one used in Chapter 2. However, in order to stay closer to the software packages, we adopt the notation used there.

Therefore,

$$\begin{aligned} \mathcal{P}f(\boldsymbol{y}_j) &\approx \sum_{k=1}^{N} f(\boldsymbol{x}_k) \sum_{\boldsymbol{\ell} \in I_n} b_{\boldsymbol{\ell}} e^{2\pi i \boldsymbol{\ell}(\boldsymbol{y}_j - \boldsymbol{x}_k)} \\ &= \sum_{\boldsymbol{\ell} \in I_n} b_{\boldsymbol{\ell}} \sum_{k=1}^{N} f(\boldsymbol{x}_k) e^{2\pi i \boldsymbol{\ell}(\boldsymbol{y}_j - \boldsymbol{x}_k)} \end{aligned}$$

Now, the exponential is split using the above mentioned property, i.e.,

$$\mathcal{P}f(\boldsymbol{y}_j) \approx \sum_{\boldsymbol{\ell} \in I_n} b_{\boldsymbol{\ell}} \sum_{k=1}^{N} f(\boldsymbol{x}_k) e^{-2\pi i \boldsymbol{\ell}\boldsymbol{x}_k} e^{2\pi i \boldsymbol{\ell}\boldsymbol{y}_j}.$$

This, however, can be viewed as a fast Fourier transform at non-uniformly spaced points, i.e.,

$$\mathcal{P}f(\boldsymbol{y}_j) \approx \sum_{\boldsymbol{\ell} \in I_n} c_{\boldsymbol{\ell}} e^{2\pi i \boldsymbol{\ell}\boldsymbol{y}_j}.$$

where the coefficients $c_{\boldsymbol{\ell}} = b_{\boldsymbol{\ell}} a_{\boldsymbol{\ell}}$ with

$$a_{\boldsymbol{\ell}} = \sum_{k=1}^{N} f(\boldsymbol{x}_k) e^{-2\pi i \boldsymbol{\ell}\boldsymbol{x}_k}$$

which is nothing but an inverse discrete Fourier transform at non-uniformly spaced points. These latter two transforms are dealt with numerically using the NFFT software.

Together, for the case of non-singular kernels Φ we have the following algorithm.

Algorithm (Fast Fourier transform evaluation)

For $\boldsymbol{\ell} \in I_n$

Compute the coefficients

$$b_{\boldsymbol{\ell}} = \frac{1}{n^s} \sum_{\boldsymbol{k} \in I_n} \Phi\left(\frac{\boldsymbol{k}}{n}\right) e^{-2\pi i \boldsymbol{k}\boldsymbol{\ell}/n}$$

by inverse FFT.

Compute the coefficients

$$a_{\boldsymbol{\ell}} = \sum_{k=1}^{N} f(\boldsymbol{x}_k) e^{-2\pi i \boldsymbol{\ell} \boldsymbol{x}_k}$$

by inverse NFFT.

Compute the coefficients $c_{\boldsymbol{\ell}} = a_{\boldsymbol{\ell}} b_{\boldsymbol{\ell}}$.

end

For $1 \leq j \leq M$

Compute the values

$$\mathcal{P} f(\boldsymbol{y}_j) \approx \sum_{\boldsymbol{\ell} \in I_n} d_{\boldsymbol{\ell}} e^{2\pi i \boldsymbol{\ell} \boldsymbol{y}_j}$$

by NFFT.

end

Remarks:

1. In the papers [354, 490, 508] the authors also suggest a special boundary regularization in case the kernel does not decay fast enough, i.e., the kernel is large near the boundary of the domain.

2. Of course, this method will only provide an approximation of the expansion (8.16) and error estimates are provided in the literature (see, e.g., [490]).

3. While we only illustrated the use of (N)FFTs for the evaluation of radial sums it should be clear that this method can also be coupled with the algorithms of the previous sections (such as preconditioned GMRES, the "greedy" algorithm, or the Faul-Powell algorithm) to efficiently solve radial basis function interpolation systems.

A few examples of the use of fast Fourier transforms for the evaluation of approximate moving least squares approximations (quasi-interpolants) are given in Figures 8.3–8.5. The graphs on the left indicate ℓ_∞ approximation errors for a Franke-type function. The graphs on the right show the execution times in seconds for direct summation (solid lines) and FFT summations (dashed lines). The colors correspond to the three different types of kernels listed in Table 8.10 below. The red curves correspond to the Gaussians (listed in the $\mathcal{O}(h^2)$ column), green curves to the function in the $\mathcal{O}(h^4)$ column (Gaussian multiplied by a linear Laguerre polynomial), and blue curves to those in the $\mathcal{O}(h^6)$ column (Gaussian multiplied by a quadratic Laguerre polynomial).

The evaluation of the results listed in Figures 8.3–8.5 occurs at 10,001, 16,641, and 2,146,689 randomly distributed points in the unit square, respectively. The 3D experiments show that there is a cross-over value of about 1,000 evaluations at which the FFT approach becomes faster than the direct approach. For the one and two-dimensional experiments this cross-over point occurs much earlier and is not detectable in the figures.

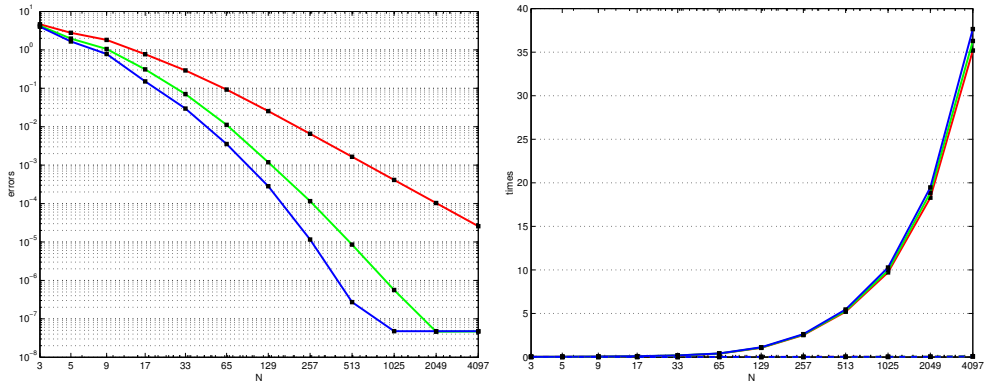

Figure 8.3: Convergence and execution times for 1D example.

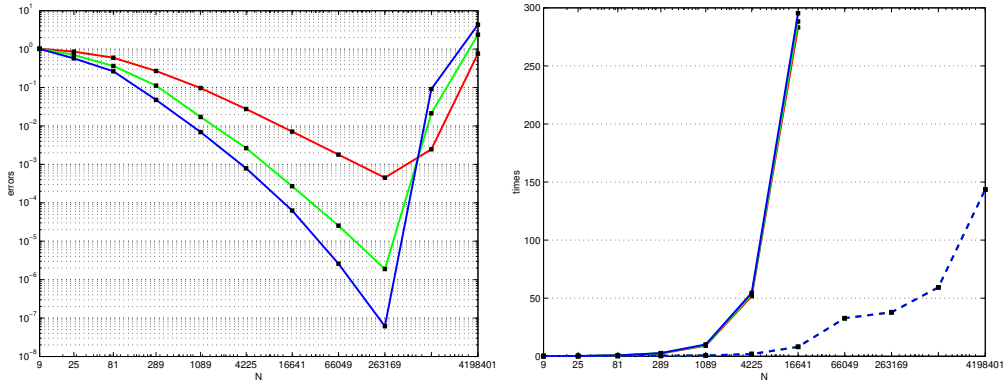

Figure 8.4: Convergence and execution times for 2D example.

The polynomial terms in Table 8.10 are given by generalized Laguerre polynomials with radial arguments. In general one can show (see, e.g., [434]) that if $L_d^{s/2}$ is used to denote the generalized Laguerre polynomial of degree d, then the smooth function f in $\mathbb{R}^s$ can be approximated with approximate approximation order $\mathcal{O}(h^{2d+2})$ by an expansion of the form

$$\mathcal{P}f(\boldsymbol{x}) = \frac{1}{(\pi\mathcal{D})^{s/2}} \sum_{k=1}^{N} f(\boldsymbol{x}_k) L_d^{s/2}\left(\frac{\|\boldsymbol{x}-\boldsymbol{x}_k\|^2}{\mathcal{D}h^2}\right) \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}_k\|^2}{\mathcal{D}h^2}\right).$$

Here $\mathcal{D}$ is a parameter that controls a so-called *saturation error*, i.e., the predicted approximation order is achieved only up to some user-controllable threshold (and therefore referred to as *approximate approximation*). This threshold is clearly visible in the convergence graphs.

8.5.3 The Fast Multipole Method

Another quite popular strategy for dealing with fast summation problems is known as the *fast multipole method*. This method was first suggested by Greengard and Rokhlin in 1987 (see, e.g., the original paper [269], the popular discussion [268], or the short

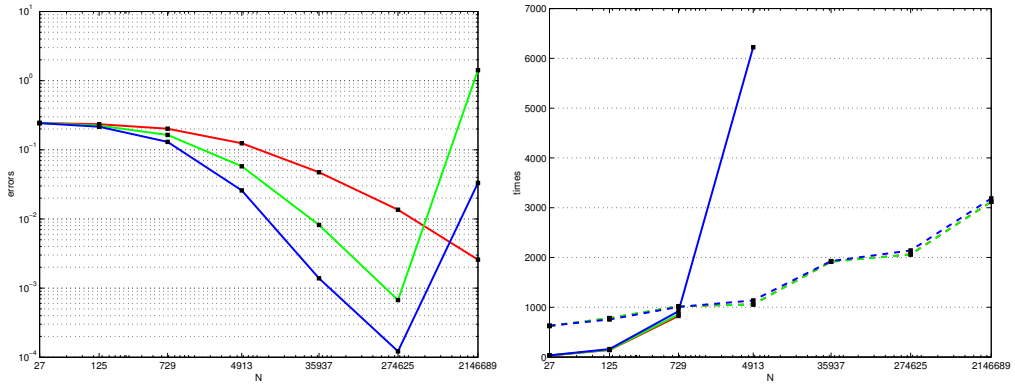

Figure 8.5: Convergence and execution times for 3D example.

s	$\mathcal{O}(h^2)$	$\mathcal{O}(h^4)$	$\mathcal{O}(h^6)$
1	$e^{-\lvert x\rvert^2}$	$\left(\frac{3}{2}-\lvert x\rvert^2\right)e^{-\lvert x\rvert^2}$	$\left(\frac{15}{8}-\frac{5}{2}\lvert x\rvert^2+\frac{1}{2}\lvert x\rvert^4\right)e^{-\lvert x\rvert^2}$
2	$e^{-\lVert\boldsymbol{x}\rVert^2}$	$\left(2-\lVert\boldsymbol{x}\rVert^2\right)e^{-\lVert\boldsymbol{x}\rVert^2}$	$\left(3-3\lVert\boldsymbol{x}\rVert^2+\frac{1}{2}\lVert\boldsymbol{x}\rVert^4\right)e^{-\lVert\boldsymbol{x}\rVert^2}$
3	$e^{-\lVert\boldsymbol{x}\rVert^2}$	$\left(\frac{5}{2}-\lVert\boldsymbol{x}\rVert^2\right)e^{-\lVert\boldsymbol{x}\rVert^2}$	$\left(\frac{35}{8}-\frac{7}{2}\lVert\boldsymbol{x}\rVert^2+\frac{1}{2}\lVert\boldsymbol{x}\rVert^4\right)e^{-\lVert\boldsymbol{x}\rVert^2}$

Table 8.10: Generating functions for approximate MLS approximation in $\mathbb{R}^s$.

course tailored to radial basis functions [37]). It has quickly become very popular in the computational sciences. The breakthrough accomplishment of this algorithm was the ability to perform fast evaluations of sums of the type

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{k=1}^{N} c_k \Phi(\boldsymbol{x}, \boldsymbol{x}_k), \qquad \boldsymbol{x} \in \mathbb{R}^s\,.$$

In particular, M such evaluations can be performed in $\mathcal{O}(M \log N)$ (or even $\mathcal{O}(M)$) operations instead of the standard $\mathcal{O}(MN)$ operations. The nonuniform fast Fourier transform of the previous section was able to do this also, and in a fairly general way for a very large class of kernels Φ, but the fast multipole method is a little older and may be more efficient since special expansions are used that are chosen with the particular kernel Φ in mind. We will now outline the basic idea of the *fast Gauss transform* [270]. This transform can be applied directly to the approximate moving least squares approximands based on Gaussians used in the previous section (see the numerical experiments reported in Table 8.11 below). The higher-order kernels consisting of Gaussians times Laguerre polynomials, however, require a completely new derivation.

Thus, using the abbreviation $\rho = \sqrt{\mathcal{D}}h$, we are now interested in a fast summation technique for M evaluations of the Gaussian quasi-interpolant (or *discrete Gauss transform*)

$$\mathcal{G}f(\boldsymbol{y}_j) = \sum_{k=1}^{N} f(\boldsymbol{x}_k) e^{-\lVert(\boldsymbol{y}_j - \boldsymbol{x}_k)/\rho\rVert^2}, \qquad j = 1, \ldots, M. \tag{8.17}$$

In [270] such an algorithm was developed, and in [591] a modification was suggested to cover also the case of variable scales ρ_k as needed for use with quasi-interpolation at scattered sites or with variable scales.

One of the central ingredients for the fast Gauss transform are the *multivariate Hermite functions*

$$h_{\boldsymbol{\alpha}}(\boldsymbol{x}) = (-1)^{|\boldsymbol{\alpha}|} D^{\boldsymbol{\alpha}} e^{-\|\boldsymbol{x}\|^2}, \tag{8.18}$$

which are related to the (multivariate) Hermite polynomials via

$$H_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \prod_{i=1}^{s} H_{\alpha_i}(x_i) = e^{\|\boldsymbol{x}\|^2} h_{\boldsymbol{\alpha}}(\boldsymbol{x}) \tag{8.19}$$

(see, e.g., the univariate formula (6.1.3) in [5]). It is beneficial that the Hermite functions can be evaluated recursively using the (univariate) recurrence relation

$$\begin{aligned} h_{n+1}(x) &= 2xh_n(x) - 2nh_{n-1}(x), \qquad n = 1, 2, \ldots, \\ h_0(x) &= e^{-|x|^2},\ h_1(x) = 2xe^{-|x|^2}, \end{aligned}$$

which follows immediately from (8.19) and the recursion relation for Hermite polynomials (see, e.g., formula (6.1.10) in [5]).

The algorithm of Greengard and Strain is based on three basic expansions which we list below as Theorems 8.5.2–8.5.4 (see [270, 271]). The main effect of these expansions is the fact that the variables $\boldsymbol{y}_j$ and $\boldsymbol{x}_k$ will be separated (this is the fundamental "trick" used with all fast summation algorithms). This will allow for the pre-computation and storage of certain *moments* below.

The first step in the algorithm is to scale the problem to the unit box $[0,1]^s$, and then subdivide the unit box into smaller boxes B and C which usually coincide. They can, however, also differ. The boxes B contain the *sources* $\boldsymbol{x}_k$, and the boxes C the *targets* $\boldsymbol{y}_j$. For each source box B one then determines its *interaction region* $IR(B)$. The interaction region of B is a set of nearest neighbors of B such that the error of truncating the sum over all boxes is below a certain threshold. Due to the fast decay of the Gaussians it is suggested (see [271]) to use the 9^s nearest neighbors for single precision and the 13^s nearest neighbors for double precision.

Theorem 8.5.2 *Let I_B be the index set denoting the sources $\boldsymbol{x}_k$ which lie in a box B with center $\boldsymbol{x}_B$ and side length ρ, and let $\boldsymbol{y}_C$ be the center of the target box C ($\in IR(B)$) of radius r_c containing the targets $\boldsymbol{y}_j$. Then the Gaussian field due to the sources in B,*

$$\mathcal{G}^{(B)} f(\boldsymbol{y}_j) = \sum_{k \in I_B} f(\boldsymbol{x}_k) e^{-\|(\boldsymbol{y}_j - \boldsymbol{x}_k)/\rho\|^2},$$

has the following Taylor expansion about $\boldsymbol{y}_C$:

$$\mathcal{G}^{(B)} f(\boldsymbol{y}_j) = \sum_{\boldsymbol{\alpha} \geq 0} A_{\boldsymbol{\alpha}}^{(B)} \left(\frac{\boldsymbol{y}_j - \boldsymbol{y}_C}{\rho} \right)^{\boldsymbol{\alpha}}, \tag{8.20}$$

where the coefficients $A_{\boldsymbol{\alpha}}^{(B)}$ are given by

$$A_{\boldsymbol{\alpha}}^{(B)} = \frac{(-1)^{|\boldsymbol{\alpha}|}}{\boldsymbol{\alpha}!} \sum_{k \in I_B} f(\boldsymbol{x}_k) h_{\boldsymbol{\alpha}} \left(\frac{\boldsymbol{x}_k - \boldsymbol{y}_C}{\rho} \right).$$

The error $E_T(p)$ due to truncating the series (8.20) after the p-th order terms satisfies the bound

$$|E_T(p)| = |\sum_{\boldsymbol{\alpha}>p} A_{\boldsymbol{\alpha}}^{(B)} \left(\frac{\boldsymbol{y}_j - \boldsymbol{y}_C}{\rho}\right)^{\boldsymbol{\alpha}}| \leq (1.09)^s F^{(B)} \frac{1}{\sqrt{(p+1)!^s}} \left[\frac{\left(\frac{r_c}{\rho}\right)^{p+1}}{1-\frac{r_c}{\rho}}\right]^s,$$

where $F^{(B)} = \sum_{k\in I_B} |f(\boldsymbol{x}_k)|$.

Remark: Here we used the multi-index notation $\boldsymbol{\alpha} \geq 0$ to denote the constraints $\alpha_i \geq 0$ for all $i = 1, \ldots, s$. More generally, for some integer p $\boldsymbol{\alpha} \geq p$ if $\alpha_i \geq p$ for all $i = 1, \ldots, s$. This implies $\boldsymbol{\alpha} > p$ for some integer p if $\boldsymbol{\alpha} \geq p$ and $\alpha_i > p$ for some i. We also use $\boldsymbol{\alpha} \geq \boldsymbol{\beta}$ if $\alpha_i \geq \beta_i$ for all $i = 1, \ldots, s$.

The expansion (8.20) will be used in the case when the source box B contains relatively few sources, but the target box C contains many targets.

By reversing the role of the Hermite functions and the shifted monomials one can write a single Gaussian as a multivariate Hermite expansion about a point $\boldsymbol{z}_0 \in \mathbb{R}^s$, i.e.,

$$e^{-\|(\boldsymbol{y}_j - \boldsymbol{x}_k)/\rho\|^2} = \sum_{\boldsymbol{\alpha}\geq 0} \frac{1}{\boldsymbol{\alpha}!} \left(\frac{\boldsymbol{x}_k - \boldsymbol{z}_0}{\rho}\right)^{\boldsymbol{\alpha}} h_{\boldsymbol{\alpha}} \left(\frac{\boldsymbol{y}_j - \boldsymbol{z}_0}{\rho}\right). \tag{8.21}$$

This is used in

Theorem 8.5.3 *(Far-field expansion) Let I_B be the index set denoting the sources $\boldsymbol{x}_k$ which lie in a box B with center $\boldsymbol{x}_B$ and side length ρ. Then the Gaussian field due to the sources in B,*

$$\mathcal{G}^{(B)} f(\boldsymbol{y}_j) = \sum_{k\in I_B} f(\boldsymbol{x}_k) e^{-\|(\boldsymbol{y}_j - \boldsymbol{x}_k)/\rho\|^2},$$

is equal to an Hermite expansion about $\boldsymbol{x}_B$:

$$\mathcal{G}^{(B)} f(\boldsymbol{y}_j) = \sum_{\boldsymbol{\alpha}\geq 0} B_{\boldsymbol{\alpha}}^{(B)} h_{\boldsymbol{\alpha}} \left(\frac{\boldsymbol{y}_j - \boldsymbol{x}_B}{\rho}\right). \tag{8.22}$$

The moments $B_{\boldsymbol{\alpha}}^{(B)}$ are given by

$$B_{\boldsymbol{\alpha}}^{(B)} = \frac{1}{\boldsymbol{\alpha}!} \sum_{k\in I_B} f(\boldsymbol{x}_k) \left(\frac{\boldsymbol{x}_k - \boldsymbol{x}_B}{\rho}\right)^{\boldsymbol{\alpha}}.$$

The error $E_H(p)$ due to truncating the series (8.22) after p-th order terms satisfies the bound

$$|E_H(p)| = |\sum_{\boldsymbol{\alpha}>p} B_{\boldsymbol{\alpha}}^{(B)} h_{\boldsymbol{\alpha}} \left(\frac{\boldsymbol{y}_j - \boldsymbol{x}_B}{\rho}\right)| \leq (1.09)^s F^{(B)} \frac{1}{\sqrt{(p+1)!^s}} \left[\frac{\left(\frac{r_c}{\rho}\right)^{p+1}}{1-\frac{r_c}{\rho}}\right]^s.$$

Theorem 8.5.3 is used when B contains many sources, but C only few targets. Finally, in the case when both B and C contain relatively many points we use

Theorem 8.5.4 *(Translation operation) Let the sources $\boldsymbol{x}_k$ lie in a box B with center $\boldsymbol{x}_B$ and side length ρ and let $\boldsymbol{y}_j$ be an evaluation point in a box C with center $\boldsymbol{y}_C$. Then the corresponding truncated Hermite expansion (8.22) can be expanded as a Taylor series of the form*

$$\mathcal{G}^{(BC)} f(\boldsymbol{y}_j) = \sum_{\boldsymbol{\beta} \geq 0} C_{\boldsymbol{\beta}} \left(\frac{\boldsymbol{y}_j - \boldsymbol{y}_C}{\rho} \right)^{\boldsymbol{\beta}}. \tag{8.23}$$

The coefficients $C_{\boldsymbol{\beta}}$ are given by

$$C_{\boldsymbol{\beta}} = \frac{(-1)^{|\boldsymbol{\beta}|}}{\boldsymbol{\beta}!} \sum_{\boldsymbol{\alpha} \leq p} B_{\boldsymbol{\alpha}}^{(B)} h_{\boldsymbol{\alpha}+\boldsymbol{\beta}} \left(\frac{\boldsymbol{x}_B - \boldsymbol{y}_C}{\rho} \right),$$

with $B_{\boldsymbol{\alpha}}^{(B)}$ as in Theorem 8.5.3. The error $E_T(p)$ due to truncating the series (8.23) after p-th order terms satisfies the bound

$$|E_T(p)| = |\sum_{\boldsymbol{\beta} > p} B_{\boldsymbol{\beta}}^{(B)} \left(\frac{\boldsymbol{x} - \boldsymbol{y}_C}{\rho} \right)^{\boldsymbol{\beta}}| \leq (1.09)^s F^{(B)} \frac{1}{\sqrt{(p+1)!^s}} \left[\frac{\left(\frac{r_c}{\rho}\right)^{p+1}}{1 - \frac{r_c}{\rho}} \right]^s.$$

Theorem 8.5.4 is based on the multivariate Taylor series expansion of the Hermite functions $h_{\boldsymbol{\alpha}}$

$$h_{\boldsymbol{\alpha}} \left(\frac{\boldsymbol{y}_j - \boldsymbol{x}_B}{\rho} \right) = \sum_{\boldsymbol{\beta} \geq 0} \frac{(-1)^{|\boldsymbol{\beta}|}}{\boldsymbol{\beta}!} \left(\frac{\boldsymbol{y}_j - \boldsymbol{y}_C}{\rho} \right)^{\boldsymbol{\beta}} h_{\boldsymbol{\alpha}+\boldsymbol{\beta}} \left(\frac{\boldsymbol{x}_B - \boldsymbol{y}_C}{\rho} \right).$$

Remarks:

1. The error estimates in the original paper on the fast Gauss transform [270] were incorrect. In the mean time a number of other authors have provided alternate error bounds in their papers (see, e.g., [31, 225, 271, 637]).

2. For 1D calculations on the order of $p = 20$ terms are required to achieve double precision accuracy. For the 2D one can get by with a smaller value of p (about 15), but the number of terms is of course much higher (on the order of p^s for s-dimensional problems).

The basic outline of the algorithm is as follows:

Algorithm:

1. If necessary, scale the problem so that the coarsest box $B_0 = [0, 1]^s$. Subdivide B_0 into smaller boxes with side length ρ parallel to the axes. Assign each source $\boldsymbol{x}_k$ to the box B in which it lies and each evaluation point $\boldsymbol{y}_j$ to the box C in which it lies.

2. Choose p so that the truncation error satisfies the desired accuracy, and for each box B compute and store the coefficients (or moments)

$$B_{\boldsymbol{\alpha}}^{(B)} = \frac{1}{\boldsymbol{\alpha}!} \sum_{k \in I_B} f(\boldsymbol{x}_k) \left(\frac{\boldsymbol{x}_k - \boldsymbol{x}_B}{\rho} \right)^{\boldsymbol{\alpha}}, \qquad \boldsymbol{\alpha} \leq p,$$

of its Hermite expansion (8.22).

3. For each evaluation box C, determine its interaction region $IR(C)$.

4. For each evaluation box C transform all Hermite expansions in source boxes within the interaction region $IR(C)$ into a single Taylor expansion using (8.23), i.e.,

$$\mathcal{G}f(\boldsymbol{y}_j) \approx \sum_{\boldsymbol{\beta} \le p} C_{\boldsymbol{\beta}} \left(\frac{\boldsymbol{y}_j - \boldsymbol{y}_C}{\rho} \right)^{\boldsymbol{\beta}},$$

where

$$C_{\boldsymbol{\beta}} = \frac{(-1)^{|\boldsymbol{\beta}|}}{\boldsymbol{\beta}!} \sum_{B \in IR(C)} \sum_{\boldsymbol{\alpha} \le p} B_{\boldsymbol{\alpha}}^{(B)} h_{\boldsymbol{\alpha}+\boldsymbol{\beta}} \left(\frac{\boldsymbol{x}_B - \boldsymbol{y}_C}{\rho} \right).$$

For a small number of points direct summation is more efficient than the fast transform. For the case $s = 1$ the fast Gauss transform should be preferable to direct summation for $N \approx 1000$ (with $M \approx 2000$ evaluation points). The break-even point in $\mathbb{R}^2$ is at about $N = 12000$ (with $M = 24000$). In particular, in $\mathbb{R}^3$, it is rather disappointing that the break-even point may not occur until about $N = 270000$ data sites (with $M = 2.16 \times 10^6$ evaluation points). This makes fast Gauss transform in its basic form virtually unusable for 3D applications.

Note that this algorithm does not use a hierarchical decomposition of space as is typical for so-called *tree codes*, as well as many other more general fast multipole algorithms. In this algorithm the interaction region is determined simply based on the fast decay of the Gaussian.

Clearly, the most work is involved in step 4. The performance of this step can be improved by using *plane wave* expansions to diagonalize the translation operators (see [271]). In order to keep matters as simple as possible, we will not discuss this feature.

A more complete algorithm (designed for radial basis function interpolation) has been developed by Beatson and co-workers (see, e.g., [43, 137]).

The numerical experiments in Table 8.11 were conducted by performing quasi-interpolation of the form

$$\mathcal{Q}_h f(x) = \mathcal{D}^{-1/2} \sum_{k=1}^{N} f(x_k) \psi \left(\frac{x - x_k}{\sqrt{\mathcal{D}} h} \right),$$

with a Gaussian ψ on $N = 2^\kappa + 1$ equally spaced points in $[0, 1]$ with the mollified test function

$$f(x) = 15 e^{\frac{-.25}{.25-(x-.5)^2}} \left[\frac{3}{4} e^{-(x-2)^2/4} + \frac{3}{4} e^{-(x+1)^2/49} + \frac{1}{2} e^{-(x-7)^2/4} - \frac{1}{5} e^{-(x-4)^2} \right].$$

All errors were computed on $M = 524289$ equally spaced points in $[0, 1]$. In the "order" column we list the number $order = \ln(e_{\kappa-1}/e_\kappa)/\ln 2$ corresponding to the exponent in the $\mathcal{O}(h^{order})$ notation. Other parameters were $\mathcal{D} = 4$, and the default values for the code of [225] (i.e., $R = 0.5$). All times are measured in seconds.

The asterisk * on the entries in the lower part of the direct column indicate estimated times. The fast Gauss transform yields a speedup of roughly a factor of 300. Another way to interpret these results is that for roughly the same amout of work we can obtain

	direct			fast			
N	ℓ_∞ error	order	time	ℓ_∞ error	order	time	speedup
5	3.018954e-00		1.93	5.495125e-00		1.07	1.80
9	2.037762e-00	0.57	3.40	2.037762e-00	1.43	5.31	0.64
17	9.617170e-01	1.08	6.39	9.617170e-01	1.08	5.33	1.20
33	3.609205e-01	1.41	12.28	3.609205e-01	1.41	5.35	2.30
65	1.190192e-01	1.60	24.72	1.190192e-01	1.60	5.39	4.59
129	3.354132e-02	1.83	53.38	3.354132e-02	1.83	5.46	10.14
257	8.702868e-03	1.95	113.35	8.702868e-03	1.95	5.61	20.20
513	2.196948e-03	1.99	226.15	2.196948e-03	1.99	5.94	38.07
1025			450*	5.505832e-04	2.00	6.67	67.47
2049			900*	1.377302e-04	2.00	7.87	114.36
4097			1800*	3.443783e-05	2.00	10.56	170.45
8193			3600*	8.609789e-06	2.00	15.78	228.14
16385			7200*	2.152468e-06	2.00	26.27	274.08
32769			14400*	5.381182e-07	2.00	47.39	303.86
65537			28800*	1.345296e-07	2.00	89.91	320.32
131073			57600*	3.363241e-08	2.00	174.74	329.63
262145			115200*	8.408103e-09	2.00	343.59	335.28

Table 8.11: 1D quasi-interpolation using fast Gauss transform.

an answer which is about 100000 times more accurate. The $\mathcal{O}(h^2)$ convergence of the Gaussian quasi-interpolant is perfectly illustrated by the entries in the "order" columns.

An alternative to fast multipole methods are so-called *fast tree codes*. These kind of algorithms originated in computational chemistry. We recommend recent papers by Krasny and co-workers (e.g., [160, 385]). The advantage of these kinds of methods is that they make use of standard Taylor expansions instead of the specialized expansions (such as, e.g, in terms of Hermite functions, spherical harmonics, spherical Hankel functions, plane waves, or hypergeometric functions [137]). This simplifies their implementation. However, their convergence properties are probably not as good as for fast mutipole expansions.

We now present a very general discussion of fast summation via Taylor expansions. The presentation of this material is motivated by the work of Krasny and co-workers (see, e.g., [160, 385]) as well as the algorithm for the fast Gauss transform reviewed above. Since we are interested in many evaluations of our quasi-interpolants (or other radial basis function expansion), we split the set of M evaluation points $\boldsymbol{y}_j$ into groups (contained in boxes C with centers $\boldsymbol{y}_C$). We also split the N data locations $\boldsymbol{x}_k$ into boxes B with center $\boldsymbol{x}_B$, and use the index set I_B to denote the points in B.

In order to set the stage for a fast summation of the quasi-interpolant

$$\begin{aligned} \mathcal{Q}f(\boldsymbol{y}_j) &= \sum_{k=1}^{N} f(\boldsymbol{x}_k)\Phi(\boldsymbol{y}_j - \boldsymbol{x}_k) \\ &= \sum_{B} \sum_{k \in I_B} f(\boldsymbol{x}_k)\Phi(\boldsymbol{y}_j - \boldsymbol{x}_k) \end{aligned} \tag{8.24}$$

with generating function Φ we require the multivariate Taylor expansion of Φ about a point $\boldsymbol{z}_0 \in \mathbb{R}^s$, i.e.,

$$\Phi(\boldsymbol{z}) = \sum_{\boldsymbol{\alpha} \geq 0} D^{\boldsymbol{\alpha}} \Phi(\boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{z}_0} \frac{(\boldsymbol{z} - \boldsymbol{z}_0)^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!}, \tag{8.25}$$

where $\boldsymbol{\alpha}$ is a multi-index. Now – as for the fast Gauss transform – we consider three basic expansions.

Theorem 8.5.5 *(Taylor Series Expansion about Centers of Target Boxes) Let I_B be the index set denoting the sources $\boldsymbol{x}_k$ which lie in a box B with center $\boldsymbol{x}_B$, and let $\boldsymbol{y}_C$ be the center of the target box C containing an evaluation point $\boldsymbol{y}_j$. Then the quasi-interpolant due to sources in B*

$$\mathcal{Q}^{(B)} f(\boldsymbol{y}_j) = \sum_{k \in I_B} f(\boldsymbol{x}_k) \Phi(\boldsymbol{y}_j - \boldsymbol{x}_k)$$

can be written as a Taylor expansion about $\boldsymbol{y}_C$:

$$\mathcal{Q}^{(B)} f(\boldsymbol{y}_j) = \sum_{\boldsymbol{\alpha} \geq 0} A_{\boldsymbol{\alpha}}^{(B)} (\boldsymbol{y}_j - \boldsymbol{y}_C)^{\boldsymbol{\alpha}},$$

where

$$A_{\boldsymbol{\alpha}}^{(B)} = \frac{(-1)^{|\boldsymbol{\alpha}|}}{\boldsymbol{\alpha}!} \sum_{k \in I_B} f(\boldsymbol{x}_k) T_{\boldsymbol{\alpha}}(\boldsymbol{y}_C, \boldsymbol{x}_k)$$

with $T_{\boldsymbol{\alpha}}(\boldsymbol{y}_C, \boldsymbol{x}_k) = (-1)^{|\boldsymbol{\alpha}|} D^{\boldsymbol{\alpha}} \Phi(\boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{y}_C - \boldsymbol{x}_k}$.

Proof: We combine the contribution for the source box B of (8.24) with (8.25), and let $\boldsymbol{z} = \boldsymbol{y}_j - \boldsymbol{x}_k$ and $\boldsymbol{z}_0 = \boldsymbol{y}_C - \boldsymbol{x}_k$. Then (8.24) becomes

$$\mathcal{Q}^{(B)} f(\boldsymbol{y}_j) = \sum_{k \in I_B} f(\boldsymbol{x}_k) \sum_{\boldsymbol{\alpha} \geq 0} D^{\boldsymbol{\alpha}} \Phi(\boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{y}_C - \boldsymbol{x}_k} \frac{(\boldsymbol{y}_j - \boldsymbol{y}_C)^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!}.$$

Using the abbreviation $T_{\boldsymbol{\alpha}}(\boldsymbol{y}_C, \boldsymbol{x}_k) = (-1)^{|\boldsymbol{\alpha}|} D^{\boldsymbol{\alpha}} \Phi(\boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{y}_C - \boldsymbol{x}_k}$ we can rewrite this as

$$\mathcal{Q}^{(B)} f(\boldsymbol{y}_j) = \sum_{\boldsymbol{\alpha} \geq 0} A_{\boldsymbol{\alpha}}^{(B)} (\boldsymbol{y}_j - \boldsymbol{y}_C)^{\boldsymbol{\alpha}},$$

where

$$A_{\boldsymbol{\alpha}}^{(B)} = \frac{(-1)^{|\boldsymbol{\alpha}|}}{\boldsymbol{\alpha}!} \sum_{k \in I_B} f(\boldsymbol{x}_k) T_{\boldsymbol{\alpha}}(\boldsymbol{y}_C, \boldsymbol{x}_k).$$

□

Example : If we take $\Phi(\boldsymbol{x}) = e^{-\|\boldsymbol{x}\|^2}$ then

$$T_{\boldsymbol{\alpha}}(\boldsymbol{y}_C, \boldsymbol{x}_k) = h_{\boldsymbol{\alpha}}(\boldsymbol{y}_C - \boldsymbol{x}_k) = h_{\boldsymbol{\alpha}}(\boldsymbol{x}_k - \boldsymbol{y}_C),$$

and Theorem 8.5.5 is equivalent to Theorem 8.5.2 given above.

Remark: We can see that the Taylor expansion has allowed us to separate the evaluation points $\boldsymbol{y}_j$ from the data points $\boldsymbol{x}_k$.

Theorem 8.5.6 *(Reversed Taylor Series Expansion about Centers of Source Boxes) Let I_B be the index set denoting the sources $\boldsymbol{x}_k$ which lie in a box B with center $\boldsymbol{x}_B$. Then the quasi-interpolant due to sources in B*

$$\mathcal{Q}^{(B)} f(\boldsymbol{y}_j) = \sum_{k \in I_B} f(\boldsymbol{x}_k) \Phi(\boldsymbol{y}_j - \boldsymbol{x}_k)$$

can be written as a reversed Taylor expansion about $\boldsymbol{x}_B$:

$$\mathcal{Q}^{(B)} f(\boldsymbol{y}_j) = \sum_{\boldsymbol{\alpha} \geq 0} B_{\boldsymbol{\alpha}}^{(B)} T_{\boldsymbol{\alpha}}(\boldsymbol{y}_j, \boldsymbol{x}_B),$$

with the moments $B_{\boldsymbol{\alpha}}^{(B)}$ *given by*

$$B_{\boldsymbol{\alpha}}^{(B)} = \frac{1}{\boldsymbol{\alpha}!} \sum_{k \in I_B} f(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{x}_B)^{\boldsymbol{\alpha}},$$

and $T_{\boldsymbol{\alpha}}(\boldsymbol{y}_j, \boldsymbol{x}_B) = (-1)^{|\boldsymbol{\alpha}|} D^{\boldsymbol{\alpha}} \Phi(\boldsymbol{z})|_{\boldsymbol{z} = \boldsymbol{y}_j - \boldsymbol{x}_B}$.

Proof: We combine the contribution for the source box B of (8.24) with (8.25), and let $\boldsymbol{z} = \boldsymbol{y}_j - \boldsymbol{x}_k$ and $\boldsymbol{z}_0 = \boldsymbol{y}_j - \boldsymbol{x}_B$. Then (8.24) becomes

$$\mathcal{Q}^{(B)} f(\boldsymbol{y}_j) = \sum_{k \in I_B} f(\boldsymbol{x}_k) \sum_{\boldsymbol{\alpha} \geq 0} D^{\boldsymbol{\alpha}} \Phi(\boldsymbol{z})|_{\boldsymbol{z} = \boldsymbol{y}_j - \boldsymbol{x}_B} (-1)^{|\boldsymbol{\alpha}|} \frac{(\boldsymbol{x}_k - \boldsymbol{x}_B)^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!}.$$

Using the abbreviation $T_{\boldsymbol{\alpha}}(\boldsymbol{y}_j, \boldsymbol{x}_B) = (-1)^{|\boldsymbol{\alpha}|} D^{\boldsymbol{\alpha}} \Phi(\boldsymbol{z})|_{\boldsymbol{z} = \boldsymbol{y}_j - \boldsymbol{x}_B}$ we can reverse the role of the Taylor coefficients and the polynomials to write this as

$$\mathcal{Q}^{(B)} f(\boldsymbol{y}_j) = \sum_{\boldsymbol{\alpha} \geq 0} B_{\boldsymbol{\alpha}}^{(B)} T_{\boldsymbol{\alpha}}(\boldsymbol{y}_j, \boldsymbol{x}_B),$$

with

$$B_{\boldsymbol{\alpha}}^{(B)} = \frac{1}{\boldsymbol{\alpha}!} \sum_{k \in I_B} f(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{x}_B)^{\boldsymbol{\alpha}}.$$

□

Example: Using $\Phi(\boldsymbol{x}) = e^{-\|\boldsymbol{x}\|^2}$ this is equivalent to Theorem 8.5.3.

Remark: The moments can be pre-computed and stored during the setup phase of the algorithm.

Theorem 8.5.7 *(Conversion of Taylor Series Expansions about Source Centers to Series about Target Centers) Let I_B be the index set denoting the sources $\boldsymbol{x}_k$ which lie in a box B with center $\boldsymbol{x}_B$, and let $\boldsymbol{y}_C$ be the center of the target box C containing $\boldsymbol{y}_j$. Then a fast summation formula for the quasi-interpolant*

$$\mathcal{Q} f(\boldsymbol{y}_j) = \sum_{k=1}^{N} f(\boldsymbol{x}_k) \Phi(\boldsymbol{y}_j - \boldsymbol{x}_k)$$

can be given as an expansion about $\boldsymbol{y}_C$*:*

$$\mathcal{Q}f(\boldsymbol{y}_j) \approx \sum_{\boldsymbol{\beta}\le p} C_{\boldsymbol{\beta}}(\boldsymbol{y}_j - \boldsymbol{y}_C)^{\boldsymbol{\beta}},$$

where

$$C_{\boldsymbol{\beta}} = \frac{(-1)^{|\boldsymbol{\beta}|}}{\boldsymbol{\beta}!} \sum_{\boldsymbol{\alpha}+\boldsymbol{\beta}\le p} \sum_{B} T_{\boldsymbol{\alpha}+\boldsymbol{\beta}}(\boldsymbol{y}_C, \boldsymbol{x}_B) B_{\boldsymbol{\alpha}}^{(B)},$$

$T_{\boldsymbol{\alpha}+\boldsymbol{\beta}}(\boldsymbol{y}_C, \boldsymbol{x}_B) = (-1)^{|\boldsymbol{\alpha}+\boldsymbol{\beta}|} D^{\boldsymbol{\alpha}+\boldsymbol{\beta}}\Phi(\boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{y}_C-\boldsymbol{x}_B}$*, and the moments* $B_{\boldsymbol{\alpha}}^{(B)}$ *are as in Theorem 8.5.6.*

Proof: We combine (8.24) with (8.25), and now replace $\boldsymbol{z}$ by $\boldsymbol{y}_j - \boldsymbol{x}_k$ and $\boldsymbol{z}_0$ by $\boldsymbol{y}_C - \boldsymbol{x}_B$. Then (8.24) becomes

$$\mathcal{Q}f(\boldsymbol{y}_j) = \sum_{B} \sum_{k\in I_B} f(\boldsymbol{x}_k) \sum_{\boldsymbol{\alpha}\ge 0} D^{\boldsymbol{\alpha}}\Phi(\boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{y}_C-\boldsymbol{x}_B} \frac{(\boldsymbol{y}_j - \boldsymbol{x}_k - (\boldsymbol{y}_C - \boldsymbol{x}_B))^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!}.$$

Using the abbreviation $T_{\boldsymbol{\alpha}}(\boldsymbol{y}_C, \boldsymbol{x}_B) = (-1)^{|\boldsymbol{\alpha}|} D^{\boldsymbol{\alpha}}\Phi(\boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{y}_C-\boldsymbol{x}_B}$ along with the multivariate binomial theorem we can rewrite this as

$$\begin{aligned}
\mathcal{Q}f(\boldsymbol{y}_j) &= \sum_{B} \sum_{k\in I_B} f(\boldsymbol{x}_k) \sum_{\boldsymbol{\alpha}\ge 0} (-1)^{|\boldsymbol{\alpha}|} \frac{T_{\boldsymbol{\alpha}}(\boldsymbol{y}_C, \boldsymbol{x}_B)}{\boldsymbol{\alpha}!} \sum_{\boldsymbol{\beta}\le\boldsymbol{\alpha}} \binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} (-1)^{|\boldsymbol{\beta}|} (\boldsymbol{y}_j - \boldsymbol{y}_C)^{\boldsymbol{\alpha}-\boldsymbol{\beta}} (\boldsymbol{x}_k - \boldsymbol{x}_B)^{\boldsymbol{\beta}} \\
&= \sum_{\boldsymbol{\alpha}\ge 0} \sum_{B} (-1)^{|\boldsymbol{\alpha}|} T_{\boldsymbol{\alpha}}(\boldsymbol{y}_C, \boldsymbol{x}_B) \sum_{\boldsymbol{\beta}\le\boldsymbol{\alpha}} (-1)^{|\boldsymbol{\beta}|} \frac{(\boldsymbol{y}_j - \boldsymbol{y}_C)^{\boldsymbol{\alpha}-\boldsymbol{\beta}}}{(\boldsymbol{\alpha}-\boldsymbol{\beta})!} \sum_{k\in I_B} f(\boldsymbol{x}_k) \frac{(\boldsymbol{x}_k - \boldsymbol{x}_B)^{\boldsymbol{\beta}}}{\boldsymbol{\beta}!}.
\end{aligned}$$

In fact, we can introduce the moments of Theorem 8.5.6 and write

$$\mathcal{Q}f(\boldsymbol{y}_j) = \sum_{\boldsymbol{\alpha}\ge 0} \sum_{B} (-1)^{|\boldsymbol{\alpha}|} T_{\boldsymbol{\alpha}}(\boldsymbol{y}_C, \boldsymbol{x}_B) \sum_{\boldsymbol{\beta}\le\boldsymbol{\alpha}} (-1)^{|\boldsymbol{\beta}|} \frac{(\boldsymbol{y}_j - \boldsymbol{y}_C)^{\boldsymbol{\alpha}-\boldsymbol{\beta}}}{(\boldsymbol{\alpha}-\boldsymbol{\beta})!} B_{\boldsymbol{\beta}}^{(B)},$$

where

$$B_{\boldsymbol{\beta}}^{(B)} = \frac{1}{\boldsymbol{\beta}!} \sum_{k\in I_B} f(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{x}_B)^{\boldsymbol{\beta}}.$$

A fast algorithm is now obtained by truncating the infinite series after the p-th order terms, i.e.,

$$\mathcal{Q}f(\boldsymbol{y}_j) \approx \sum_{\boldsymbol{\alpha}\le p} \sum_{B} (-1)^{|\boldsymbol{\alpha}|} T_{\boldsymbol{\alpha}}(\boldsymbol{y}_C, \boldsymbol{x}_B) \sum_{\boldsymbol{\beta}\le\boldsymbol{\alpha}} (-1)^{|\boldsymbol{\beta}|} \frac{(\boldsymbol{y}_j - \boldsymbol{y}_C)^{\boldsymbol{\alpha}-\boldsymbol{\beta}}}{(\boldsymbol{\alpha}-\boldsymbol{\beta})!} B_{\boldsymbol{\beta}}^{(B)}.$$

Using the fact that

$$\sum_{\boldsymbol{\alpha}\le p} a_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}\le\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}-\boldsymbol{\beta}} = \sum_{\boldsymbol{\alpha}\le p} b_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\alpha}\le\boldsymbol{\beta}\le p} a_{\boldsymbol{\beta}} = \sum_{\boldsymbol{\alpha}\le p} b_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\alpha}+\boldsymbol{\beta}\le p} a_{\boldsymbol{\alpha}+\boldsymbol{\beta}},$$

which can be verified by a simple rearrangement of the summations and an index transformation, we obtain (interchanging the role of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$) the following fast summation formula:

$$\mathcal{Q}f(\boldsymbol{y}_j) \approx \sum_{\boldsymbol{\beta}\le p} \sum_{\boldsymbol{\alpha}+\boldsymbol{\beta}\le p} (-1)^{|\boldsymbol{\alpha}|} \frac{1}{\boldsymbol{\beta}!} \sum_{B} (-1)^{|\boldsymbol{\alpha}+\boldsymbol{\beta}|} T_{\boldsymbol{\alpha}+\boldsymbol{\beta}}(\boldsymbol{y}_C, \boldsymbol{x}_B) B_{\boldsymbol{\alpha}}^{(B)} (\boldsymbol{y}_j - \boldsymbol{y}_C)^{\boldsymbol{\beta}}.$$

This is equivalent to the statement of the theorem. □

Example: Using $\Phi(\boldsymbol{x}) = e^{-\|\boldsymbol{x}\|^2}$ this is almost equivalent to Theorem 8.5.3. However, our alternate formula is more efficient since only Hermite functions up to order p are required (as opposed to order $2p$ in the Greengard/Strain version). This gain is achieved by using the binomial theorem instead of a second Taylor expansion (the Hermite series expansion used in the traditional fast Gauss transform is equivalent to a Taylor expansion).

Remarks:

1. Note that the Taylor coefficients $T_{\boldsymbol{\alpha}}(\boldsymbol{y}_C, \boldsymbol{x}_B)$ depend only on the box centers $\boldsymbol{y}_C$ and $\boldsymbol{x}_B$.

2. In order to make the algorithm efficient one will use a decision rule (as in Strain's code for the fast Gauss transform) to decide when to use which of the three expansions. Error estimation is very similar to Greengard/Strain. The only difference is that one needs bounds on the Taylor coefficients instead of the Hermite functions.

3. In order to adapt this fast transform to Gauss-Laguerre generating functions of the previous sections (or any other generating function) one needs to compute the required Taylor coefficients. This is a task that goes beyond the scope of this manuscript.

8.6 Domain Decomposition

Finally, another method commonly used to deal with large computational problems is the *domain decomposition* method. The domain decomposition method is frequently implemented on parallel computers in order to speed up the computation even more. A standard reference (based mostly on finite difference and finite element methods) is the book by Smith, Bjørstad and Gropp [584]. For radial basis functions there is a recent paper by Beatson, Light and Billings [42].

The main aim of the paper [42] is to solve the radial basis function interpolation problem discussed multiple times in previous sections. In particular, a so-called *multiplicative Schwarz* algorithm (which is analogous to Gauss-Seidel iteration) is presented, and linear convergence of the algorithm is proved. A section with numerical experiments reports results for an *additive Schwarz* method (which is analogous to Jacobi iteration).

In particular, the authors implemented polyharmonic radial basis functions, and used the scale invariant basis discussed in Section 8.4.

The classical additive Schwarz algorithm is usually discussed in the context of partial differential equations, and it is known that one should add a coarse level correction in order to ensure convergence and to filter out some of the low-frequency oscillations (see, e.g., [584]).

In [42] a two-level additive algorithm for interpolation problems was presented. One begins by subdividing the set on interpolation point $\mathcal{X}$ into M smaller sets $\mathcal{X}_i$, $i = 1, \ldots, M$, whose pairwise intersection is non-empty. The points that belong to one

set $\mathcal{X}_i$ only are called *inner points* of $\mathcal{X}_i$. Those points in the intersection of more than one set need to be assigned in some way as inner points to only one of the subsets $\mathcal{X}_i$ so that the collection of all inner points yields the entire set $\mathcal{X}$. This corresponds to the concept of *overlapping domains*. One also needs to choose a coarse grid $\mathcal{Y}$ that contains points from all of the inner point sets.

In the setup phase of the algorithm the radial basis function interpolation matrices for the smaller problems on each of the subsets $\mathcal{X}_i$, $i = 1, \ldots, M$, are computed and factored. At this point one can use the homogeneous basis of Section 8.4 to ensure numerical stability. Now the algorithm proceeds as follows:

Algorithm:

Input: Data f, point sets $\mathcal{X}_i$ and factored interpolation matrices A_i, $i = 1, \ldots, M$, tolerance ϵ

Initialize $r = f$, $s = 0$

While $\|r\| > \epsilon$ do

For $i = 1$ to m (i.e., for each subset $\mathcal{X}_i$) do

Determine the coefficients $\boldsymbol{c}_i$ of the interpolant to the residual $r|_{\mathcal{X}_i}$ on $\mathcal{X}_i$.

end

Make $\boldsymbol{c}$ orthogonal to Π^s_{m-1}.

Assemble an intermediate approximation $s_1 = \sum_{j=1}^{N} c_j \Phi(\cdot, \boldsymbol{x}_j)$.

Compute the residual on the coarse grid, i.e.,

$$r_1 = r - s_1|_{\mathcal{Y}}.$$

Interpolate to r_1 on the coarse grid $\mathcal{Y}$ using a radial basis function expansion s_2.

Update $s = s + s_1 + s_2$.

Reevaluate the global residual $r = f - s$ on the whole set $\mathcal{X}$

end

Remarks:

1. In [42] it is proved that a multiplicative version of this algorithm converges at least linearly. However, the additive version can be more easily implemented on a parallel computer.

2. If strictly positive definite kernels such as Gaussians are used, then it is not necessary to make the coefficients $\boldsymbol{c}$ orthogonal to polynomials.

3. As in many algorithms before, the evaluation of the residuals needs to be made "fast" using either a fast multipole method or a version of the fast Fourier transform.

4. In the case of very large data sets it may be necessary to use more than two levels so that one ends up with a *multigrid* algorithm.

5. The authors of [42] report having solved interpolation problems with several millions of points using the domain decomposition algorithm above.

6. A number of other papers discussing domain decomposition methods for radial basis functions have recently appeared in the literature (see, e.g., [166, 306, 313, 379, 396, 647]). However, most of these papers contain little theory, focussing mostly on numerical experiments.

Chapter 9

Applications

9.1 Solving Partial Differential Equations via Collocation

In this section we discuss the numerical solution of elliptic partial differential equations using a collocation approach based on radial basis functions. To make the discussion transparent we will focus on the case of a time independent linear elliptic partial differential equation in $\mathbb{R}^2$.

9.1.1 Kansa's Approach

In [340] Kansa suggested a now very popular non-symmetric method for the solution of elliptic PDEs with radial basis functions. In order to be able to clearly point out the differences between Kansa's method and a symmetric approach proposed in [194] we recall some of the basics of scattered data interpolation with radial basis functions in $\mathbb{R}^s$.

In this context we are given data $\{\boldsymbol{x}_i, f_i\}$, $i = 1, \ldots, N$, $\boldsymbol{x}_i \in \mathbb{R}^s$, where we can think of the values f_i being sampled from a function $f : \mathbb{R}^s \to \mathbb{R}$. The goal is to find an interpolant of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{N} c_j \varphi(\|\boldsymbol{x} - \boldsymbol{x}_j\|), \qquad \boldsymbol{x} \in \mathbb{R}^s, \tag{9.1}$$

such that

$$\mathcal{P}f(\boldsymbol{x}_i) = f_i, \qquad i = 1, \ldots, N.$$

The solution of this problem leads to a linear system $A\boldsymbol{c} = \boldsymbol{f}$ with the entries of A given by

$$A_{ij} = \varphi(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \qquad i, j = 1, \ldots, N. \tag{9.2}$$

As discussed earlier, the matrix A is non-singular for a large class of radial functions including (inverse) multiquadrics, Gaussians, and the strictly positive definite compactly supported functions of Wendland, Wu, or Buhmann. In the case of strictly conditionally positive definite functions such as thin plate splines the problem needs to be augmented by polynomials.

We now switch to the collocation solution of partial differential equations. Assume we are given a domain $\Omega \subset \mathbb{R}^s$, and a linear elliptic partial differential equation of the form

$$L[u](\boldsymbol{x}) = f(\boldsymbol{x}), \qquad \boldsymbol{x} \text{ in } \Omega, \tag{9.3}$$

with (for simplicity of description) Dirichlet boundary conditions

$$u(\boldsymbol{x}) = g(\boldsymbol{x}), \qquad \boldsymbol{x} \text{ on } \partial\Omega. \tag{9.4}$$

For Kansa's collocation method we then choose to represent u by a radial basis function expansion analogous to that used for scattered data interpolation, i.e.,

$$u(\boldsymbol{x}) = \sum_{j=1}^{N} c_j \varphi(\|\boldsymbol{x} - \boldsymbol{\xi}_j\|), \tag{9.5}$$

where we now introduce the points $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N$ as centers for the radial basis functions. They will usually be selected to coincide with the collocation points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$. However, the discussion below is clearer if we formally distinguish between centers $\boldsymbol{\xi}_j$ and collocation points $\boldsymbol{x}_i$. We assume the simplest possible setting here, i.e., no polynomial terms are added to the expansion (9.5). The collocation matrix which arises when matching the differential equation (9.3) and the boundary conditions (9.4) at the collocation points $\mathcal{X}$ will be of the form

$$A = \begin{bmatrix} \Phi \\ L[\Phi] \end{bmatrix}, \tag{9.6}$$

where the two blocks are generated as follows:

$$\begin{aligned} \Phi_{ij} &= \varphi(\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|), \qquad \boldsymbol{x}_i \in B,\ \boldsymbol{\xi}_j \in \mathcal{X}, \\ L[\Phi]_{ij} &= L[\varphi](\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|), \qquad \boldsymbol{x}_i \in I,\ \boldsymbol{\xi}_j \in \mathcal{X}. \end{aligned}$$

Here we have identified (as we will do throughout this section) the set of centers with the set of collocation points. The set $\mathcal{X}$ is split into a set I of interior points, and B of boundary points. The problem is well-posed if the linear system $A\boldsymbol{c} = \boldsymbol{y}$, with $\boldsymbol{y}$ a vector consisting of entries $g(\boldsymbol{x}_i)$, $\boldsymbol{x}_i \in B$, followed by $f(\boldsymbol{x}_i)$, $\boldsymbol{x}_i \in I$, has a unique solution.

We note that a change in the boundary conditions (9.4) is as simple as changing a few rows in the matrix A in (9.6) as well as on the right-hand side $\boldsymbol{y}$. We also point out that Kansa only proposed to use multiquadrics in (9.5), and for that method suggested the use of varying parameters α_j, $j = 1, \ldots, N$, which improves the accuracy of the method when compared to using only one constant value of α (see [340]).

A problem with Kansa's method is that – for a constant multiquadric shape parameter α – the matrix A may for certain configurations of the centers $\boldsymbol{\xi}_j$ be singular. Originally, Kansa assumed that the non-singularity results for interpolation matrices would carry over to the PDE case. However, as the numerical experiments of Hon and Schaback [304] show, this is not so. This is to be expected since the matrix for the collocation problem is composed of rows which are built from *different* functions (which – depending on the differential operator L – might not even be radial). The results for

the non-singularity of interpolation matrices, however, are based on the fact that A is generated by a *single* function φ.

An indication of the success of Kansa's method (which has not yet been shown to be well-posed) are the early papers [165, 166, 262, 341, 467] and many more since. In his paper [340] Kansa describes three sets of experiments using his method and comments on the superior performance of multiquadrics in terms of computational complexity and accuracy when compared to finite difference methods. Therefore, it remains an interesting open question whether the well-posedness of Kansa's method can be established at least for certain configurations of centers. Moreover, Kansa's suggestion to use variable shape parameters α_j in order to improve accuracy and stability of the problem has very little theoretical support. Except for one paper by Bozzini, Lenarduzzi and Schaback [68] (which addresses only the interpolation setting) this problem has not been addressed in the literature.

Before we describe an alternate approach which does ensure well-posedness of the resulting collocation matrix and which is based on basis functions suitable for scattered Hermite interpolation we would like to point out that in [467] the authors suggest how Kansa's method can be applied to other types of partial differential equation problems such as non-linear elliptic PDEs, systems of elliptic PDEs, and time-dependent parabolic or hyperbolic PDEs.

9.1.2 An Hermite-based Approach

The following symmetric approach is based on scattered Hermite interpolation (see, e.g., [315, 484, 598, 651]), which we now also quickly review. In this context we are given data $\{\boldsymbol{x}_i, L_i f\}$, $i = 1, \ldots, N$, $\boldsymbol{x}_i \in \mathbb{R}^s$ where $\mathcal{L} = \{L_1, \ldots, L_N\}$ is a linearly independent set of continuous linear functionals. We try to find an interpolant of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{N} c_j L_j^{\boldsymbol{\xi}} \varphi(\|\boldsymbol{x} - \boldsymbol{\xi}\|), \qquad \boldsymbol{x} \in \mathbb{R}^s, \tag{9.7}$$

satisfying

$$L_i \mathcal{P}f = L_i f, \qquad i = 1, \ldots, N.$$

We have used $L^{\boldsymbol{\xi}}$ to indicate that the functional L acts on φ viewed as a function of the second argument $\boldsymbol{\xi}$. The linear system $A\boldsymbol{c} = Lf$ which arises in this case has matrix entries

$$A_{ij} = L_i L_j^{\boldsymbol{\xi}} \varphi, \qquad i, j = 1, \ldots, N. \tag{9.8}$$

In the references mentioned at the beginning of this subsection it is shown that A is non-singular for the same classes of φ as given for scattered data interpolation in our earlier chapters.

Remark: It should be pointed out that this formulation of Hermite interpolation is very general and goes considerably beyond the standard notion of Hermite interpolation (which refers to interpolation of successive derivative values). Here any kind of linear functional are allowed as long as the set $\mathcal{L}$ is linearly independent.

We illustrate this approach with a simple example using derivative functionals.

Example: Let data $\{\boldsymbol{x}_i, f(\boldsymbol{x}_i)\}_{i=1}^n$ and $\{\boldsymbol{x}_i, \frac{\partial f}{\partial x}(\boldsymbol{x}_i)\}_{i=n+1}^N$ with $\boldsymbol{x} = (x, y) \in \mathbb{R}^2$ be given. Then

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{n} c_j \varphi(\|\boldsymbol{x} - \boldsymbol{x}_j\|) - \sum_{j=n+1}^{N} c_j \frac{\partial \varphi}{\partial x}(\|\boldsymbol{x} - \boldsymbol{x}_j\|),$$

and

$$A = \begin{bmatrix} \Phi & -\Phi_x \\ \Phi_x & -\Phi_{xx} \end{bmatrix},$$

with

$$\begin{aligned}
\Phi_{ij} &= \varphi(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \qquad i, j = 1, \ldots, n, \\
-\Phi_{x,ij} &= -\frac{\partial \varphi}{\partial x}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \qquad i = 1, \ldots, n,\ j = n+1, \ldots, N, \\
\Phi_{x,ij} &= \frac{\partial \varphi}{\partial x}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \qquad i = n+1, \ldots, N,\ j = 1, \ldots, n, \\
\Phi_{xx,ij} &= \frac{\partial^2 \varphi}{\partial x^2}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \qquad i, j = n+1, \ldots, N.
\end{aligned}$$

Now we describe an alternative collocation method based on the generalized interpolation theory just reviewed. Assume we are given the same PDE (9.3) with boundary conditions (9.4) as in the section on Kansa's method. In order to be able to apply the results from scattered Hermite interpolation to ensure the non-singularity of the collocation matrix we propose the following expansion for the unknown function u:

$$u(\boldsymbol{x}) = \sum_{j=1}^{\#B} c_j \varphi(\|\boldsymbol{x} - \boldsymbol{\xi}_j\|) + \sum_{j=\#B+1}^{N} c_j L^{\boldsymbol{\xi}}[\varphi](\|\boldsymbol{x} - \boldsymbol{\xi}_j\|), \tag{9.9}$$

where $\#B$ denotes the number of nodes on the boundary of Ω, and $L^{\boldsymbol{\xi}}$ is the differential operator used in (9.3), but acting on φ viewed as a function of the second argument, i.e., $L[\varphi]$ is equal to $L^{\boldsymbol{\xi}}[\varphi]$ up to a possible difference in sign. Note the difference in notation. In (9.7) L is a linear functional, and in (9.9) a differential operator.

This expansion for u leads to a collocation matrix A which is of the form

$$A = \begin{bmatrix} \Phi & L^{\boldsymbol{\xi}}[\Phi] \\ L[\Phi] & L[L^{\boldsymbol{\xi}}[\Phi]] \end{bmatrix}, \tag{9.10}$$

where the four blocks are generated as follows:

$$\begin{aligned}
\Phi_{ij} &= \varphi(\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|), \qquad \boldsymbol{x}_i, \boldsymbol{\xi}_j \in B, \\
L^{\boldsymbol{\xi}}[\Phi]_{ij} &= L^{\boldsymbol{\xi}}[\varphi](\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|), \qquad \boldsymbol{x}_i, \in B,\ \boldsymbol{\xi}_j \in I, \\
L[\Phi]_{ij} &= L[\varphi](\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|), \qquad \boldsymbol{x}_i \in I,\ \boldsymbol{\xi}_j \in B, \\
L[L^{\boldsymbol{\xi}}[\Phi]]_{ij} &= L[L^{\boldsymbol{\xi}}[\varphi]](\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|), \qquad \boldsymbol{x}_i, \boldsymbol{\xi}_j \in I.
\end{aligned}$$

The matrix (9.10) is of the same type as the scattered Hermite interpolation matrices (9.8), and therefore non-singular as long as φ is chosen appropriately. Thus, viewed using the new expansion (9.9) for u, the collocation approach is certainly well-posed.

	α	ρ_K	ρ_H	$\text{cond}_K(A)$	$\text{cond}_H(A)$
5×3	1.0	5.248447e-02	2.004420e-01	2.599606e+03	1.627432e+03
8×4	1.0	1.126843e-02	1.124710e-02	2.325758e+05	8.167527e+04
10×6	1.0	5.809472e-03	6.481697e-03	4.321740e+07	1.808001e+07
16×8	1.0	1.347863e-03	1.720007e-03	8.685785e+10	1.496772e+10
20×12	1.0	5.053090e-04	5.973294e-04	5.161540e+15	1.234633e+15

Table 9.1: Error progression for increasingly denser data sets (Ex.1, fixed α).

Another point in favor of the Hermite based approach is that the matrix (9.10) is (anti)-symmetric as opposed to the completely unstructured matrix (9.6) of the same size. This property should be of value when trying to devise an efficient implementation of the collocation method. Also note that although A consists of four blocks now, it still is of the same size, namely $N \times N$, as the collocation matrix (9.6) obtained for Kansa's approach.

Remark: One attempt to obtain an efficient implementation of the Hermite based collocation method is a version of the greedy algorithm described in Section 8.5.1 by Hon, Schaback and Zhou [305].

9.1.3 Numerical Examples

The following test examples are taken from [194]. We restrict ourselves to two-dimensional Poisson problems whose analytic solution is readily available and therefore can easily be verified. We will refer to a point in $\mathbb{R}^2$ as (x, y). In all of the following tests we used multiquadrics in the expansions (9.5) and (9.9) of the unknown function u.

Example 1: Consider the Poisson equation

$$\Delta u(x, y) = y(1 - y) \sin^3 x, \qquad x \in (0, \pi),\ y \in (0, 1),$$

with Dirichlet boundary conditions

$$u(x, 0) = u(x, 1) = u(0, y) = u(\pi, y) = 0.$$

For this test problem we selected various uniform grids as listed in Tables 9.1 and 9.2 on $[0, \pi] \times [0, 1]$. Tables 9.1 and 9.2 show the values of the multiquadric parameter α, the relative maximum errors ρ computed on a fine grid of 60×60 points, and the approximate condition numbers of A. The range of u on the evaluation grid is approximately $[-0.021023, 0.0]$. The "optimal" value for α was determined by trial and error. The subscripts K and H refer to Kansa's and the Hermite based method, respectively.

Figure 9.1 shows the distribution of the errors $|u(x) - s(x)|$ on the evaluation grid for the two methods on the 8×4 grid used in Table 9.2. The scale used for the shading is displayed on the right.

Example 2: Consider the Poisson equation

$$\Delta u(x, y) = \sin x - \sin^3 x, \qquad x \in (0, \frac{\pi}{2}),\ y \in (0, 2),$$

	α_K	α_H	ρ_K	ρ_H	$\text{cond}_K(A)$	$\text{cond}_H(A)$
5×3	1.18	1.39	1.627193e-02	4.180428e-02	5.592238e+03	5.231279e+03
8×4	1.04	1.11	1.103747e-02	1.062891e-02	3.175078e+05	1.735482e+05
10×6	4.80	3.84	2.739293e-03	3.451799e-03	1.193586e+18	1.414927e+15
16×8	3.12	3.12	2.707006e-04	2.082886e-04	1.209487e+19	6.609375e+18
20×12	2.00	2.30	3.894511e-05	1.273363e-05	3.739554e+19	6.750955e+18

Table 9.2: Error progression for increasingly denser data sets (Ex.1, "optimal" α).

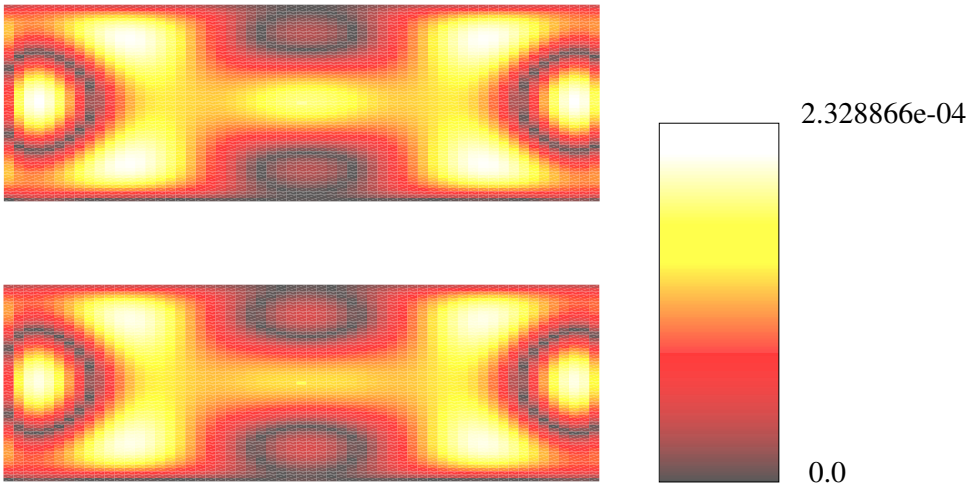

Figure 9.1: Error for Kansa's (top), Hermite (bottom) solution for Ex. 1 on 8×4 grid.

with mixed Dirichlet and Neumann boundary conditions

$$u(0, y) = u_x(\frac{\pi}{2}, y) = u_y(x, 0) = u_y(x, 2) = 0.$$

For this example we selected uniform grids on $[0, \pi/2] \times [0, 2]$ as listed in Table 9.3. This time we only list the results for the "optimal" choice of α. The values listed are analogous to those in Ex. 1.

All in all the Hermite method seems to perform slightly better than Kansa's method. Especially for the cases in which we used relatively many interior points (which is where the methods differ). Also, the matrices for the Hermite method generally have smaller condition numbers. An advantage of the Hermite approach over Kansa's method is

	α_K	α_H	ρ_K	ρ_H	$\text{cond}_K(A)$	$\text{cond}_H(A)$
3×3	109.0	2.19	9.628085e-01	1.141043e-01	1.592286e+16	5.560886e+02
5×5	1.80	1.73	2.181029e-02	4.327029e-02	2.395293e+06	1.271196e+05
7×7	1.58	3.56	6.910084e-03	1.871798e-04	5.762316e+08	1.854850e+12
10×10	2.80	3.29	9.265197e-05	5.126676e-05	2.842111e+18	7.070804e+17
14×14	2.28	2.62	1.138751e-05	1.725526e-06	6.573143e+19	5.891454e+18
20×20	1.53	1.91	5.501057e-06	6.217559e-07	5.889491e+19	7.576112e+19

Table 9.3: Error progression for increasingly denser data sets (Ex.2, "optimal" α).

that for the differential operator L used here, the collocation matrices resulting from the Hermite approach are symmetric. Therefore the amount of computation can be reduced considerably, which is important for larger problems. Kansa's method has the advantage of being simpler to implement (since less derivatives of the basis functions are required).

Remarks:

1. Both of the methods described in this section have been implemented for many different applications. A thorough comparison of the two methods was reported in [520].

2. Since the methods described above were both originally used with globally supported basis functions, the same concerns as for interpolation problems about stability and numerical efficiency apply. Two recent papers by Ling and Kansa [395, 396] address these issues. In particular, they develop a preconditioner in the spirit of the one described in Section 8.3.3, and describe their experience with a domain decomposition algorithm.

3. A convergence analysis for the symmetric method was established by Franke and Schaback [229, 230]. The error estimates established in [229, 230] require the solution of the PDE to be very smooth. Therefore, one should be able to use meshfree radial basis function collocation techniques especially well for (high-dimensional) PDE problems with smooth solutions on possibly irregular domains. Due to the known counterexamples [304] for the non-symmetric method, a convergence analysis is still lacking for that method.

4. Recently, Miranda [462] has shown that Kansa's method will be well-posed if it is combined with so-called *R-functions*. This idea was also used by Höllig and his co-workers in their development of WEB-splines (see, e.g., [299]).

5. Kansa's method has the advantage of being easily adapted for nonlinear elliptic PDEs (see, e.g., [201, 467]).

Some numerical evidence for convergence rates of the symmetric collocation method is given by the examples above, and in the papers [336, 520]. The example above shows very high convergence rates (as predicted by the estimate in [230]) when using multiquadrics on a problem which has a smooth solution. In [336] thin plate splines as well as Wendland's C^4 compactly supported RBF $\varphi_{3,2}$ were tested. The results for thin plate splines are in good agreement with the theory. However, the numerical experiments using the Wendland function show $\mathcal{O}(h^3)$ convergence instead of $\mathcal{O}(h)$ as predicted by the lower bounds of [230] combined with the error bound for Wendland functions. This could suggest that a sharper error estimate may be possible when using compactly supported RBFs.

Other recent papers investigating various aspects of radial basis function collocation are, e.g., [135] by Cheng, Golberg, Kansa and Zammito, [215] by Fedoseyev, Friedman and Kansa, [345] by Kansa and Hon, [360] by Larsson and Fornberg, [365] by Leitão, and [424] by Mai-Duy and Tran-Cong.

For example, in the paper [215] it is suggested that the collocation points on the boundary are also used to satisfy the PDE. However, this adds a set of extra equations

to the problem, and therefore one should also use some additional basis functions in the expansion (9.5). It is suggested in [215] that these centers lie outside the domain Ω. The motivation for this modification is the well-known fact that both for interpolation and collocation with radial basis functions the error is largest near the boundary. In various numerical experiments this strategy is shown to improve the accuracy of Kansa's basic non-symmetric method. It should be noted that there is once more no theoretical foundation for this method.

Larsson and Fornberg [360] compare Kansa's basic collocation method, the modification just described, and the Hermite-based symmetric approach mentioned earlier. Using multiquadric basis functions in a standard implementation they conclude that the symmetric method is the most accurate, followed by the non-symmetric method with boundary collocation. The reason for this is the better conditioning of the system for the symmetric method. Larsson and Fornberg also discuss an implementation of the three methods using the complex Contour-Padé integration method mentioned in Section 8.1. With this technique stability problems are overcome, and it turns out that both the symmetric and the non-symmetric method perform with comparable accuracy. Boundary collocation of the PDE yields an improvement only if these conditions are used as additional equations, i.e., by increasing the problem size. It should also be noted that often the most accurate results were achieved with values of the multiquadric shape parameter α which would lead to severe ill-conditioning using a standard implementation, and therefore these results could be achieved only using the complex integration method. Moreover, in [360] radial basis function collocation is deemed to be far superior in accuracy than standard second-order finite differences or a standard Fourier-Chebyshev pseudospectral method.

Leitão [365] applies the symmetric collocation method to a fourth-order Kirchhoff plate bending problem, and emphasizes the simplicity of the implementation of the radial basis function collocation method. And, finally, Mai-Duy and Tran-Cong [424] suggest a collocation method for which the basis functions are taken to be anti-derivatives of the usual radial basis functions.

All of the experiments just mentioned were conducted without using a multilevel approach. In particular, in order to achieve convergence with the Wendland functions the support had to be chosen so large that only problems with a very modest number of centers could be handled (see [336]). So, as for scattered data interpolation, a multilevel approach is needed to obtain computational efficiency.

We would like to end the discussion of the collocation approach by looking at a multilevel implementation with compactly supported functions.

The most significant difference between the use of compactly supported RBFs for scattered data interpolation and for the numerical solution of PDEs by collocation appears when we turn to the multilevel approach. Recall that the use of the multilevel method is motivated by our desire to obtain a convergent scheme while at the same time keeping the bandwidth fixed, and thus the computational complexity at $\mathcal{O}(N)$.

Here is an adaptation of the basic multilevel algorithm of Section 8.2 to the case of a collocation solution of the problem $Lu = f$:

mesh	ℓ_2-error	rate
5	3.637579e-04	
9	1.892007e-05	4.26
17	3.055339e-06	2.63
33	2.111403e-06	0.53
65	2.062621e-06	0.03
129	2.066411e-06	0.00
257	2.070168e-06	0.00
513	2.072171e-06	0.00
1025	2.073182e-06	0.00
2049	2.073688e-06	0.00

Table 9.4: Multilevel collocation algorithm for symmetric collocation with constant bandwidth.

Algorithm (Multilevel Collocation)

$u_0 = 0$.

For k from 1 to K do

Find $u_k \in \mathcal{S}_{\mathcal{X}_k}$ such that $Lu_k = (f - Lu_{k-1})$ on grid $\mathcal{X}_k$.

Update $u_k \leftarrow u_{k-1} + u_k$.

end

Here $\mathcal{S}_{\mathcal{X}_k}$ is the space of functions used for expansion (9.5) or (9.9) on grid $\mathcal{X}_k$. Whereas we noted above that there is strong numerical (and limited theoretical) evidence that the basic multilevel interpolation algorithm converges (at least linearly), the following example shows that *we cannot in general expect the multilevel collocation algorithm to converge at all.*

Example: Consider the boundary-value problem

$$\begin{aligned} -u''(x) + \pi^2 u(x) &= 2\pi^2 \sin \pi x, \qquad x \in (0,1), \\ u(0) &= u(1) = 0, \end{aligned}$$

with solution $u(x) = \sin \pi x$. As computational grids $\mathcal{X}_k$ we take $2^{k+1} + 1$ uniformly spaced points on $[0, 1]$ as indicated in Table 9.4. We use the C^6 compactly supported Wendland function $\varphi_{3,3}$ and the conjugate gradient method with Jacobi preconditioning is used to solve the resulting linear systems. We take the support size on the first grid to be so large that the resulting matrix is a dense matrix. During subsequent iterations the support size is halved (as is the meshsize) in order to maintain a constant bandwidth of 17 (i.e., work in the stationary setting). Even though the first three iterations seem to indicate significant rates of convergence, the convergence behavior quickly changes, and by the fifth iteration there is virtually no improvement of the error (the fact that the errors actually increase is due to the fact that they are computed on increasingly finer grids).

We note that the same behavior can be observed if the non-symmetric approach is used instead. However, then the convergence ceases at a slightly later stage. We also note that the same phenomenon was observed by Wendland in the context of a multilevel Galerkin algorithm for compactly supported RBFs (see [631] as well as our discussion in the next section).

Remarks:

1. It has been suggested that the convergence behavior of the multilevel collocation algorithm may be linked to the phenomenon of approximate approximation. However, so far no connection has been established.

2. As was shown in [198] a possible remedy for the non-convergence problem is *smoothing*. One might also expect that a slightly different scaling of the support sizes of the basis functions (such that the bandwidth of the matrix is allowed to increase slowly from one iteration to the next, i.e., moving to the non-stationary setting) will lead to better results. In [198] it was shown that this is in fact true. However, smoothing further improved the convergence. A discussion of the idea of post-conditioning via smoothing is beyond the scope of this text. We refer the reader to the paper [209].

9.2 Galerkin Methods

A variational approach to the solution of PDEs with RBFs has so far only been considered by Wendland [630, 631]. In [631] he studies the Helmholtz equation with natural boundary conditions, i.e.,

$$\begin{aligned} -\Delta u + u &= f \qquad \text{in } \Omega, \\ \frac{\partial}{\partial \nu} u &= 0 \qquad \text{on } \partial\Omega, \end{aligned}$$

where ν denotes the outer unit normal vector. The classical Galerkin formulation then leads to the problem of finding a function $u \in H^1(\Omega)$ such that

$$a(u, v) = (f, v)_{L_2(\Omega)} \qquad \text{for all } v \in H^1(\Omega),$$

where $(f, v)_{L_2(\Omega)}$ is the usual L_2 inner product, and for the Helmholtz equation the bilinear form a is given by

$$a(u, v) = \int_\Omega (\nabla u \cdot \nabla v + uv) dx.$$

In order to obtain a numerical scheme the infinite-dimensional space $H^1(\Omega)$ is replaced by some finite-dimensional subspace $\mathcal{S}_\mathcal{X} \subseteq H^1(\Omega)$, where $\mathcal{X}$ is some computational grid to be used for the solution. In the context of RBFs $\mathcal{S}_\mathcal{X}$ is taken as

$$\mathcal{S}_\mathcal{X} = \text{span}\{\phi(\| \cdot - x_j \|_2),\ x_j \in \mathcal{X}\}.$$

This results in a square system of linear equations for the coefficients of $u_{\mathcal{X}} \in \mathcal{S}_{\mathcal{X}}$ determined by

$$a(u_{\mathcal{X}}, v) = (f, v)_{L_2(\Omega)} \qquad \text{for all } v \in \mathcal{S}_{\mathcal{X}}.$$

For more on the Galerkin method (in the context of finite elements) see, e.g., [69, 70]. It was shown in [630] that for those RBFs (globally as well as locally supported) whose Fourier transform decays like $(1 + \| \cdot \|_2)^{-2\beta}$ the following convergence estimate holds:

$$\|u - u_{\mathcal{X}}\|_{H^1(\Omega)} \leq C h^{\sigma-1} \|u\|_{H^\sigma(\Omega)}, \tag{9.11}$$

where h is the meshsize of $\mathcal{X}$, the solution satisfies the regularity requirements $u \in H^\sigma(\Omega)$, and where the convergence rate is determined by $\beta \geq \sigma > s/2 + 1$. For Wendland's compactly supported RBFs this implies that functions which are in $C^{2\kappa}$ and strictly positive definite on $\mathbb{R}^s$ satisfying $\kappa \geq \sigma - \frac{s+1}{2}$ will have $\mathcal{O}(h^{\kappa+(s-1)/2})$ convergence order, i.e., the C^0 function $\varphi_{3,0} = (1-r)_+^2$ yields $\mathcal{O}(h)$ and the C^2 function $\varphi_{3,1} = (1-r)_+^4(4r+1)$ delivers $\mathcal{O}(h^2)$ convergence in $\mathbb{R}^3$. As with the convergence estimate for symmetric collocation there is a link between the regularity requirements on the solution and the space dimension s. Also, so far, the theory is only established for PDEs with natural boundary conditions.

The convergence estimate (9.11) holds for the non-stationary setting, i.e., if we are using compactly supported basis functions, for fixed support radii. By the same argumentation as used in Section 8, one will want to switch to the stationary setting and employ a multilevel algorithm in which the solution at each step is updated by a fit to the most recent residual. This should ensure both convergence and numerical efficiency.

Here is the variant of the stationary multilevel collocation algorithm listed above for the weak formulation (see [631]):

Algorithm (Multilevel Galerkin)

$u_0 = 0$.

For k from 1 to K do

Find $u_k \in \mathcal{S}_{\mathcal{X}_k}$ such that $a(u_k, v) = (f, v) - a(u_{k-1}, v)$ for all $v \in \mathcal{S}_{\mathcal{X}_k}$.

Update $u_k \leftarrow u_{k-1} + u_k$.

end

This algorithm *does not converge in general* (see Tab. 1 in [631]).

Since the weak formulation can be interpreted as a Hilbert space projection method, Wendland was able to show that a modified version of the multilevel Galerkin algorithm, namely

Algorithm (Nested Multilevel Galerkin)

Fix K and $M \in \mathbb{N}$, and set $v_0 = 0$.

For j from 0 while *resiudal* > *tolerance* to M do

Set $u_0 = v_j$.

Apply the k-loop of the previous algorithm and denote the result with $\hat{u}(v_j)$.

Set $v_{j+1} = \hat{u}(v_j)$.

end

does converge. In fact, using this algorithm Wendland proves, and also observes numerically, convergence which is at least linear (see Theorem 3 and Tab. 2 in [631]). The important difference between the two multilevel Galerkin algorithms is the added outer iteration in the nested version which is a well-known idea from linear algebra introduced in 1937 by Kaczmarz [337]. A proof of the linear convergence for general Hilbert space projection methods coupled with Kaczmarz iteration can be found in [585]. This alternate projection idea is also the fundamental ingredient in the convergence proof of the domain decomposition method of Beatson, Light and Billings [42] described in the previous chapter. We mention here that in the multigrid literature Kaczmarz' method is frequently used as a smoother (see e.g. [435]).

Remarks:

1. Aside from difficulties with Dirichlet (or sometimes called *essential*) boundary conditions, Wendland reports that the numerical evaluation of the weak-form integrals presents a major problem for the radial basis function Galerkin approach. Both of these difficulties are also well-known in many other flavors of meshfree weak-form methods. An especially promising solution to the issue of Dirichlet boundary conditions seems to be the use of R-functions as proposed by Höllig and Reif in the context of WEB-splines (see, e.g., [299] or our earlier discussion in the context of collocation methods).

2. In a recent paper by Schaback [559] the author presents a framework for the radial basis function solution of problems both in the strong (collocation) and weak (Galerkin) form.

Many other meshfree methods for the solution of partial differential equations in the weak form appear in the (mostly engineering) literature. These methods come under such names as smoothed particle hydrodynamics (SPH) (e.g., [463]), reproducing kernel particle method (RKPM) (see, e.g., [380, 399]), point interpolation method (PIM) (see, [397]), element free Galerkin method (EFG) (see, e.g., [49]), meshless local Petrov-Galerkin method (MLPG) [14], h-p-cloud method [164], partition of unity finite element method (PUFEM) [16, 443], or generalized finite element method (GFEM) [15]. Most of these methods are based on the moving least squares approximation method discussed in Chapter 7.

There are two recent books by Atluri [12] and Liu [397] summarizing many of these methods. However, these books focus mostly on a survey of the various methods and related computational and implementation issues with little emphasis on the mathematical foundation of these methods. The recent survey paper [15] by Babuška, Banerjee and Osborn, fills a large part of this void.

Bibliography

[1] F.M. Acosta, *Radial basis functions and related models*, Signal Proc., 45 (1995), 37–58.

[2] P. Alfeld, *Scattered data interpolation in three or more variables*, in Mathematical Methods in Computer Aided Geometric Design, T. Lyche and L. Schumaker (eds.), Academic Press, New York, 1989, 1–33.

[3] G. Allasia and P. Giolito, *Fast evaluation of cardinal radial basis interpolants*, in Surface Fitting and Multiresolution Methods, A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press, Nashville TN, 1997, 1–8.

[4] J. Allison, *Multiquadratic radial basis functions for representing multidimensional high energy physics data*, Comp. Phys. Comm., 77 (1993), 377–395.

[5] G. E. Andrews, R. Askey, and R. Roy, *Special Functions*, Cambridge University Press (Cambridge), 1999.

[6] N. Arad, N. Dyn, and D. Reisfeld, *Image warping by radial basis functions: applications to facial expressions*, CVGIP: Graphical models and image processing, 56 (1994), 161–172.

[7] M. G. Armentano, *Error estimates in Sobolev spaces for moving least square approximations*, SIAM J. Numer. Anal. 39 (2001), no. 1, 38–51.

[8] R. Arnott, *An adaptive radial basis function diversity combiner*, IEE Electronics Letters, 29 (1993), 1092–1094.

[9] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc., 68 (1950), 337–404.

[10] R. Askey, *Radial characteristic functions*, TSR #1262, University of Wisconsin-Madison, 1973.

[11] R. Askey, *Orthogonal Polynomials and Special Functions*, Reg. Conf. Ser. in Appl. Math. (SIAM), 1975.

[12] S. N. Atluri and S. Shen, *The Meshless Local Petrov-Galerkin (MLPG) Method*, Tech Science Press, Encino, CA, 2002.

[13] S. N. Atluri and S. Shen, *The meshless local Petrov-Galerkin (MLPG) method: a simple & less costly alternative to the finite element and boundary element methods*, Comput. Model. Eng. Sci. 3 (2002), 11–51.

[14] S. N. Atluri and T. Zhu, *A new meshless local Petrov-Galerkin (MLPG) approach in computational mechanics*, Comput. Mech. 22 (1998), 117–127.

[15] I. Babuška, U. Banerjee and J. E. Osborn, Survey of meshless and generalized finite element methods: A unified approach, University of Texas, Austin, TICAM Report 02-03, 2002.

[16] I. Babuška and M. Melenk, *The partition of unity method*, Int. J. Numer. Meths. Eng. (40) 1997, 727–758.

[17] G. Backus and F. Gilbert, *The resolving power of gross earth data*, Geophys. J. R. Astr. Soc. (16) 1968, 169–205.

[18] K. Ball, *Eigenvalues of Euclidean distance matrices*, J. Approx. Theory, 68 (1992), 74–82.

[19] K. Ball, N. Sivakumar, and J. D. Ward, *On the sensitivity of radial basis interpolation to minimal data separation distance*, Constr. Approx., 8 (1992), 401–426.

[20] R. J. Barnes, *A modified conjugate gradient algorithm for scattered data interpolation using radial basis functions*, J. Appl. Sci. Comput., 1 (1994), 227–238.

[21] R. E. Barnhill and H. S. Ou, *Surfaces defined on surfaces*, Comput. Aided Geom. Design, 7 (1990), 323–336.

[22] A. R. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Trans. Inform. Theory 39 (1993), 930–945.

[23] D. Bates, M. Lindstrom, G. Wahba, and B. Yandell, *GCVPACK - routines for generalized cross validation*, Commun. Statist. B - Simulation and Computation, 16 (1987), 263–297.

[24] D. Bates and G. Wahba, *Computational methods for generalized cross validation with large data sets*, in Treatment of Integral Equations by Numerical Methods, C. T. H. Baker, and G. F. Miller (eds.), Academic Press, New York, 1982, 283–296.

[25] R. Baule, *Moving Least Squares Approximation mit parameterabhängigen Gewichtsfunktionen*, Diplomarbeit, Universität Göttingen, 2000.

[26] B. J. C. Baxter, *Conditionally positive functions and p-norm distance matrices*, Constr. Approx., 7 (1991), 427–440.

[27] B. J. C. Baxter, *Norm estimates for inverses of distance matrices*, in Mathematical Methods in Computer Aided Geometric Design II, T. Lyche and L. Schumaker (eds.), Academic Press, New York, 1992, 9–18.

[28] B. J. C. Baxter, *The asymptotic cardinal function of the multiquadric* $\phi(r) = (r^2 + c^2)^{1/2}$ *as* $c \to \infty$, Comput. Math. Appl., 24 (1992), 1–6.

[29] B. J. C. Baxter, *The interpolation theory of radial basis functions*, University of Cambridge, 1992.

[30] B. J. C. Baxter, *Preconditioned conjugate gradients, radial basis functions, and Toeplitz matrices*, Comput. Math. Appl. 43 (2002), 305–318.

[31] B. J. C. Baxter and G. Roussos, *A new error estimate of the fast Gauss transform*, SIAM J. Sci. Comput. 24 (2002), no. 1, 257–259.

[32] B. J. C. Baxter and N. Sivakumar, *On shifted cardinal interpolation by Gaussians and multiquadrics*, J. Approx. Theory, 87 (1996), 36–59.

[33] B. J. C. Baxter, N. Sivakumar, and J. D. Ward, *Regarding the p-norms of radial basis interpolation matrices*, Constr. Approx., 10 (1994), 451–468.

[34] R. K. Beatson, J. B. Cherrie, and C. T. Mouat, *Fast fitting of radial basis functions: methods based on preconditioned GMRES iteration*, Adv. Comput. Math. 11 (1999), 253–270.

[35] R. K. Beatson and N. Dyn, *Multiquadric B-splines*, J. Approx. Theory, 87 (1996), 1–24.

[36] R. K. Beatson, G. Goodsell, and M. J. D. Powell, *On multigrid techniques for thin plate spline interpolation in two dimensions*, in The Mathematics of Numerical Analysis, Lectures in Appl. Math., 32, Amer. Math. Soc., Providence, RI, 1996, 77–97.

[37] R. K. Beatson and L. Greengard, *A short course on fast multipole methods*, in Wavelets, Multilevel Methods and Elliptic PDEs (Leicester, 1996), M. Ainsworth, J. Levesley, M. Marletta and W. A. Light (eds.), Numer. Math. Sci. Comput., Oxford Univ. Press, New York, 1997, 1–37.

[38] R. K. Beatson and J. Levesley, *Good point/bad point iterations for solving the thin-plate spline interpolation equations*, in Approximation Theory X, C. K. Chui, L. L. Schumaker, and J. Stöckler (eds.), Vanderbilt Univ. Press, Nashville, TN, 2002, 17–25.

[39] R. K. Beatson and W. A. Light, *Quasi-interpolation in the absence of polynomial reproduction*, in Numerical Methods in Approximation Theory, ISNM 105, D. Braess, L. L. Schumaker (ed.), Birkhäuser, Basel, 1992, 21–39.

[40] R. K. Beatson and W. A. Light, *Quasi-interpolation with thin plate splines on a square*, Constr. Approx., 9 (1993), 407–433.

[41] R. K. Beatson and W. A. Light, *Fast evaluation of radial basis functions: methods for two-dimensional polyharmonic splines*, IMA J. Numer. Anal., 17 (1997), 343–372.

[42] R. K. Beatson, W. A. Light, and S. Billings, *Fast solution of the radial basis function interpolation equations: domain decomposition methods*, SIAM J. Sci. Comput. 22 (2000), 1717–1740.

[43] R. K. Beatson and G. N. Newsam, *Fast evaluation of radial basis functions: I*, Comput. Math. Appl., 24 (1992), 7–19.

[44] R. K. Beatson and M. J. D. Powell, *Univariate multiquadric approximation: quasi-interpolation to scattered data*, Constr. Approx., 8 (1992), 275–288.

[45] R. K. Beatson and M. J. D. Powell, *Univariate interpolation on a regular finite grid by a multiquadric plus a linear polynomial*, J. Inst. Math. Applics., 12 (1992), 107–133.

[46] R. K. Beatson and M. J. D. Powell, *An iterative method for thin plate spline interpolation that employs approximations to Lagrange functions*, in Numerical Analysis 1993, Dundee, 1993), xx (ed.), Pitman Res. Notes Math. Ser., 303, Longman Sci. Tech., Harlow, 1994, 17–39.

[47] J. Behrens, A. Iske, and M. Käser, *Adaptive meshfree method of backward characteristics for nonlinear transport equations*, in Lecture Notes in Computer Science and Engineering Vol.26: Meshfree Methods for Partial Differential Equations, M. Griebel and M. A. Schweitzer (eds.), Springer Verlag, 2002, 21–36.

[48] A. Bejancu, Jr., *Local accuracy for radial basis function interpolation on finite uniform grids*, J. Approx. Theory 99 (1999), 242–257.

[49] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming and P. Krysl, *Meshless methods: an overview and recent developments*, Comp. Meth. Appl. Mech. Eng. (139) 1996, 3–47.

[50] B. Ben Moussa, N. Lanson, and J. P. Vila, *Convergence of meshless methods for conservation laws. Applications to Euler equations*, preprint.

[51] C. Berg, J. P. R. Christensen, and P. Ressel, *Harmonic Analysis on Semigroups*, Springer, Berlin, 1984.

[52] A. Beyer, *Optimale Centerverteilung bei Interpolation mit radialen Basisfunktionen*, Diplomarbeit, Universität Göttingen, 1994.

[53] P. Binev and K. Jetter, *Estimating the condition number for multivariate interpolation problems*, in Numerical Methods in Approximation Theory, ISNM 105, D. Braess, L. L. Schumaker (ed.), Birkhäuser, Basel, 1992, 41–52.

[54] N. H. Bingham, *Positive definite functions on spheres*, Proc. Camb. Phil. Soc., 73 (1973), 145–156.

[55] C. Bishop, *Improving the generalization properties of radial basis function neural networks*, Neural Computation, 3 (1991), 579–588.

[56] L. M. Blumenthal, *Distance Geometries*, Univ. of Missouri Studies, 13 (1938), 142pp.

[57] S. Bochner, *Vorlesungen über Fouriersche Integrale*, Akademische Verlagsgesellschaft, Leipzig, 1932.

[58] S. Bochner, *Monotone Funktionen, Stieltjes Integrale und harmonische Analyse*, Math. Ann., 108 (1933), 378–410.

[59] S. Bochner, *Hilbert distances and positive definite functions*, Ann. of Math., 42 (1941), 647–656.

[60] C. de Boor, *Approximation order without quasi-interpolants*, in Approximation Theory VII, E. W. Cheney, C. Chui, and L. Schumaker (eds.), Academic Press, New York, 1992, 1–18.

[61] C. de Boor, R. A. DeVore, and A. Ron, *Approximation from shift-invariant subspaces of* $L_2(\mathbb{R}^d)$, Trans. Amer. Math. Soc., 341 (1994), 787–806.

[62] 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., 119 (1994), 37–78.

[63] C. de Boor and A. Ron, *Fourier analysis of the approximation power of principal shift-invariant spaces*, Constr. Approx., 8 (1992), 427–462.

[64] A. G. Bors and I. Pitas, *Median radial basis function neural network*, IEEE Trans. Neural Networks, 7 (1996), 1351–1364.

[65] L. P. Bos and U. Maier, *On the asymptotics of Fekete-type points for univariate radial basis interpolation*, J. Approx. Theory, 119 (2002), 252–270.

[66] L. P. Bos and K. Šalkauskas, *On the matrix* $\left[|x_i - x_j|^3\right]$ *and the cubic spline continuity equations*, J. Approx. Theory, 51 (1987), 81–88.

[67] L. P. Bos and K. Šalkauskas, *Moving least-squares are Backus-Gilbert optimal*, J. Approx. Theory 59 (1989), 267–275.

[68] M. Bozzini, L. Lenarduzzi, and R. Schaback, *Adaptive interpolation by scaled multiquadrics*, Adv. in Comp. Math. 16 (2002), 375–387.

[69] D. Braess, *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*, Cambridge University Press (Cambridge), 1997.

[70] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, Springer Verlag (New York), 1994.

[71] D. S. Broomhead and D. Lowe, *Multivariate functional interpolation and adaptive networks*, Complex Systems, 2 (1988), 321–355.

[72] A. L. Brown, *Uniform approximation by radial basis functions, Appendix B to Radial basis functions in 1990*, in Advances in Numerical Analysis II: Wavelets, Subdivision, and Radial Basis Functions, W. Light (ed.), Oxford University Press, Oxford, 1992, 203–206.

[73] M. J. Buckley, *Fast computation of a discretized thin-plate smoothing spline for image data*, Biometrika, 81 (1994), 247–258.

[74] M. D. Buhmann, *Convergence of univariate quasi-interpolation using multiquadrics*, IMA J. Numer. Anal., 8 (1988), 365–383.

[75] M. D. Buhmann, *Multivariate interpolation using radial basis functions*, Ph.D. Dissertation, University of Cambridge, 1989.

[76] M. D. Buhmann, *Cardinal interpolation with radial basis functions: an integral approach*, in Multivariate Approximation Theory IV, ISNM 90, C. Chui, W. Schempp, and K. Zeller (eds.), Birkhäuser Verlag, Basel, 1989, 41–64.

[77] M. D. Buhmann, *Multivariate interpolation in odd-dimensional Euclidean spaces using multiquadrics*, Constr. Approx., 6 (1990), 21–34.

[78] M. D. Buhmann, *Multivariate cardinal interpolation with radial basis functions*, Constr. Approx., 6 (1990), 225–255.

[79] M. D. Buhmann, *New developments in the theory of radial basis function interpolation*, in Multivariate Approximation: From CAGD to Wavelets, Kurt Jetter and Florencio Utreras (eds.), World Scientific Publishing, Singapore, 1993, 35–75.

[80] M. D. Buhmann, *On quasi-interpolation with radial basis functions*, J. Approx. Theory, 72 (1993), 103–130.

[81] M. D. Buhmann, *Discrete least squares approximation and prewavelets from radial function spaces*, Proc. Camb. Phil. Soc., 114 (1993), 533–558.

[82] M. D. Buhmann, *Multiquadric prewavelets on nonequally spaced knots in one dimension*, Math. Comp., 64 (1995), 1611–1625.

[83] M. D. Buhmann, *Pre-wavelets on scattered knots and from radial function spaces: a review*, in The Mathematics of Surfaces VI, R. R. Martin (ed.), Clarendon Press, Oxford, 1995, xxx–xxx.

[84] M. D. Buhmann, *Radial functions on compact support*, Proc. Edin. Math. Soc. II 41 (1998), 33-46.

[85] M. D. Buhmann, *Radial basis functions*, in Acta Numerica 2000, Vol. 9, Cambridge University Press (2000), 1–38.

[86] M. D. Buhmann, *Radial Basis Functions : Theory and Implementations*, Cambridge University Press, Cambridge, 2003.

[87] M. D. Buhmann and C. K. Chui, *A note on the local stability of translates of radial basis functions*, J. Approx. Theory, 74 (1993), 36–40.

[88] M. D. Buhmann, F. Derrien, and A. Le Méhauté, *Spectral properties and knot removal for interpolation by pure radial sums*, in Mathematical Methods for Curves and Surfaces, M. Dæhlen, T. Lyche, and L. Schumaker (eds.), Vanderbilt University Press, Nashville, 1995, 55–62.

[89] M. D. Buhmann and N. Dyn, *Error estimates for multiquadric interpolation*, in Curves and Surfaces, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), Academic Press, New York, 1991, 51–58.

[90] M. D. Buhmann and N. Dyn, *Spectral convergence of multiquadric interpolation*, Proc. Edinburgh Math. Soc., 36 (1993), 319–333.

[91] M. D. Buhmann, N. Dyn, and D. Levin, *On quasi-interpolation with radial basis functions on non-regular grids*, Constr. Approx., 11 (1995), 239–254.

[92] M. D. Buhmann and A. Le Méhauté, *Knot removal with radial basis function interpolantion*, C. R. Acad. Sci. Paris Sér. I Math., 320 (1995), 501–506.

[93] M. D. Buhmann and C. A. Micchelli, *Completely monotonic functions for cardinal interpolation*, in Approximation Theory VI, C. Chui, L. Schumaker, and J. Ward (eds.), Academic Press, New York, 1989, 1–4.

[94] M. D. Buhmann and C. A. Micchelli, *Multiply monotone functions for cardinal interpolation*, Adv. in Appl. Math., 12 (1991), 358–386.

[95] M. D. Buhmann and C. A. Micchelli, *Multiquadric interpolation improved*, Comput. Math. Appl., 24 (1992), 21–25.

[96] M. D. Buhmann and C. A. Micchelli, *On radial basis approximation on periodic grids*, Proc. Camb. Phil. Soc., 112 (1992), 317–334.

[97] M. D. Buhmann and A. Pinkus, *On a recovery problem*, Annals of Numerical Mathematics, 4 (1997), 129–142,

[98] M. D. Buhmann and M. J. D. Powell, *Radial basis function interpolation on an infinite regular grid*, in Algorithms for Approximation II, M. G. Cox and J. C. Mason (eds.), Chapman & Hall, London, 1990, 146–169.

[99] M. D. Buhmann and A. Ron, *Radial basis functions: L^p-approximation orders with scattered centres*, in Wavelets, Images, and Surface Fitting, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), A. K. Peters, Wellesley MA, 1994, 93–112.

[100] A. Caiti, G. Magenes, T. Panisini, and R. Simpson, *Smooth approximation by radial basis functions: three case studies*, J. Appl. Sc. Comp., 1 (1994), 88–113.

[101] R. E. Carlson and T. A. Foley, *The parameter R^2 in multiquadric interpolation*, Comput. Math. Appl., 21 (1991), 29–42.

[102] R. E. Carlson and T. A. Foley, *Interpolation of track data with radial basis methods*, Comput. Math. Appl., 24 (1992), 27–34.

[103] R. E. Carlson and B. K. Natarajan, *Sparse approximate multiquadric interpolation*, Comput. Math. Appl., 27 (1994), 99–108.

[104] J. C. Carr, R. K. Beatson, J. B. Cherie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans, *Reconstruction and representation of 3D objects with radial basis functions*, in *SGGRAPH 2001*, 67–76.

[105] I. Cha and S. A. Kassam, *Blind equalisation using radial basis function networks*, Proc. Canadian Conference Elec. Eng. and Comp. Sci., 1 (1992), TM.3.13.14.

[106] I. Cha and S. A. Kassam, *Channel equalization using adaptive complex radial basis function networks*, IEEE J. on Selected Areas in Communications, 13 (1995), 121–131.

[107] I. Cha and S. A. Kassam, *Interference canellation using radial basis function networks*, EURASIP Signal Processing, 147 (1995), 247–268.

[108] S. V. Chakravarthy and J. Ghosh, *Scale-based clustering using the radial basis function network*, IEEE Trans. Neural Networks, 7 (1996), 1250–1261.

[109] A. K. Chan, C. K. Chui, and L. T. Guan, *Radial basis function approach to interpolation of large reflecting surfaces*, in Curves and Surfaces in Computer Vision and Graphics, L. A. Ferrari and R. J. P. de Figueiredo (eds.), SPIE (Vol. 1251), 1990, xxx–xxx.

[110] R. Chan and G. Strang, *Toeplitz equations by conjugate gradients with circulant preconditioners*, SIAM J. Sci. Statist. Comput. 10 (1989), 104–119.

[111] T. F. Chan and D. E. Foulser, *Effectively well-conditioned linear systems*, SIAM J. Sci. Statist. Comput. 9 (1988), 963–969.

[112] K. F. Chang, *Strictly positive definite functions*, J. Approx. Theory, 87 (1996), 148–158. Addendum: J. Approx. Theory, 88 (1997), 384.

[113] C. S. Chen, and C. A. Brebbia, *The dual reciprocity method for Helmholtz-type operators*, in Boundary Elements XX, A. Kassab, C. A. Brebbia and M. Chopra (eds.), Computational Mechanics Publications, 1998, 495–504.

[114] C. S. Chen, C. A. Brebbia and H. Power, *Boundary element methods using compactly supported radial basis functions*, Comm. Numer. Meths. Eng. (15) 1999, 137–150.

[115] C. S. Chen, M. Ganesh, M. A. Golberg, and A. H.-D. Cheng, *Multilevel compact radial functions based computational schemes for some elliptic problems*, Comput. Math. Appl. 43 (2002), 359–378.

[116] C. S. Chen, M. A. Golberg and R. Schaback, *Recent developments in the dual reciprocity method using compactly supported radial basis functions*, preprint, University of Nevada.

[117] J. S. Chen, C. T. Wu, S. Yoon, and Y. You, *A stabilized conforming nodal integration for Galerkin meshfree methods*, Int. J. Numer. Meth. Engng to appear.

[118] S. Chen, *Radial basis functions for signal prediction and system modelling*, J. Appl. Sci. Comput., 1 (1994), 172–207.

[119] S. Chen, S. A. Billings, C. F. N. Cowan, and P. M. Grant, *Non-linear systems identification using radial basis functions*, Int. J. Systems Science, 21 (1990), 2513–2539.

[120] S. Chen, S. A. Billings, and P. M. Grant, *Recursive hybrid algorithm for nonlinear system identification using radial basis function networks*, Int. J. Control, 55 (1992), 1051–1070.

[121] S. Chen, C. F. N. Cowan, and P. M. Grant, *Orthogonal least squares learning algorithm for radial basis function networks*, IEEE Trans. Neural Networks, 2 (1991), 302–309.

[122] S. Chen, G. J. Gibson, C. F. N. Cowan, and P. M. Grant, *Reconstruction of binary signals using an adaptive radial-basis-function equaliser*, EURASIP Signal Processing, 22 (1991), 77–93.

[123] S. Chen, S. McLaughlin, P. M. Grant, and B. Mulgrew, *Reduced-complexity multistage blind clustering equaliser*, IEEE Int. Communications Conference (ICC) Proceedings, 1993, 1149–1153.

[124] S. Chen, S. McLaughlin, and B. Mulgrew, *Complex valued radial basis function networks: network architecture and learning algorithms (part I)*, EURASIP Signal Processing J., 25 (1994), 19–31.

[125] S. Chen, S. McLaughlin, and B. Mulgrew, *Complex valued radial basis function networks: application to digital communications channel equalisation (part II)*, EURASIP Signal Processing J., 36 (1994), 175–188.

[126] S. Chen and B. Mulgrew, *Overcoming co-channel interference using an adaptive radial basis function equaliser*, EURASIP Signal Processing J., 28 (1992), 91–107.

[127] T. Chen and H. Chen, *Approximation capability to functions of several variables, nonlinear functionals and operators by radial basis function neural networks*, IEEE Trans. Neural Networks, 6 (1995), 904–910.

[128] T. Chen and H. Chen, *Denseness of radial-basis functions in* $L^2(\mathbb{R}^n)$ *and its applications in neural networks*, in Approximation Theory VIII, Vol. 1: Approximation and Interpolation, C. Chui, and L. Schumaker (eds.), World Scientific Publishing, Singapore, 1995, 137–144.

[129] E. W. Cheney, *Approximation by functions of nonclassical form*, in Approximation Theory, Spline Functions and Applications, S. P. Singh (ed.), Kluwer, Dordrecht, Netherlands, 1992, 1–18.

[130] E. W. Cheney, *Approximation and interpolation on spheres*, in Approximation Theory, Wavelets and Applications, S. P. Singh (ed.), Kluwer, Dordrecht, Netherlands, 1995, 47–53.

[131] E. W. Cheney, *Approximation using positive definite functions*, in Approximation Theory VIII, Vol. 1: Approximation and Interpolation, C. Chui, and L. Schumaker (eds.), World Scientific Publishing, Singapore, 1995, 145–168.

[132] E. W. Cheney and W. A. Light, *A Course in Approximation Theory*, Brooks Cole, 1999.

[133] E. W. Cheney, W. A. Light, and Y. Xu, *On kernels and approximation orders*, in Proc. Sixth Southeastern Approximation Theory Conference, George Anastassiou (ed.), Lecture Notes in Pure and Applied Mathematics, Vol. 138, 1992, 227–242.

[134] E. W. Cheney and Y. Xu, *A set of research problems in approximation theory*, in Topics in Polynomials of One and Several Variables and Their Applications, T. M. Rassias, H. M. Srivastava, and A. Yanushauskas (eds.), World Scientific Publishers, London, 1993, 109–123.

[135] A. H. D. Cheng, M. A. Golberg, E. J. Kansa, and G. Zammito, *Exponential convergence and H-c multiquadric collocation method for partial differential equations*, Numer. Methods Partial Differential Equations 19 (2003), no. 5, 571–594.

[136] C. C. Cheng and Y. F. Zheng, *Thin plate spline surface approximation using Coons' patches*, Comput. Aided Geom. Design, 11 (1994), 269–287.

[137] J. B. Cherrie, R. K. Beatson, and G. N. Newsam, *Fast evaluation of radial basis functions: methods for generalized multiquadrics in* $\mathbb{R}^n$, SIAM J. Sci. Comput. 23 (2002), 1549–1571.

[138] E. S. Chng, S. Chen, and B. Mulgrew, *Gradient radial basis function networks for nonlinear and nonstationary time series prediction*, IEEE Trans. Neural Networks, 7 (1996), 190–194.

[139] E. S. Chng, S. Chen, B. Mulgrew, and G. J. Gibson, *Realising the Bayesian decision boundary for channel equalisation using radial basis function network and linear equaliser*, Mathematics for Neural Networks and Applications, Oxford, 1995.

[140] C. K. Chui, *Multivariate Splines*, Reg. Conf. Ser. in Appl. Math. (SIAM), 1988.

[141] C. K. Chui, J. Stöckler, and J. D. Ward, *Analytic wavelets generated by radial functions*, Adv. Comput. Math., 5 (1996), 95–123.

[142] C. K. Chui, J. D. Ward, and K. Jetter, *Cardinal interpolation with differences of tempered functions*, Comput. Math. Appl., 24 (1992), 35–48.

[143] K. C. Chung and T. H. Yao, *On lattices admitting unique Lagrange interpolations*, SIAM J. Numer. Anal., 14 (1977), 735–743.

[144] J. Cid-Sueira, A. Artes-Rodriguez, and A. R. Figueiras-Vidal, *Recurrent radial basis function networks for optimal symbol-by-symbol equalisation*, EURASIP Signal Processing, 40 (1994), 53–63.

[145] J. Cid-Sueiro and A. R. Figueiras-Vidal, *Recurrent radial basis function networks for optimal equalisation*, Proc. IEEE Workshop on Neural Networks for Signal Processing, 1993, 562–571.

[146] W. S. Cleveland and C. L. Loader, *Smoothing by local regression: Principles and methods*, in Statistical Theory and Computational Aspects of Smoothing, W. Haerdle and M. G. Schimek (eds.), Springer, (1996), 10–49.

[147] R. Cuppens, *Decomposition of Multivariate Probability*, Academic Press (New York), 1975.

[148] E. L. De Forest, *On some methods of interpolation applicable to the graduation of irregular series*, Annual Report of the Board of Regents of the Smithsonian Institution for 1871, (1873) 275–339.

[149] E. L. De Forest, *Additions to a memoir on methods of interpolation applicable to the graduation of irregular series*, Annual Report of the Board of Regents of the Smithsonian Institution for 1873, (1874) 319–353.

[150] F. J. Delvos, *Convergence of interpolation by translation*, Colloq. Math. Soc. J. Bolya, 49 (1985), 273–287.

[151] F. J. Delvos, *Periodic interpolation on uniform meshes*, J. Approx. Theory, 51 (1987), 71–80.

[152] S. De Marchi, R. Schaback, and H. Wendland, *Optimal data-independent point locations for radial basis function interpolation*, preprint, Universität Göttingen, 2003.

[153] G. Dilts, *Equivalence of the SPH method and a space-time Galerkin moving particle method*, Los Alamos National Laboratory, 1996.

[154] J. G. Dix and R. D. Ogden, *An interpolation scheme with radial basis in Sobolev spaces* $H^s(\mathbb{R}^n)$, Rocky Mountain J. Math., 24 (1994), 1319–1337.

[155] D. Djokovič and R. van Damme, *Radial splines and moving grids*, preprint, 1996.

[156] D. Djokovič and R. van Damme, *Moving node methods for PDE's using radial basis functions and B-splines*, preprint, 1997.

[157] J. Dolbow and T. Belytschko, *An introduction to programming the meshless element free Galerkin method*, Northwestern University, 1997.

[158] W. F. Donoghue, *Monotone Matrix Functions and Analytic Continuation*, Springer (Berlin), 1974.

[159] T. A. Driscoll and B. Fornberg, *Interpolation in the limit of increasingly flat radial basis functions*, Comput. Math. Appl. 43 (2002), 413–422.

[160] Z.-H. Duan, and R. Krasny, *An adaptive treecode for computing nonbonded potential energy in classical molecular systems*, J. Comput. Chemistry 22 (2001) 184–195.

[161] C. A. Duarte, *A review of some meshless methods to solve partial differential equations*, University of Texas, 1995.

[162] C. A. Duarte and J. T. Oden, *Hp clouds – a meshless method to solve boundary-value problems*, University of Texas, 1995.

[163] C. A. Duarte and J. T. Oden, *An h-p adaptive method using clouds*, University of Texas, 1995.

[164] C. A. Duarte and J. T. Oden, *H-p clouds – an h-p meshless method*, Num. Meth. for Part. Diff. Eq. 4, John Wiley & Sons, 1996.

[165] M. R. Dubal, *Construction of three-dimensional black-hole initial data via multiquadrics*, Phys. Rev. D, 45 (1992), 1178–1187.

[166] M. R. Dubal, *Domain decomposition and local refinement for multiquadric approximations. I: Second-order equations in one-dimension*, J. Appl. Sc. Comp., 1 (1994), 146–171.

[167] M. R. Dubal, S. R. Oliveira, and R. A. Matzner, *Solution of elliptic equations in numerical relativity using multiquadrics*, in Approaches to Numerical Relativity, R. d'Inverno (ed.), Cambridge University Press, Cambridge, 1992, 265–280.

[168] J. Duchon, *Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces*, Rev. Francaise Automat. Informat. Rech. Opér., Anal. Numer., 10 (1976), 5–12.

[169] J. Duchon, *Splines minimizing rotation-invariant semi-norms in Sobolev spaces*, in Constructive Theory of Functions of Several Variables, Oberwolfach 1976, W. Schempp and K. Zeller (eds.), Springer Lecture Notes in Math. 571, Springer-Verlag, Berlin, 1977, 85–100.

[170] J. Duchon, *Sur l'erreur d'interpolation des fonctions de plusieurs variables par les D^m-splines*, Rev. Francaise Automat. Informat. Rech. Opér., Anal. Numer., 12 (1978), 325–334.

[171] J. Duchon, *Fontions splines homogènes à plusiers variables*, Université de Grenoble, 1980.

[172] N. Dyn, *Interpolation of scattered data by radial functions*, in Topics in Multivariate Approximation, C. K. Chui, L. L. Schumaker, and F. Utreras (eds.), Academic Press, New York, 1987, 47–61.

[173] N. Dyn, *Interpolation and approximation by radial and related functions*, in Approximation Theory VI, C. Chui, L. Schumaker, and J. Ward (eds.), Academic Press, New York, 1989, 211–234.

[174] N. Dyn, T. Goodman, and C. A. Micchelli, *Positive powers of certain conditionally negative definite matrices*, Indaga. Math. Proc. Ser.A., 89 (1986), 163–178.

[175] N. Dyn, I. R. H. Jackson, D. Levin, and A. Ron, *On multivariate approximation by the integer translates of a basis function*, Israel J. Math., 78 (1992), 95–130.

[176] N. Dyn and D. Levin, *Bell shaped basis functions for surface fitting*, in Approximation Theory and Applications, Z. Ziegler (ed.), Academic Press, New York, 1981, 113–129.

[177] N. Dyn and D. Levin, *Construction of surface spline interpolants of scattered data over finite domains*, Rev. Francaise Automat. Informat. Rech. Opér., Anal. Numer., 16 (1982), 201–209.

[178] N. Dyn and D. Levin, *Iterative solution of systems originating from integral equations and surface interpolation*, SIAM J. Numer. Anal., 20 (1983), 377–390.

[179] N. Dyn, D. Levin, and S. Rippa, *Surface interpolation and smoothing by "thin plate" splines*, in Approximation Theory IV, C. Chui, L. Schumaker, and J. Ward (eds.), Academic Press, New York, 1983, 445–449.

[180] N. Dyn, D. Levin, and S. Rippa, *Numerical procedures for surface fitting of scattered data by radial functions*, SIAM J. Sci. Statist. Comput., 7 (1986), 639–659.

[181] N. Dyn, W. A. Light, and E. W. Cheney, *Interpolation by piecewise-linear radial basis functions, I*, J. Approx. Theory, 59 (1989), 202–223.

[182] N. Dyn and C. A. Micchelli, *Interpolation by sums of radial functions*, Numer. Math., 58 (1990), 1–9.

[183] N. Dyn and A. Ron, *Radial basis function approximation: from gridded centers to scattered centers*, Proc. London Math. Soc., 71 (1995), 76–108.

[184] N. Dyn and A. Ron, *Multiresolution analysis by infinitely differentiable compactly supported functions*, Appl. Comput. Harmon. Anal., 2 (1995), 15–20.

[185] T. Evgeniou, M. Pontil, and T. Poggio, *Regularization networks and support vector machines*, Adv. Comput. Math. 13 (2000), no. 1, 1–50.

[186] J. Fan and I. Gijbels, *Local Polynomial Modelling and its Applications*, Chapman & Hall (New York), 1996.

[187] R. Farwig, *Multivariate interpolation of arbitrarily spaced data by moving least squares methods*, J. Comput. Appl. Math. 16 (1986), 79–93.

[188] R. Farwig, *Multivariate interpolation of scattered data by moving least squares methods*, Algorithms for Approximation, Oxford Univ. Press, New York, 193–211, 1987.

[189] R. Farwig, *Rate of convergence of moving least squares interpolation methods: the univariate case*, Progress in Approximation Theory, Academic Press, Boston, MA, 313–327, 1991.

[190] G. E. Fasshauer, *Scattered data interpolation with radial basis functions on the sphere*, Vanderbilt University, 1994.

[191] G. E. Fasshauer, *On the density of certain classes of radial basis functions on spheres*, Vanderbilt University, 1994.

[192] G. E. Fasshauer, *Adaptive least squares fitting with radial basis functions on the sphere*, in Mathematical Methods for Curves and Surfaces, M. Dæhlen, T. Lyche, and L. Schumaker (eds.), Vanderbilt University Press, Nashville, 1995, 141–150.

[193] G. E. Fasshauer, *Radial Basis Functions on Spheres*, Ph.D. Dissertation, Vanderbilt University, 1995.

[194] G. E. Fasshauer, *Solving partial differential equations by collocation with radial basis functions*, in Surface Fitting and Multiresolution Methods, A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press, Nashville TN, 1997, 131–138.

[195] G. E. Fasshauer, *On smoothing for multilevel approximation with radial basis functions*, in Approximation Theory IX, Vol.II: Computational Aspects, Charles K. Chui, and L. L. Schumaker (eds.), Vanderbilt University Press, 1999, 55–62.

[196] G. E. Fasshauer, *Hermite interpolation with radial basis functions on spheres*, Adv. in Comp. Math. (10) 1999, 81–96.

[197] G. E. Fasshauer, *On the numerical solution of differential equations with radial basis functions*, in Boundary Element Technology XIII, C. S. Chen, C. A. Brebbia, and D. W. Pepper (eds.), WIT Press, 1999, 291–300.

[198] G. E. Fasshauer, *Solving differential equations with radial basis functions: multilevel methods and smoothing*, Adv. in Comp. Math. (11) 1999, 139–159.

[199] G. E. Fasshauer, *Nonsymmetric multilevel RBF collocation within an operator Newton framework for nonlinear PDEs*, in Trends in Approximation Theory, K. Kopotun, T. Lyche, and M. Neamtu (eds.), Vanderbilt University Press, 2001, 103–112.

[200] G. E. Fasshauer, *High-order moving least-squares approximation via fast radial Laguerre transforms*, in preparation, 32 pp..

[201] G. E. Fasshauer, *Newton iteration with multiquadrics for the solution of nonlinear PDEs*, Comput. Math. Applic. (43) 2002, 423–438.

[202] G. E. Fasshauer, *Matrix-free multilevel moving least-squares methods*, in Approximation Theory X: Wavelets, Splines, and Applications, C. K. Chui, L. L. Schumaker, and J. Stöckler (eds.), Vanderbilt University Press, 2002, 271–281.

[203] G. E. Fasshauer, *Approximate moving least-squares approximation with compactly supported weights*, in Lecture Notes in Computer Science and Engineering Vol.26: Meshfree Methods for Partial Differential Equations, M. Griebel and M. A. Schweitzer (eds.), Springer Verlag, 2002, 105-116.

[204] G. E. Fasshauer, *Approximate moving least-squares approximation for time-dependent PDEs*, in WCCM V, Fifth World Congress on Computational Mechanics (http://wccm.tuwien.ac.at), H. A. Mang, F. G. Rammerstorfer, and J. Eberhardsteiner (eds.), Vienna University of Technology, 2002.

[205] G. E. Fasshauer, *Approximate moving least-squares approximation: A fast and accurate multivariate approximation method*, in Curve and Surface Fitting: Saint-Malo 2002, A. Cohen, J.-L. Merrien, and L. L. Schumaker (eds.), Nashboro Press, 2003, 139–148.

[206] G. E. Fasshauer, *Toward approximate moving least squares approximation with irregularly spaced centers*, to appear in Computer Methods in Applied Mechanics & Engineering.

[207] G. E. Fasshauer, E. C. Gartland and J. W. Jerome, *Algorithms defined by Nash iteration: some implementations via multilevel collocation and smoothing*, J. Comp. Appl. Math. (119) 2000, 161–183.

[208] G. E. Fasshauer, E. C. Gartland and J. W. Jerome, *Newton iteration for partial differential equations and the approximation of the identity*, Numerical Algorithms (25) 2000, 181–195.

[209] G. E. Fasshauer and J. W. Jerome, *Multistep approximation algorithms: Improved convergence rates through postconditioning with smoothing kernels*, Adv. in Comput. Math. (10) 1999, 1–27.

[210] G. E. Fasshauer, A. Q. M. Khaliq, and D. A. Voss, *Using meshfree approximation for multi-asset American option problems*, submitted.

[211] G. E. Fasshauer and L. L. Schumaker, *Scattered data fitting on the sphere*, in Mathematical Methods for Curves and Surfaces II, M. Dæhlen, T. Lyche, and L. L. Schumaker (eds.), Vanderbilt University Press, 1998, 117–166.

[212] A. C. Faul and M. J. D. Powell, *Proof of convergence of an iterative technique for thin plate spline interpolation in two dimensions*, preprint, 1998.

[213] A. C. Faul and M. J. D. Powell, *Krylov subspace methods for radial basis function interpolation*, Numerical Analysis 1999 (Dundee), Chapman & Hall/CRC, Boca Raton, FL, 2000, 115–141,

[214] A. I. Fedoseyev, M. J. Friedman and E. J. Kansa, *Continuation for nonlinear elliptic partial differential equations discretized by the multiquadric method*, preprint, 1998.

[215] A. I. Fedoseyev, M. J. Friedman and E. J. Kansa, *Improved multiquadric method for elliptic partial differential equations via PDE collocation on the boundary*, Comput. Math. Applic. 43 (2002), 439–455.

[216] M. S. Floater and A. Iske, *Thinning and approximation of large sets of scattered data*, in Advanced Topics in Multivariate Approximation, F. Fontanella, K. Jetter, and P.-J. Laurent (eds.), World Scientific Publishing, Singapore, 1996, xxx–xxx.

[217] M. S. Floater and A. Iske, *Multistep scattered data interpolation using compactly supported radial basis functions*, J. Comput. Applied Math., 73 (1996), 65–78.

[218] T. A. Foley, *Interpolation and approximation of* 3*-D and* 4*-D scattered data*, Comput. Math. Appl., 13 (1987), 711–740.

[219] T. A. Foley, *Interpolation of scattered data on a spherical domain*, in Algorithms for Approximation II, M. G. Cox and J. C. Mason (eds.), Chapman & Hall, London, 1990, 303–310.

[220] T. A. Foley, *The map and blend scattered data interpolant on the sphere*, Comput. Math. Appl., 24 (1992), 49–60.

[221] T. A. Foley, *Near optimal parameter selection for multiquadric interpolation*, J. Appl. Sc. Comp., 1 (1994), 54–69.

[222] T. A. Foley, S. Dayanand, and D. Zeckzer, *Localized radial basis methods using rational triangle patches*, in Geometric Modelling – Dagstuhl 1993, H. Hagen, G. Farin and H. Noltemeier (eds.), Springer Verlag, Berlin, 1995, 163–176.

[223] T. A. Foley, D. A. Lane, G. M. Nielson, R. Franke, and H. Hagen, *Interpolation of scattered data on closed surfaces*, Comput. Aided Geom. Design, 7 (1990), 303–312.

[224] T. A. Foley, D. A. Lane, G. M. Nielson, R. Ramaraj, *Visualizing functions over a sphere*, Comp. Graphics and Applics., 10 (1990), 32–40.

[225] A. G. Florence C. F. van Loan, *A Kronecker product formulation of the fast Gauss transform*, preprint 2000.

[226] B. Fornberg, T. A. Driscoll, G. Wright, and R. Charles, *Observations on the behavior of radial basis function approximations near boundaries*, Comput. Math. Appl. 43 (2002), 473–490.

[227] B. Fornberg and N. Flyer, *Accuracy of radial basis function interpolation and derivative approximations on 1-D infinite grids*, preprint, University of Colorado, 2003.

[228] B. Fornberg and G. Wright, *Stable computation of multiquadric interpolants for all values of the shape parameter*, preprint, University of Colorado.

[229] C. Franke and R. Schaback, *Solving partial differential equations by collocation using radial basis functions*, Appl. Math. Comp. (93) 1998, 73–82.

[230] C. Franke and R. Schaback, *Convergence orders of meshless collocation methods using radial basis functions*, Adv. in Comput. Math. (8) 1998, 381–399.

[231] R. Franke, *Scattered data interpolation: tests of some methods*, Math. Comp., 48 (1982), 181–200.

[232] R. Franke, *Smooth interpolation of scattered data by local thin plate splines*, Comput. Math. Appl., 8 (1982), 273–281.

[233] R. Franke, *Thin plate splines with tension*, Comput. Aided Geom. Design, 2 (1985), 87–95.

[234] R. Franke, *Recent advances in the approximation of surfaces from scattered data*, in Topics in Multivariate Approximation, C. K. Chui, L. L. Schumaker, and F. Utreras (eds.), Academic Press, New York, 1987, 79–98.

[235] R. Franke, *Approximation of scattered data for meteorological applications*, in Multivariate Approximation and Interpolation, ISNM 94, W. Haussman and K. Jetter (eds.), Birkhäuser, Basel, 1990, 107–120.

[236] R. Franke, *Paper presented at SIAM Geometric Modeling Conference in Nashville Nov. 1995*, SIAM, 1995.

[237] R. Franke, H. Hagen, and G. M. Nielson, *Least squares surface approximation to scattered data using multiquadric functions*, Adv. in Comput. Math., 2 (1994), 81-99.

[238] R. Franke, H. Hagen, and G. M. Nielson, *Repeated knots in least squares multiquadric functions*, in Geometric Modelling – Dagstuhl 1993, H. Hagen, G. Farin and H. Noltemeier (eds.), Springer Verlag, Berlin, 1995, 177-187.

[239] R. Franke and G. M. Nielson, *Scattered data interpolation and applications: A tutorial and survey*, in Geometric Modeling, Methods and Applications, H. Hagen and D. Roller (eds.), Springer Verlag, Berlin, 1991, 131–160.

[240] R. Franke and K. Šalkauskas, *Localization of multivariate interpolation and smoothing methods*, Paper presented at SIAM Geometric Modeling Conference, Nashville, 1995.

[241] W. Freeden, *Spherical spline interpoaltion – basic theory and computational aspects*, J. Comput. Appl. Math., 11 (1984), 367–375.

[242] W. Freeden, T. Gervens, and M. Schreiner, *Constructive Approximation on the Sphere*, Oxford University Press, Oxford, 1998.

[243] W. Freeden, M. Schreiner, and R. Franke, *A survey on spherical spline approximation*, Surv. Math. Ind. (7) 1997, 29–85.

[244] J. A. Freeman and D. Saad, *Learning and generalization in radial basis function networks*, Neural Computation, 7 (1995), 1000-1020.

[245] M. Frigo and S. G. Johnson, *FFTW: The fastest Fourier transform in the West (C library)*, `http://www.fftw.org/`.

[246] E. A. Galperin, E. J. Kansa, A. Makroglou, and S. A. Nelson, *Numerical solutions of weakly singular Volterra equations using global optimization techniques and radial basis functions*, University of Quebec, 1996.

[247] E. A. Galperin and Q. Zheng *Solution and control of PDE via global optimization methods*, Comput. Math. Appl., 25 (1993), 103–118.

[248] G. Gasper, *Positivity and special functions*, in Theory and Application of Special Functions, R. Askey (ed.), Academic Press, New York, 1975, 375–433.

[249] I. M. Gel'fand and G. E. Schilow, *Verallgemeinerte Funktionen, Vol. 1*, Deutscher Verlag der Wissenschaften, Leipzig, 1960.

[250] I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions Vol. 4*, Academic Press, New York, 1964.

[251] F. Girosi, *Some extensions of radial basis functions and their applications in artificial intelligence*, Comput. Math. Appl., 24 (1992), 61–80.

[252] F. Girosi, *An equivalence between sparse approximation and support vector machines*, MIT Artificial Intelligence Lab, Report 1606, 1997.

[253] F. Girosi and G. Anzellotti, *Rates of convergence for radial basis functions and neural networks*, in Artificial Networks for Speech and Vision, R. J. Mammone (ed.), Chapman & Hall, London, 1993, 97–114.

[254] F. Girosi, M. Jones, and T. Poggio, *Regularization theory and neural network architectures*, MIT AI memo 1430, 1993.

[255] M. A. Golberg, *A note on the sparse representation of discrete integral operators*, Appl. Math. Comp., 70 (1995), 97–118.

[256] M. A. Golberg, *Recent developments in the numerical evaluation of particular solutions in the boudary element method*, Appl. Math. Comp., 75 (1996), 91–101.

[257] M. A. Golberg and C. S. Chen, *The theory of radial basis functions applied to the BEM for inhomogeneous partial differential equations*, Boundary Elements Comm., 5 (1994), 57–61.

[258] M. A. Golberg and C. S. Chen, *A bibliography on radial basis function approximation*, UNLV, 1996.

[259] M. A. Golberg and C. S. Chen, *Discrete Projection Methods for Integral Equations*, Computational Mechanics Publications (Southampton), 1997.

[260] M. A. Golberg, C. S. Chen, and H. Bowman, *Some recent results and proposals for the use of radial basis functions in the BEM*, Eng. Anal. with Bound. Elem. (23) 1999, 285–296.

[261] M. A. Golberg, C. S. Chen, H. Bowman, and H. Power, *Some comments on the use of radial basis functions in the dual reciprocity method*, Comput. Mech. (21) 1998, 141–148.

[262] M. A. Golberg, C. S. Chen, and S. R. Karur, *Improved multiquadric approximation for partial differential equations*, Eng. Anal. with Bound. Elem. (18) 1996, 9–17.

[263] M. von Golitschek and L. L. Schumaker, *Data fitting by penalized least squares*, in Algorithms for Approximation II, M. G. Cax and J. C. Mason (eds.), Chapman & Hall, London, 1990, 210–227.

[264] M. Golomb and H. F. Weinberger, *Optimal approximation and error bounds*, in On numerical approximation, R. E. Langer(ed.), University of Wisconsin Press, 1959, 117–190.

[265] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1989.

[266] G. Goodsell, *A multigrid-type method for thin plate spline interpolation on a circle*, IMA J. Numer. Anal., 17 (1997), 321–327.

[267] J. P. Gram, *Über Entwicklung reeler Functionen in Reihen mittelst der Methode der kleinsten Quadrate*, J. Math. 94 (1883), 41–73.

[268] L. Greengard, *Fast algorithms for classical physics*, Science 265 (1994), 909–914.

[269] L. Greengard and V. Rokhlin, *A fast algorithm for particle simulations*, J. Comput. Phys. 73 (1987), no. 2, 325–348.

[270] L. Greengard and J. Strain, *The fast Gauss transform*, SIAM J. Sci. Statist. Comput. (12) 1991, 79–94.

[271] L. F. Greengard and X. Sun, *A new version of the fast Gauss transform*, Doc. Math. J. DMV, Extra Volume ICM, 1998, 575–584.

[272] C. Gu and G. Wahba, *Semiparametric analysis of variance with tensor product thin plate splines*, J. Roy. Statist. Soc. Ser. B, 55 (1993), 353–368.

[273] K. Guo, S. Hu, and X. Sun, *Conditionally positive definite functions and Laplace-Stieltjes integrals*, J. Approx. Theory, 74 (1993), 249–265.

[274] K. Guo, S. Hu, and X. Sun, *Cardinal interpolation using linear combinations of translates of conditionally positive definite functions*, Numer. Func. Anal. Optim., 14 (1993), 371–381.

[275] K. Guo and X. Sun, *Scattered data interpolation by linear combinations of translates of conditionally positive definite functions*, Numer. Func. Anal. Optim., 12 (1991), 137–152.

[276] T. Gutzmer, *Interpolation by positive definite functions on locally compact groups with application to* SO(3), Results Math., 29 (1996), 69–77.

[277] R. E. Hagan, *Using multiquadric and general finite differences in ground water modeling*, Utah State University, 19xx.

[278] R. E. Hagan, *Multiquadrics in engineering modeling*, in Proceedings ASCE National Conference on Irrigation and Drainage Engineering, Park City, Utah, 1993.

[279] R. E. Hagan and E. J. Kansa, *Studies of the R parameter in the multiquadric function applied to ground water pumping*, J. Appl. Sci. Comput., 1 (1994), 266–281.

[280] S. J. Hales and J. Levesley, *On compactly supported, positive definite, radial basis functions*, Technical Report 2000/30, University of Leicester, UK.

[281] S. J. Hales and J. Levesley, *Error estimates for multilevel approximation using polyharmonic splines*, Numer. Algorithms 30 (2002), 1–10.

[282] S. K. Halgamuge, W. Poechmueller, and M. Glesner, *An alternative approach for generation of membership functions and fuzzy rules based on radial and cubic basis function networks*, Internat. J. Approx. Reason. 12 (1995), 279–298.

[283] E. J. Halton and W. A. Light, *On local and controlled approximation order*, J. Approx. Theory, 72 (1993), 268–277.

[284] D. Handscomb, *Local recovery of a solenoidal vector field by an extension of the thin-plate spline technique*, Algorithms for approximation, III, Oxford, 1992, Numer. Algorithms 5, 1993, no. 1-4, 121–129.

[285] R. L. Harder and R. N. Desmarais, *Interpolation using surface splines*, J. Aircraft, 9 (1972), 189–191.

[286] R. L. Hardy, *Multiquadric equations of topography and other irregular surfaces*, J. Geophys. Res., 76 (1971), 1905–1915.

[287] R. L. Hardy, *Geodetic applications of multiquadric analysis*, AVN Allg. Vermess. Nachr., 79 (1972), 389–406.

[288] R. L. Hardy, *Research results in the application of multiquadric equations to surveying and mapping problems*, Survg. Mapp., 35 (1975), 321–332.

[289] R. L. Hardy, *Least squares prediction*, Photogramm. Eng. and Remote Sensing, 43 (1977), 475–492.

[290] R. L. Hardy, *Theory and applications of the multiquadric-biharmonic method*, Comput. Math. Appl., 19 (1990), 163–208.

[291] R. L. Hardy, *Comments on the discovery and evolution of the multiquadric method*, Comput. Math. Appl., 24 (1992), xi–xii.

[292] R. L. Hardy, *A contribution of the multiquadric method: interpolation of potential inside the earth*, Comput. Math. Appl., 24 (1992), 81–97.

[293] R. L. Hardy and W. M. Göpfert, *Least squares prediction of gravity anomalies, geodial undulations, and deflections of the vertical with multiquadric harmonic functions*, Geophys. Res. Letters, 2 (1975), 423–426.

[294] R. L. Hardy and S. A. Nelson, *A multiquadric-biharmonic representation and approximation of disturbing potential*, Geophys. Res. Letters, 13 (1986), 18–21.

[295] R. L. Hardy and S. Siranyone, *The multiquadric-biharmonic method of a three dimensional mapping inside ore deposits*, in Proceedings of Multinational Conference on Mine Planning and Design, University of Kentucky, 1989.

[296] E. J. Hartman, J. D. Keeler, and J. M. Kowalski, *Layered neural networks with Gaussian hidden units as universal approximations*, Neural Computations, 2 (1990), 210–215.

[297] S. Hartmann, *Multilevel-Fehlerabschätzung bei der Interpoaltion mit radialen Basisfunktionen*, Ph.D. dissertation, Universität Göttingen, 1998.

[298] M. Heiss and S. Kampl, *Multiplication-free radial basis function network*, IEEE Trans. Neural Networks, 7 (1996), 1461–1464.

[299] K. Höllig, *Finite Element Methods With B-Splines*, SIAM Frontiers in Applied Mathematics no. 26, Philadelphia, 2003.

[300] Y. C. Hon, K. F. Cheung, and X. Z. Mao, *A multiquadric model for tide and storm surge predictions*, City University of Hong Kong, 1996.

[301] Y. C. Hon and X. Z. Mao, *An efficient numerical scheme for Burgers' equation*, Appl. Math. Comput. (95) 1998, 37–50.

[302] Y. C. Hon and X. Z. Mao, *A multiquadric interpolation method for solving initial value problems*, City University of Hong Kong, 1996.

[303] Y. C. Hon and X. Z. Mao, *A radial basis function method for solving options pricing model*, Financial Engineering (8) 1999, 31–49.

[304] Y. C. Hon and R. Schaback, *On nonsymmetric collocation by radial basis functions*, Appl. Math. Comput. 119 (2001), 177–186.

[305] Y. C. Hon, R. Schaback, and X. Zhou, *An adaptive greedy algorithm for solving large RBF collocation problems*, Numer. Algorithms 32 (2003), 13–25.

[306] Y. C. Hon and Z. Wu, *Additive Schwarz domain decomposition with a radial basis approximation*, Int. J. Appl. Math. 4 (2000), 81–98.

[307] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press (Cambridge), 1985.

[308] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press (Cambridge), 1991.

[309] K. J. Hunt, R. Haas, and R. Murray Smith, *Extending the functional equivalence of radial basis function networks and fuzzy inference systems*, IEEE Trans. Neural Networks, 7 (1996), 776–781.

[310] M. F. Hutchinson, *The application of thin plate smoothing splines to continent-wide data assimilation*, in BMRC Research Report No.27, Data Assimilation Systems, J. D. Jasper (ed.), Bureau of Meteorology, Melbourne, 1991, 104–113.

[311] M. F. Hutchinson, *On thin plate splines and kriging*, in Computing and Science in Statistics 25, M. E. Tarter and M. D. Lock (eds.), Interface Foundation of North America, University of California, Berkeley, 1993, 55–62.

[312] M. F. Hutchinson, *Interpolating mean rainfall using thin plate smoothing splines*, Int. J. GIS, 9 (1995), 305–403.

[313] M. S. Ingber, C. S. Chen, and J. A. Tanski, *A mesh free approach using radial basis functions and parallel domain decomposition for solving three dimensional diffusion equations*, University of Nevada, preprint 2003.

[314] A. Iske, *Charakterisierung bedingt positiv definiter Funktionen für multivariate Interpolationsmethoden mit radial Basisfunktionen*, Ph.D. dissertation, Universität Göttingen, 1994.

[315] A. Iske, *Reconstruction of functions from generalized Hermite-Birkhoff data*, in Approximation Theory VIII, Vol. 1: Approximation and Interpolation, C. Chui, and L. Schumaker (eds.), World Scientific Publishing, Singapore, 1995, 257–264.

[316] A. Iske, *Perfect centre placement for radial basis function methods*, preprint, Technische Universität München.

[317] A. Iske, *Reconstruction of smooth signals from irregular samples by using radial basis function approximation*, in Proceeding of the 1999 International Workshop on Sampling Theory and Applications, Y. Lyubaskii (ed.), The Norwegian University of Science and Technology, Trondheim, 1999, 82–87.

[318] A. Iske, *Hierarchical scattered data filtering for multilevel interpolation schemes*, in Mathematical Methods for Curves and Surfaces: Oslo 2000, T. Lyche and L. L. Schumaker (eds.), Vanderbilt University Press (2001), 211–220.

[319] A. Iske, *Multiresolution Methods in Scattered Data Modelling*, Habilitation Thesis, Universität München, 2002.

[320] A. Iske, *On the approximation order and numerical stability of local Lagrange interpolation by polyharmonic splines*, in Modern Developments in Multivariate Approximation, W. Haussmann, K. Jetter, M. Reimer, and J. Stöckler (eds.), ISNM 145, Birkhäuser Verlag (2003), 153–165.

[321] A. Iske and T. Sonar, *On the structure of function spaces in optimal recovery of point functionals for ENO-schemes by radial basis functions*, Numer. Math., 74 (1996), 177–201.

[322] T. Ivanov, V. Maz'ya, and G. Schmidt, *Boundary layer approximate approximations and cubature of potentials in domains*, Adv. in Comput. Math. (10) 1999, 311–342.

[323] I. R. H. Jackson, *Approximation to continuous functions using radial basis functions*, in The Mathematics of Surfaces II, R. R. Martin (ed.), Clarendon Press, Oxford, 1987, 115–135.

[324] I. R. H. Jackson, *Radial basis function methods from multivariate approximation*, University of Cambridge, 1988.

[325] I. R. H. Jackson, *Convergence properties of radial basis functions*, Constr. Approx., 4 (1988), 243–264.

[326] I. R. H. Jackson, *An order of convergence for some radial basis functions*, IMA J. Numer. Anal., 9 (1989), 567–587.

[327] I. R. H. Jackson, *Radial basis functions – a survey and new results*, in Mathematics of Surfaces III, D. C. Handscomb (ed.), Clarendon Press, Oxford, 1989, 115–133.

[328] C. Jekeli, *Hardy's multiquadric-biharmonic method for gravity field predictions*, Comput. Math. Appl., 28 (1994), 43–46.

[329] K. Jetter, *Multivariate approximation from the cardinal interpolation point of view*, in Approximation Theory VII, E. W. Cheney, C. Chui, and L. Schumaker (eds.), Academic Press, New York, 1992, 131–161.

[330] K. Jetter, *Riesz bounds in scattered data interpolation and L_2-approximation*, in Multivariate Approximation: From CAGD to Wavelets, Kurt Jetter and Florencio Utreras (eds.), World Scientific Publishing, Singapore, 1993, 167–177.

[331] K. Jetter, *Conditionally lower Riesz bounds in scattered data interpolation*, in Wavelets, Images, and Surface Fitting, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), A. K. Peters, Wellesley MA, 1994, 295–302.

[332] K. Jetter and J. Stöckler, *Algorithms for cardinal interpolation using box splines and radial basis functions*, Numer. Math., 60 (1991), 97–114.

[333] K. Jetter and J. Stöckler, *Topics in scattered data interpolation and non-uniform sampling*, in Surface Fitting and Multiresolution, A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press (1997), 191–208.

[334] R. Q. Jia and J. J. Lei, *Approximation by multiinteger translates of functions having global support*, J. Approx. Theory, 72 (1991), 2–23.

[335] R. Q. Jia and J. J. Lei, *A new version of the Strang-Fix conditions*, J. Approx. Theory, 74 (1993), 221-225.

[336] B. Jumarhon, S. Amini and K. Chen, *Hermite type collocation method using radial basis functions*, preprint, 1998.

[337] S. Kaczmarz, *Approximate solution of systems of linear equations* (originally published as: Angenäherte Auflösung von Systemen linearer Gleichungen, Bulletin International de l'Academie Polonaise des Sciences, Lett. A, 1937, 355–357), Int. J. Control 57 (1993) 1269–1271.

[338] E. J. Kansa, *Application of Hardy's multiquadric interpolation to hydrodynamics*, Proc. 1986 Simul. Conf., Vol. 4, (1986), 111–117.

[339] E. J. Kansa, *Multiquadrics – A scattered data approximation scheme with applications to computational fluid-dynamics – I: Surface approximations and partial derivative estimates*, Comput. Math. Appl., 19 (1990), 127–145.

[340] E. J. Kansa, *Multiquadrics – A scattered data approximation scheme with applications to computational fluid-dynamics – II: Solutions to parabolic, hyperbolic and elliptic partial differential equations*, Comput. Math. Appl., 19 (1990), 147–161.

[341] E. J. Kansa, *A strictly conservative spatial approximation scheme for the governing engineering and physics equations over irregular regions and inhomogeneous scattered nodes*, Comput. Math. Appl., 24 (1992), 169–190.

[342] E. J. Kansa and R. E. Carlson, *Improved accuracy of multiquadric interpolation using variable shape parameters*, Comput. Math. Appl., 24 (1992), 99–120.

[343] E. J. Kansa and R. E. Carlson, *Radial basis functions: a class of grid free scattered data approximations*, Computational Fluid Dynamics J., 3 (1995), 479–496.

[344] E. J. Kansa, G. E. Fasshauer, H. Power, and L. Ling, *A volumetric integral radial basis function method for time-dependent partial differential equations: I. Formulation*, to appear in Journal of Engineering Analysis with Boundary Elements.

[345] E. J. Kansa and Y. C. Hon, *Circumventing the ill-conditioning problem with multiquadric radial basis functions: Applications to elliptic partial differential equations*, Comput. Math. Applic. 39 (2000), 123–137.

[346] S. Karlin, *Total Positivity*, Stanford University Press, Stanford, 1968.

[347] S. R. Karur and P. A. Ramachandran, *Radial basis function approximation in the dual reciprocity method*, Math. Comput. Modell., 20 (1994), 59–70.

[348] S. R. Karur and P. A. Ramachandran, *Augmented thin plate spline approximation in DRM*, Boundary Elements Comm., 6 (1995), 55–58.

[349] J. T. Kent and K. V. Mardia, *The link between kriging and thin-plate splines, Probability, statistics and optimisation*, Wiley Ser. Probab. Math. Statist., Wiley, Chichester, 1994, 325–339.

[350] D. Kincaid and W. Cheney, *Numerical Analysis: Mathematics of Scientific Computing* (2nd ed.), Brooks/Cole, Pacific Grove, CA, 1996.

[351] O. Kounchev, *Multivariate Polysplines: Applications to Numerical and Wavelet Analysis*, Academic Press, New York, 2001.

[352] A. Krzyzak, T. Linder, and G. Lugosi, *Nonparametric estimation and classification using radial basis function nets and empirical risk minimization*, IEEE Trans. Neural Networks, 7 (1996), 475–487.

[353] S. Kunis and D. Potts, *NFFT, Softwarepackage (C-library)*, Universität Lübeck, `http://www.math.uni-luebeck.de/potts/nfft/`, 2002.

[354] S. Kunis, D. Potts, and G. Steidl, *Fast Fourier transforms at nonequispaced knots: A user's guide to a C-library*, Universität Lübeck, `http://www.math.uni-luebeck.de/potts/nfft/`, 2002.

[355] A. J. Kurdila, F. J. Narcowich, and J. D. Ward, *Persistency of excitation in identification using radial basis function approximants*, SIAM J. Control Optim., 33 (1995), 625–642.

[356] J. Kybic, T. Blu, and M. Unser, *Generalized sampling: A variational approach – Part I: Theory*, IEEE Trans. Signal Proc., 50 (2002), 1965–1976.

[357] J. Kybic, T. Blu, and M. Unser, *Generalized sampling: A variational approach – Part II: Applications*, IEEE Trans. Signal Proc., 50 (2002), 1977–1985.

[358] P. Lancaster and K. Šalkauskas, *Surfaces generated by moving least squares methods*, Math. Comp. (37) 1981, 141–158.

[359] P. Lancaster and K. Šalkauskas, *Curve and Surface Fitting*, Academic Press, New York, 1986.

[360] E. Larsson and B. Fornberg, *A numerical study of some radial basis function based solution methods for elliptic PDEs*, Comput. Math. Appl., to appear.

[361] E. Larsson and B. Fornberg, *Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions*, preprint, University of Colorado.

[362] Y. Lee, *Handwritten digit recognition using K nearest-neighbor, radial-basis function, and backpropagation neural networks*, Neural Computation, 3 (1991), 440–449.

[363] S. Lee and J. C. J. Pan, *Unconstrained handwritten numeral recognition based on radial basis competitive and cooperative networks with spatiotemporal feature representation*, IEEE Trans. Neural Networks, 7 (1996), 455–474.

[364] D. Lee and J. H. Shiau, *Thin plate splines with discontinuities and fast algorithms for their computation*, SIAM J. Scient. Computing, 15 (1994), 1311-1330.

[365] V. M. A. Leitão, *A meshless method for Kirchhoff plate bending problems*, Int. J. Numer. Meth. Engng. 52 (2001), 1107–1130.

[366] A. Le Méhauté, *Knot removal for scattered data*, in Approximation Theory, Wavelets and Applications, S. P. Singh (ed.), Kluwer, Dordrecht, Netherlands, 1995, 197–213.

[367] A. Le Méhauté, *Inf-Convolution and radial basis functions*, in New Developments in Approximation Theory, M. W. Müller, M. D. Buhmann, D. H. Mache and M. Felten (eds.), Birkhäuser, Basel, 1999, 95–108.

[368] J. A. Leonard, M. A. Kramer, abd J. H. Ungar, *Using radial basis functions to approximate a function and its error bounds*, IEEE Trans. on Neural Networks, 3 (1992), 624–627.

[369] J. Levesley, *Convolution kernels for approximation by radial basis functions*, in Wavelets, Images, and Surface Fitting, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), A. K. Peters, Wellesley MA, 1994, 343–350.

[370] J. Levesley, *Pointwise estimates for multivariate interpolation using conditionally positive definite functions*, in Approximation Theory, Wavelets and Applications, S. P. Singh (ed.), Kluwer, Dordrecht, Netherlands, 1995, 381–401.

[371] J. Levesley, *Convolution kernels based on thin plate splines*, Numer. Algorithms, 10 (1995), 401–419.

[372] J. Levesley and A. K. Kushpel, *Interpolation on compact Abelian groups using generalised sk-splies*, in Approximation Theory VIII, Vol. 1: Approximation and Interpolation, C. Chui, and L. Schumaker (eds.), World Scientific Publishing, Singapore, 1995, 317–324.

[373] J. Levesley and A. K. Kushpel, *Generalised sk-spline interpolation on compact Abelian groups*, University of Leicester, 1995.

[374] J. Levesley, W. Light, D. Ragozin and X. Sun, *A simple approach to the variational theory for interpolation on spheres*, in New Developments in Approximation Theory, M. W. Müller, M. D. Buhmann, D. H. Mache and M. Felten (eds.), Birkhäuser, Basel, 1999, 117-144.

[375] J. Levesley and M. Roach, *Quasi-interpolation on compact domains*, in Approximation Theory, Wavelets and Applications, S. P. Singh (ed.), Kluwer, Dordrecht, Netherlands, 1995, 557–566.

[376] J. Levesley and X. Sun, *Scattered Hermite interpolation by ridge functions*, Numer. Func. Anal. Optim., 16 (1995), 989–1001.

[377] J. Levesley, Y. Xu, W. Light, and W. Cheney, *Convolution operators for radial basis approximation*, SIAM J. Math. Anal., 27 (1996), 286–304.

[378] D. Levin, *The approximation power of moving least-squares*, Math. Comp. 67 (1998), 1517–1531.

[379] J. Li and Y. C. Hon, *Domain decomposition for radial basis meshless methods*, preprint, University of Nevada.

[380] S. Li and W. K. Liu, *Moving least-square reproducing kernel method Part II: Fourier analysis*, Comp. Meth. Appl. Mech. Eng. (139) 1996, 159–193.

[381] S. Li and W. K. Liu, *Meshfree and particle methods and their applications*, Applied Mechanics Review (55) 2002, 1–34.

[382] X. Li, *On simultaneous approximations by radial basis function neural networks*, Appl. Math. Comput. 95 (1998), 75–89.

[383] X. Li, C. H. Ho, and C. S. Chen, *Computational test of approximation of functions and their derivatives by radial basis functions*, Neural Parallel Sci. Comput. 10 (2002), 25–46.

[384] X. Li and C. A. Micchelli, *Approximation by radial bases and neural networks*, Numer. Algorithms 25 (2000), 241–262.

[385] K. Lindsay and R. Krasny, *A particle method and adaptive treecode for vortex sheet motion in three-dimensional flow*, J. Comput. Physics 172 (2001), 879–907.

[386] W. A. Light, *Recent developments in the Strang-Fix theory for approximation orders*, in Curves and Surfaces, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), Academic Press, New York, 1991, 285–292.

[387] W. A. Light, *Some aspects of radial basis function approximation*, in Approximation Theory, Spline Functions and Applications, S. P. Singh (ed.), Kluwer, Dordrecht, Netherlands, 1992, 163–190.

[388] W. A. Light, *Using radial functions on compact domains*, in Wavelets, Images, and Surface Fitting, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), A. K. Peters, Wellesley MA, 1994, 351–370.

[389] W. A. Light, *Variational error bounds for radial basis functions*, in Numerical Analysis 1995, D. F. Griffiths and G. A. Watson (eds.), Longman (Harlow), 1996, 94–106.

[390] W. A. Light and E. W. Cheney, *Interpolation by piecewise-linear radial basis functions, II*, J. Approx. Theory, 64 (1991), 38–54.

[391] W. A. Light and E. W. Cheney, *Interpolation by periodic radial basis functions*, J. Math. Anal. Appl., 168 (1992), 111–130.

[392] W. A. Light and E. W. Cheney, *Quasi-interpolation with translates of a function having noncompact support*, Constr. Approx., 8 (1992), 35–48.

[393] W. A. Light and H. Wayne, *Error estimates for approximation by radial basis functions*, in Approximation Theory, Wavelets and Applications, S. P. Singh (ed.), Kluwer, Dordrecht, Netherlands, 1995, 215–246.

[394] W. A. Light and H. Wayne, *On power functions and error estimates for radial basis function interpolation*, J. Approx. Theory 92 (1998), 245–266.

[395] L. Ling and E. J. Kansa, *A least-squares preconditioner for radial basis functions collocation methods*, preprint, Simon Fraser University, 2002.

[396] L. Ling and E. J. Kansa, *Preconditioning for radial basis functions with domain decomposition methods*, preprint, Simon Fraser University, 2003.

[397] G. R. Liu, *Mesh Free Methods: Moving beyond the Finite Element Method*, CRC Press, Boca Raton, FL, 2002.

[398] W. K. Liu, Y. Chen, R. Aziz Uras, and C. T. Chang, *Generalized multiple scale reproducing kernel particle methods*, Comp. Meth. Appl. Mech. Eng. (139) 1996, 91–157.

[399] W. K. Liu, S. Li, and T. Belytschko, *Moving least square reproducing kernel methods (I) Methodology and convergence*, Comp. Meth. Appl. Mech. Eng. (143) 1997, 113–154.

[400] F. Locher, *Interpolation on uniform meshes by the translates of one function and related attenuation factors*, Math. Comp., 37 (1981), 403–416.

[401] R. Lorentz, F. J. Narcowich, and J. D. Ward, *Collocation discretization of the transport equation with radial basis functions*, preprint, Texas A & M University.

[402] D. Louse and R. Matthews, *Shakespeare vs. Fletcher: a stylistic analysis by radial basis functions*, Computers and Humanities, 29 (1995), 449–xxx.

[403] D. Lowe, *Novel 'topographic' nonlinear feature extraction using radial basis functions for concentration coding in the 'artificial nose'*, in 3rd IEE International Conference on Artificial Neural Networks, IEE, London, 1993.

[404] D. Lowe and A. McLachlan, *Modelling of nonstationary processes using radial basis function networks*, Neural Computing Research Group, Aston University, UK, 199x.

[405] D. Lowe and M. E. Tipping, *A novel neural network technique for exploratory data analysis*, in Proceedings of ICANN '95, Vol. 1, EC2 & Cie, Paris, 1995, 339–344.

[406] D. Lowe and M. E. Tipping, *Feed-forward neural networks and topographic mappings for exploratory data analysis*, Neural Computing Research Group, Aston University, UK, 1996.

[407] S. Lowitzsch, *Approximation and Interpolation Employing Divergence-Free Radial Basis Functions with Applications*, Ph.D. Dissertation, Texas A&M University, 2002.

[408] S. Lowitzsch, *Matrix-valued radial basis functions: Stability estimates and applications*, preprint, 2003.

[409] E. Lukacs, *Characteristic Functions*, Griffin, London, 1970.

[410] Z. Luo, *The invertibility of interpolation functions and the fundamentality of their translates on spheres*, preprint, 1994.

[411] Z. Luo, *Characterization and eigenvalue estimation of conditionally positive definite functions on spheres*, preprint, 1994.

[412] T. Lyche, *Knot Removal for Spline Curves and Surfaces*, in Approximation Theory VII, E. W. Cheney, C. Chui, and L. Schumaker (eds.), Academic Press, New York, 1992, 207–226.

[413] W. R. Madych, *Cardinal interpolation with polyharmonic splines*, in Multivariate Approximation Theory IV, ISNM 90, C. Chui, W. Schempp, and K. Zeller (eds.), Birkhäuser Verlag, Basel, 1989, 241–248.

[414] W. R. Madych, *Polyharmonic splines, multiscale analysis, and entire functions*, in Multivariate Approximation and Interpolation, ISNM 94, W. Haussman and K. Jetter (eds.), Birkhäuser, Basel, 1990, 205–216.

[415] W. R. Madych, *Error estimates for interpolation by generalized splines*, in Curves and Surfaces, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), Academic Press, New York, 1991, 297–306.

[416] W. R. Madych, *Miscellaneous error bounds for multiquadric and related interpolators*, Comput. Math. Appl., 24 (1992), 121–138.

[417] W. R. Madych and S. A. Nelson, *Multivariate interpolation: a variational theory*, manuscript, 1983.

[418] W. R. Madych and S. A. Nelson, *Multivariate interpolation and conditionally positive definite functions*, Approx. Theory Appl., 4 (1988), 77–89.

[419] W. R. Madych and S. A. Nelson, *Error bounds for multiquadric interpolation*, in Approximation Theory VI, C. Chui, L. Schumaker, and J. Ward (eds.), Academic Press, New York, 1989, 413–416.

[420] W. R. Madych and S. A. Nelson, *Multivariate interpolation and conditionally positive definite functions, II*, Math. Comp., 54 (1990), 211–230.

[421] W. R. Madych and S. A. Nelson, *Polyharmonic cardinal splines*, J. Approx. Theory, 60 (1990), 141–156.

[422] W. R. Madych and S. A. Nelson, *Polyharmonic cardinal splines: A minimization property*, J. Approx. Theory, 63 (1990), 303–320.

[423] W. R. Madych and S. A. Nelson, *Bounds on multivariate polynomials and exponential error estimates for multiquadric interpolation*, J. Approx. Theory, 70 (1992), 94–114.

[424] N. Mai-Duy and T. Tran-Cong, *Numerical solution of differential equations using multiquadric radial basis function networks*, Neural Networks 14 (2001), 185–199.

[425] J. C. Mairhuber, *On Haar's theorem concerning Chebyshev approximation problems having unique solutions*, Proc. Am. Math. Soc. (7) 1956, 609–615.

[426] A. Makroglu, *Radial basis functions in the numerical solution of Fredholm and integro-differential equations*, in Advances in Computer Methods for Partial Differential Equations – VII, R. Vichnevetsky, D. Knight, and G. Richter (eds.), IMACS, New Brunswick, 1992, 478–484.

[427] A. Makroglu, *Radial basis functions in the numerical solution of nonlinear Volterra integral equations*, J. Appl. Sci. Comput., 1 (1994), 33–53.

[428] A. Makroglu, *Multiquadric collocation methods in the numerical solution of Volterra integral and integro-differential equations*, in Mathematics of Computation 1943–1993, Proc. Sympos. Appl. Math., 48, Amer. Math. Soc., Providence, RI, 1994, 337–341.

[429] A. Makroglu and E. J. Kansa, *Multiquadric collocation methods in the numerical solution of Volterra integral equations with weakly singular kernels*, Lawrence Livermore National Laboratory, 1993.

[430] M. D. Marcozzi, S. Choi and C. S. Chen, *RBF and optimal stopping problems: an application to the pricing of vanilla options on one risky asset*, in Boundary Element Technology XIII, C. S. Chen, C. A. Brebbia, and D. W. Pepper (eds.), WIT Press, 1999, 345–354.

[431] M. Marcus and H. Minc, *Introduction to Linear Algebra*, Dover Publications, New York, 1965 (republished 1988).

[432] M. Mathias, *Über positive Fourier-Integrale*, Math. Zeit., 16 (1923), 103–125.

[433] V. Maz'ya and G. Schmidt, *On approximate approximations using Gaussian kernels*, IMA J. Numer. Anal. (16) 1996, 13–29.

[434] V. Maz'ya and G. Schmidt, *On quasi-interpolation with non-uniformly distributed centers on domains and manifolds*, J. Approx. Theory (110) 2001, 125–145.

[435] S. F. McCormick, *Multilevel Projection Methods for Partial Differential Equations*, CBMS-NSF Regional Conference Series in Applied Mathematics 62, SIAM (Philadelphia), 1992.

[436] D. H. McLain, *Drawing contours from arbitrary data points*, Comput. J. (17) 1974, 318–324.

[437] J. McMahon, *Knot selection for least squares approximation using thin plate splines*, M.S. Thesis, Naval Postgraduate School, 1986.

[438] J. McMahon and R. Franke, *Knot selection for least squares thin plate splines*, SIAM J. Sci. Statist. Comput., 13 (1992), 484–498.

[439] J. Meinguet, *Multivariate interpolation at arbitrary points made simple*, Z. Angew. Math. Phys., 30 (1979), 292–304.

[440] J. Meinguet, *An intrinsic approach to multivariate spline interpolation at arbitrary points*, in Polynomial and Spline Approximations, N. B. Sahney (ed.), Reidel, Dordrecht, 1979, 163–190.

[441] J. Meinguet, *Basic mathematical aspects of surface spline interpolation*, in Numerische Integration, G. Hämmerlin (ed.), Birkhäuser, Basel, 1979, 211–220.

[442] J. Meinguet, *Surface spline interpolation: basic theory and computational aspects*, in Approximation Theory and Spline Functions, S. P. Singh, J. H. W. Burry, and B. Watson (eds.), Reidel, Dordrecht, 1984, 127–142.

[443] J. M. Melenk and I. Babuška, *The partition of unity finite element method: basic theory and applications*, ETH Zürich, 1996.

[444] V. A. Menegatto, *Interpolation on spherical domains*, Analysis, 14 (1994), 415–424.

[445] V. A. Menegatto, *Strictly positive definite kernels on the Hilbert sphere*, Appl. Anal., 55 (1994), 91–101.

[446] V. A. Menegatto, *Interpolation on the complex Hilbert sphere*, Approx. Theory Appl., 11 (1995), 1–9.

[447] V. A. Menegatto, *Strictly conditionally negative definite functions on the sphere*, University of Texas Austin, 1992.

[448] V. A. Menegatto, *Interpolation on Spherical Spaces*, University of Texas Austin, 1992.

[449] V. A. Menegatto, *Strictly conditionally positive definite functions on the unit Hilbert sphere*, University of Texas Austin, 1992.

[450] V. A. Menegatto, *Strictly positive definite kernels on the circle*, Rocky Mountain J. Math., 25 (1995), 1149–1163.

[451] V. A. Menegatto, *Interpolation on the complex Hilbert sphere using positive definite and conditionally negative definite functions*, manuscript, 1995.

[452] V. A. Menegatto and A. P. Peron, *Conditionally positive definite kernels on Euclidean domains*, preprint, 2003.

[453] K. Menger, *Die Metrik des Hilbertschen Raumes*, Anzeiger der Akad. der Wissenschaften in Wien, Nat. Kl. 65 (1928), 159–160.

[454] H. N. Mhaskar, *Weighted polynomials, radial basis functions and potentials on locally compact spaces*, Numer. Func. Anal. Optim., 11 (1990), 987–1017.

[455] H. N. Mhaskar and C. A. Micchelli, *Approximation by superposition of sigmoidal and radial basis functions*, Adv. in Appl. Math., 13 (1992), 350–373.

[456] C. A. Micchelli, *Interpolation of scattered data: distance matrices and conditionally positive definite functions*, Constr. Approx., 2 (1986), 11–22.

[457] C. A. Micchelli and T. J. Rivlin, *A survey of optimal recovery*, in Optimal estimation in approximation theory (Proc. Internat. Sympos., Freudenstadt, 1976), C. A. Micchelli and T. J. Rivlin (eds.), Plenum Press, New York, 1977, 1–54.

[458] C. A. Micchelli and T. J. Rivlin, *Optimal recovery of best approximations*, Resultate Math. 3 (1980), no. 1, 25–32.

[459] C. A. Micchelli and T. J. Rivlin, *Lectures on optimal recovery*, in Numerical analysis, Lancaster 1984, P. R. Turner (ed.), Lecture Notes in Math., 1129, Springer, Berlin, 1985, 21–93.

[460] C. A. Micchelli, T. J. Rivlin, and S. Winograd, *The optimal recovery of smooth functions*, Numer. Math. 26 (1976), no. 2, 191–200.

[461] M. J. Milroy, G. W. Vickers, and C. Bradley, *An adaptive radial basis function approach to modeling scattered data*, xxx, to appear.

[462] J. Miranda, *Incorporating R-functions to the Theory of Positive Definite Functions to Solve Partial Differential Equations*, Ph.D. Dissertation, Illinois Institute of Technology, 2004.

[463] J. J. Monaghan, *An introduction to SPH*, Comp. Phys. Comm. 48 (1988), 89–96.

[464] P. Montès, *Local kriging interpolation: Application to scattered data on the sphere*, in Curves and Surfaces, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), Academic Press, New York, 1991, 325–329.

[465] P. Montès, *Smoothing noisy data by kriging with nugget effects*, in Wavelets, Images, and Surface Fitting, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), A. K. Peters, Wellesley MA, 1994, 371–378.

[466] G. J. Moridis and E. J. Kansa, *The method of Laplace transform multiquadrics for the solution of the groundwater flow equation*, in Advances in Computer Methods for Partial Differential Equations – VII, R. Vichnevetsky, D. Knight, and G. Richter (eds.), IMACS, New Brunswick, 1992, xxx–xxx.

[467] G. J. Moridis and E. J. Kansa, *The Laplace transform multiquadric method: A highly accurate scheme for the numerical solution of linear partial differential equations*, J. Appl. Sc. Comp., 1 (1994), 375–407.

[468] C. T. Mouat, *Fast algorithms and preconditioning techniques for fitting radial basis functions*, Ph.D. Dissertation, University of Canterbury, NZ, 2001.

[469] C. Müller, *Spherical Harmonics*, Springer Lecture Notes in Mathematics (Vol. 17), 1966.

[470] B. Mulgrew, *Applying radial basis functions*, IEEE Signal Proc. Magazine, 13 (1996), 50–65.

[471] M. T. Musavi, W. Ahmed, K. H. Chan, K. B. Faris, and D. M. Hummels, *On the training of radial basis function classifiers*, Neural Networks, 5 (1992), 595–603.

[472] R. Muttiah, M. Hutchinson, and P. Dyke, *Climate surfaces of Texas using thin plate splines*, in Approximation Theory VIII, Vol. 1: Approximation and Interpolation, C. Chui, and L. Schumaker (eds.), World Scientific Publishing, Singapore, 1995, 427–434.

[473] D. E. Myers, *Interpolation with positive definite functions*, Sciences de la Terre, 28 (1988), 252–265.

[474] D. E. Myers, *Kriging, co-Kriging, radial basis functions and the role of positive definiteness*, Comput. Math. Appl., 24 (1992), 139–148.

[475] D. E. Myers, *Selection of a radial basis function for interpolation*, in Advances in Computer Methods for Partial Differential Equations – VII, R. Vichnevetsky, D. Knight, and G. Richter (eds.), IMACS, New Brunswick, 1993, xxx–xxx.

[476] E. A. Nadaraya, *On estimating regression*, Theor. Probab. Appl. 9 (1964), 141–142.

[477] F. J. Narcowich, *Generalized Hermite interpolation and positive definite kernels on a Riemannian manifold*, J. Math. Anal. Appl., 190 (1995), 165–193.

[478] F. J. Narcowich, R. Schaback and J. D. Ward, *Multilevel interpolation and approximation*, Appl. Comput. Harmon. Anal. 7 (1999), 243–261.

[479] F. J. Narcowich, N. Sivakumar, and J. D. Ward, *On condition numbers associated with radial-function interpolation*, J. Math. Anal. Appl., 186 (1994), 457–485.

[480] F. J. Narcowich, P. Smith, and J. D. Ward, *Density of translates of radial functions on compact sets*, in Approximation Theory VIII, Vol. 1: Approximation and Interpolation, C. Chui, and L. Schumaker (eds.), World Scientific Publishing, Singapore, 1995, 435–442.

[481] F. J. Narcowich and J. D. Ward, *Norms of inverses and condition numbers for matrices associated with scattered data*, J. Approx. Theory, 64 (1991), 69–94.

[482] F. J. Narcowich and J. D. Ward, *Norms of inverses for matrices associated with scattered data*, in Curves and Surfaces, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), Academic Press, New York, 1991, 341–348.

[483] F. J. Narcowich and J. D. Ward, *Norm estimates for the inverses of a general class of scattered-data radial-function interpolation matrices*, J. Approx. Theory, 69 (1992), 84–109.

[484] F. J. Narcowich and J. D. Ward, *Generalized Hermite interpolation via matrix-valued conditionally positive definite functions*, Math. Comp., 63 (1994), 661–687.

[485] F. J. Narcowich and J. D. Ward, *Wavelets associated with periodic basis functions*, CAT #340, Texas A&M University, 1994.

[486] F. J. Narcowich and J. D. Ward, *Nonstationary wavelets on the m-sphere for scattered data*, Texas A&M University, 1995.

[487] F. J. Narcowich and J. D. Ward, *Nonstationary spherical wavelets for scattered data*, in Approximation Theory VIII, Vol. 2: Wavelets and Multilevel Approximation, C. Chui, and L. Schumaker (eds.), World Scientific Publishing, Singapore, 1995, 301–308.

[488] G. M. Nielson, *Coordinate free scattered data interpolation*, in Topics in Multivariate Approximation, C. K. Chui, L. L. Schumaker, and F. Utreras (eds.), Academic Press, New York, 1987, 175–184.

[489] G. M. Nielson and T. A. Foley, *A survey of applications of an affine invariant norm*, in Mathematical Methods in Computer Aided Geometric Design, T. Lyche and L. Schumaker (eds.), Academic Press, New York, 1989, 445–467.

[490] A. Nieslony, D. Potts, and G. Steidl, *Rapid evaluation of radial functions by fast Fourier transforms at nonequispaced knots*, Universität Mannheim, preprint 2003.

[491] W. A. Nuss, *Three dimensional meteorological analysis using multiquadric interpolation*, Int. J. Sci. Comp. Model., to appear.

[492] W. A. Nuss and D. W. Titley, *Use of multiquadric interpolation for meteorological objective analysis*, Monthly Weather Review, 12 (1994), 1011–1031.

[493] Y. Ohtake, A. Belyaev, and H. P. Seidel, *A multi-scale approach to 3D scattered data interpolation with compactly supported basis functions*, preprint, Max-Planck-Institut für Informatik, Saarbrücken, 2003.

[494] Y. Ohtake, A. Belyaev, M. Alexa, G. Turk, and H. P. Seidel, *Multi-level partition of unity implicits*, preprint, Max-Planck-Institut für Informatik, Saarbrücken, 2003.

[495] M. J. L. Orr, *Regularization in the selection of radial basis function centres*, Neural Computation, 7 (1995), 606–623.

[496] M. J. L. Orr, *Introduction to radial basis function networks*, Centre for Cognitive Sciences, University of Edinburgh, 1996.

[497] M. Orr, J. Hallam, K. Takezawa, A. Murray, S. Ninomiya, M. Oide, and T. Leonard, *Combining regression trees and radial basis function networks*, Int. J. Neural Systems 10 (2000), 453–465.

[498] U. Pahner and K. Hameyer, *Adaptive coupling of differential evolution and multiquadrics approximation for the tuning of the optimization process*, Katholieke Universiteit Leuven, 1999.

[499] J. Park and I. W. Sandberg, *Universal approximation using radial bais function networks*, Neural Computation, 3 (1991), 246–257.

[500] J. Park and I. W. Sandberg, *Approximation and radial basis function networks*, Neural Computation, 5 (1993), 305–316.

[501] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist. 33 (1962), 1065–1076.

[502] C. Pegram and D. S. Pegram, *Integration of rainfall via multiquadric surfaces over polygons*, J. Hydraulic Eng., 114 (1993), 151–163.

[503] R. B. Platte and T. A. Driscoll, *Computing eigenmodes of elliptic operators using radial basis functions*, preprint, University of Delaware, 2003.

[504] T. Poggio and F. Girosi, *Networks for approximation and learning*, Proc. IEEE, 78 (1990), 1481–1497.

[505] E. H. Potter, *Multivariate polyharmonic spline interpolation*, Iowa State University, 1981.

[506] H. Pottmann and M. Eck, *Modified multiquadric methods for scattered data interpolation over a sphere*, Comput. Aided Geom. Design, 7 (1990), 313–321.

[507] M. Pottmann and H. P. Jörgl, *Radial basis function networks for internal model control*, Appl. Math. Comput., 70 (1995), 283–298.

[508] D. Potts and G. Steidl, *Fast summation at nonequispaced knots by NFFTs*, SIAM J. Sci. Comput. 24 (2003), 2013 - 2037.

[509] M. J. D. Powell, *Radial basis functions for multivariable interpolation: a review*, in Algorithms for the Approximation of Functions and Data, J. C. Mason and M. G. Cox (eds.), Oxford Univ. Press, Oxford, 1987, 143–167.

[510] M. J. D. Powell, *Radial basis function approximations to polynomials*, in Numerical Analysis 1987, D. F. Griffiths and G. A. Watson (eds.), Longman Scientific and Technical (essex, 1987, 223–241.

[511] M. J. D. Powell, *Univariate multiquadric approximation: reproduction of linear polynomials*, in Multivariate Approximation and Interpolation, ISNM 94, W. Haussman and K. Jetter (eds.), Birkhäuser, Basel, 1990, 227–240.

[512] M. J. D. Powell, *Univariate multivariate interpolation: some recent results*, in Curves and Surfaces, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), Academic Press, New York, 1991, 371–382.

[513] M. J. D. Powell, *The theory of radial basis functions in 1990*, in Advances in Numerical Analysis II: Wavelets, Subdivision, and Radial Basis Functions, W. Light (ed.), Oxford University Press, Oxford, 1992, 105–210.

[514] M. J. D. Powell, *Tabulation of thin-plate splines on a very fine two-dimensional grid*, in Numerical Methods in Approximation Theory, ISNM 105, D. Braess, L. L. Schumaker (ed.), Birkhäuser, Basel, 1992, 221–244.

[515] M. J. D. Powell, *Truncated Laurent expansions for the fast evaluation of thin plate splines*, in Algorithms for Approximation III, J. C. Mason and M. G. Cox (eds.), Chapman and Hall, London, 1992, 99–120.

[516] M. J. D. Powell, *Some algorithms for thin plate spline interpolation to functions of two variables*, in Advances in Computational Mathematics (New Delhi, 1993), H. P. Dikshit and C. A. Micchelli (eds.), Ser. Approx. Decompos., 4 (1994), World Sci. Publishing, 303–319.

[517] M. J. D. Powell, *The uniform convergence of thin plate spline interpolation in two dimensions*, Numer. Math., 68 (1994), 107–128.

[518] M. J. D. Powell, *A new iterative algorithm for thin plate spline interpolation in two dimensions*, Annals of Numerical Mathematics, 4 (1997), 519–528.

[519] M. J. D. Powell, *Recent research at Cambridge on radial basis functions*, in New Developments in Approximation Theory, M. W. Müller, M. D. Buhmann, D. H. Mache and M. Felten (eds.), Birkhäuser, Basel, 1999, 215–232.

[520] H. Power and V. Barraco, *A comparison analysis between unsymmetric and symmetric radial basis function collocation methods for the numerical solution of partial differential equations*, Comput. Math. Appl. 43 (2002), 551–583.

[521] E. Quak, N. Sivakumar, and J. D. Ward, *Least squares approximation by radial functions*, SIAM J. Numer. Anal., 24 (1993), 1043-1066.

[522] C. Rabut, *Fast quasi-interpolation of surfaces with generalized B-splines on regular nets*, in Mathematics of Surfaces III, D. C. Handscomb (ed.), Clarendon Press, Oxford, 1989, 429–449.

[523] C. Rabut, *B-splines polyharmoniques cardinales: Interpolation, quasi-interpolation, filtrage*, University of Toulouse, 1990.

[524] C. Rabut, *An introduction to Schoenberg's approximation*, Comput. Math. Appl., 24 (1992), 149–175.

[525] C. Rabut, *Elementary m-harmonic cardinal B-splines*, Numer. Algorithms, 2 (1992), 39–67.

[526] C. Rabut, *High level m-harmonic cardinal B-splines*, Numer. Algorithms, 2 (1992), 68–84.

[527] D. L. Ragozin and J. Levesley, *Zonal kernels, approximations and positive definiteness on spheres and compact homogeneous spaces*, in Curves and Surfaces with Applications in CAGD, A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press, Nashville TN, 1997, 371–378.

[528] C. H. Reinsch, *Smoothing by spline functions*, Numer. Math. 10 (1967), 177–183.

[529] F. Riesz and B. Sz.-Nagy, *Functional Analysis*, Dover Publications, New York, 1955 (republished 1990).

[530] S. Rippa, *Interpolation and smoothing of scattered data by radial basis functions*, M.S. Thesis, Tel Aviv University, 1984.

[531] S. Rippa, *An algorithm for selecting a good value for the parameter c in radial basis function interpolation*, Adv. Comput. Math. 11 (1999), 193–210.

[532] A. Ron, *The L_2-approximation orders of principal shift-invariant spaces generated by a radial basis function*, in Numerical Methods in Approximation Theory, ISNM 105, D. Braess, L. L. Schumaker (ed.), Birkhäuser, Basel, 1992, 245–268.

[533] A. Ron and X. Sun, *Strictly positive definite functions on spheres*, Math. Comp. 65 (1996), no. $\tilde{2}16$, 1513–1530.

[534] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, Ann. Math. Statist. 27 (1956), 832–837.

[535] M. Rosenblum and L. S. Davis, *An improved radial basis function network for visual autonomous road following*, IEEE Trans. Neural Networks, 7 (1996), 1111–1120.

[536] M. Rosenblum, Y. Yacoob, and L. S. Davis, *Human expression recognition from motion using a radial basis function network architecture*, IEEE Trans. Neural Networks, 7 (1996), 1121–1138.

[537] W. Rudin, *Functional Analysis*, McGraw-Hill, New York, 1973.

[538] Y. Saad and M. H. Schultz, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear equations*, SIAM J. Sci. Statist. Comput. 7 (1986), 856–869.

[539] A. Saha and C. L. Wu, *Approximation, dimension reduction and nonconvex optimization using linear superpositions of Gaussians*, IEEE Trans. on Computers, 42 (1993), 1222–1233.

[540] K. Šalkauskas, *Moving least squares interpolation with thin-plate splines and radial basis functions*, Comput. Math. Appl., 24 (1992), 177–185.

[541] K. Šalkauskas and L. Bos, *Weighted splines as optimal interpolants*, Rocky Mountain J. Math., 22 (1992), 205–217.

[542] R. M. Sanner and J. J. E. Slotine, *Gaussian networks for direct adaptive control*, IEEE Trans. Neural Networks, 3 (1992), 837–863.

[543] S. A. Sarra, *Accurate derivative approximations using radial basis functions*, preprint, Marshall University, 2003.

[544] H. C. Saundersen, *Multiquadric interpolation of fluid speeds in a natural river channel*, Comput. Math. Appl., 24 (1992), 187–193.

[545] R. Schaback, *Comparison of radial basis function interpolants*, in Multivariate Approximation: From CAGD to Wavelets, Kurt Jetter and Florencio Utreras (eds.), World Scientific Publishing, Singapore, 1993, 293–305.

[546] R. Schaback, *Approximation of polynomials by radial basis functions*, in Wavelets, Images, and Surface Fitting, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), A. K. Peters, Wellesley MA, 1994, 459–466.

[547] R. Schaback, *Lower bounds for norms of inverses of interpolation matrices for radial basis functions*, J. Approx. Theory, 79 (1994), 287–306.

[548] R. Schaback, *Creating surfaces from scattered data using radial basis functions*, in Mathematical Methods for Curves and Surfaces, M. Dæhlen, T. Lyche, and L. Schumaker (eds.), Vanderbilt University Press, Nashville, 1995, 477–496.

[549] R. Schaback, *Error estimates and condition numbers for radial basis function interpolation*, Adv. in Comput. Math., 3 (1995), 251–264.

[550] R. Schaback, *Multivariate interpolation and approximation by translates of a basis function*, in Approximation Theory VIII, Vol. 1: Approximation and Interpolation, C. Chui, and L. Schumaker (eds.), World Scientific Publishing, Singapore, 1995, 491–514.

[551] R. Schaback, *Optimal recovery in translation-invariant spaces of functions*, Annals of Numerical Mathematics, 4 (1997), 547–556.

[552] R. Schaback, *Approximation by radial basis functions with finitely many centers*, Constr. Approx., 12 (1996), 331–340.

[553] R. Schaback, *On the efficiency of interpolation by radial basis functions*, in Surface Fitting and Multiresolution Methods, A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press, Nashville TN, 1997, 309–318.

[554] R. Schaback, *Native Hilbert spaces for radial basis functions I*, in New Developments in Approximation Theory, M. W. Müller, M. D. Buhmann, D. H. Mache and M. Felten (eds.), Birkhäuser, Basel, 1999, 255–282.

[555] R. Schaback, *A unified theory of radial basis functions. Native Hilbert spaces for radial basis functions II*, J. Comput. Appl. Math. 121 (2000), 165–177.

[556] R. Schaback, *Remarks on meshless local construction of surfaces*, in The Mathematics of Surfaces, IX, R. Cipolla and R. Martin (eds.), Springer, 2000, 34–58.

[557] R. Schaback, *Stability of radial basis function interpolants*, Approximation Theory X, C. K. Chui, L. L. Schumaker, and J. Stöckler (eds.), Vanderbilt Univ. Press, Nashville, TN, 2002, 433–440.

[558] R. Schaback, *Multivariate interpolation by polynomials and radial basis functions*, preprint, Universität Göttingen, 2002.

[559] R. Schaback, *On the scope of meshless kernel methods*, preprint, Universität Göttingen, 2003.

[560] R. Schaback and H. Wendland, *Using compactly supported radial basis functions to solve partial differential equations*, in Boundary Element Technology XIII, C. S. Chen, C. A. Brebbia, and D. W. Pepper (eds.), WIT Press, 1999, 311–324.

[561] R. Schaback and H. Wendland, *Characterization and construction of radial basis functions*, in Multivariate Approximation and Applications, N. Dyn, D. Leviatan, D. Levin, and A. Pinkus (eds.), Cambridge Univ. Press, Cambridge (2001), 1–24.

[562] R. Schaback and H. Wendland, *Numerical techniques based on radial basis functions*, Curve and Surface Fitting: Saint-Malo 1999, A. Cohen, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press, Nashville, TN, 2000, 359-374.

[563] R. Schaback and H. Wendland, *Adaptive greedy techniques for approximate solution of large RBF systems*, Numer. Algorithms 24 (2000), 239–254.

[564] R. Schaback and Z. Wu, *Operators on radial functions*, J. Comput. Appl. Math. 73 (1996), 257–270.

[565] I. P. Schagen, *Sequential exploration of unknown multidimensional functions as an aid to optimization*, IMA J. Numer. Anal., 4 (1984), 337–347.

[566] R. Schimming and M. Belger, *Polyharmonic radial functions on a Riemannian manifold*, Math. Nachr., 153 (1991), 207–216.

[567] R. Schiro and G. Williams, *An adaptive application of multiquadric interpolants for numerically modeling large numbers of irregularly spaced hydrographic data*, Survg. Mapp., 44 (1984), 365–381.

[568] I. J. Schoenberg, *On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space*, Ann. of Math., 38 (1937), 787–793.

[569] I. J. Schoenberg, *Metric spaces and completely monotone functions*, Ann. of Math., 39 (1938), 811–841.

[570] I. J. Schoenberg, *Metric spaces and positive definite functions*, Trans. Amer. Math. Soc., 44 (1938), 522–536.

[571] I. J. Schoenberg, *Positive definite functions on spheres*, Duke Math. J., 9 (1942), 96–108.

[572] I. J. Schoenberg, *Spline functions and the problem of graduation*, Proc. Nat. Acad. Sci. 52 (1964), 947–950.

[573] M. Schreiner, *On a new condition for strictly positive definite functions on spheres*, Proc. Amer. Math. Soc., to appear.

[574] L. L. Schumaker, *Spline Functions: Basic Theory*, John Wiley & Sons, New York, 1981 (reprinted by Krieger Publishing 1993).

[575] C. Shannon, *Communication in the presence of noise*, Proc. IRE 37 (1949), 10–21.

[576] M. Sharan, E. J. Kansa, and S. Gupta, *Application of the multiquadric method for numerical solution of elliptic partial differential equations*, Appl. Math. Comp., to appear.

[577] E. W. Shaw and P. P. Lynn, *Area rainfall evaluation using two surface fitting techniques*, Bull. Int. Ass. Hydrol. Sci., 17 (1972), 419–433.

[578] D. Shepard, *A two dimensional interpolation function for irregularly spaced data*, Proc. 23rd Nat. Conf. ACM 1968, 517–523.

[579] T. J. Shepherd and D. S. Broomhead, *Nonlinear signal processing using radial basis functions, SPE Advanced Signal Processing Algorithms, Architectures, and Implementations*, 1348 (1990), 51–61.

[580] A. Sherstinsky and R. W. Picard, *On the efficiency of the orthogonal least squares training method for radial basis function networks*, IEEE Trans. Neural Networks, 7 (1996), 195–200.

[581] C. Shu, H. Ding, and K. S. Yeo, *Local radial basis function-based differential quadrature method and its application to solve two-dimensional incompressible Navier-Stokes equations*, Comput. Methods Appl. Mech. Engrg. 192 (2003), 941–954.

[582] R. Sibson and G. Stone, *Computation of thin-plate splines*, SIAM J. Sci. Statist. Comput., 12 (1991), 1304–1313.

[583] N. Sivakumar and J. D. Ward, *On the least squares fit by radial functions to multidimensional scattered data*, Numer. Math., 65 (1993), 219–243.

[584] B. F. Smith, P. Bjørstad, and W. D. Gropp, *Domain Decomposition. Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Cambridge University Press, Cambridge, 1996.

[585] K. T. Smith, D. C. Solmon and S. L. Wagner, *Practical and mathematical aspects of the problem of reconstructing objects from radiographs*, Bull. Amer. Math. Soc. (83) 1977, 1227–1270.

[586] T. Sonar, *Optimal recovery using thin plate splines in finite volume methods for the numerical solution of hyperbolic conservation laws*, IMA J. Numer. Anal., 16 (1996), 549–581.

[587] S. Stead, *Estimation of gradients from scattered data*, Rocky Mountain J. Math., 14 (1984), 265–279.

[588] N. C. Steele, C. R. Reeves, M. Nicholas, and P. J. King, *Radial basis function artificial neural networks for the inference process in fuzzy logic based control*, Computing, 54 (1995), 99–117.

[589] E. M. Stein and G. Weiss, *Introduction to Fourier Analysis in Euclidean Spaces*, Princeton University Press, Princeton, 1971.

[590] J. Stewart, *Positive definite functions and generalizations, an historical survey*, Rocky Mountain J. Math., 6 (1976), 409–434.

[591] J. Strain, *The fast Gauss transform with variable scales*, SIAM J. Sci. Statist. Comput.12 (1991) 1131–1139.s

[592] X. Sun, *On the solvability of radial function interpolation*, in Approximation Theory VI, C. Chui, L. Schumaker, and J. Ward (eds.), Academic Press, New York, 1989, 643–646.

[593] X. Sun, *Multivariate interpolation using ridge or related functions*, University of Texas, 1990.

[594] X. Sun, *Norm estimates for inverses of Euclidean distance matrices*, J. Approx. Theory, 70 (1992), 339–347.

[595] X. Sun, *Cardinal and scattered-cardinal interpolation by functions having non-compact support*, Comput. Math. Appl., 24 (1992), 195–200.

[596] X. Sun, *Solvability of multivariate interpolation by radial or related functions*, J. Approx. Theory, 72 (1993), 252–267.

[597] X. Sun, *Conditionally positive definite functions and their application to multivariate interpolation*, J. Approx. Theory, 74 (1993), 159–180.

[598] X. Sun, *Scattered Hermite interpolation using radial basis functions*, Linear Algebra Appl., 207 (1994), 135–146.

[599] X. Sun, *The fundamentality of translates of a continuous function on spheres*, Numerical Algorithms, 8 (1994), 131–134.

[600] X. Sun, *Cardinal Hermite interpolation using positive definite functions*, Numerical Algorithms, 7 (1994), 253–268.

[601] X. Sun, *Conditional positive definiteness and complete monotonicity*, in Approximation Theory VIII, Vol. 1: Approximation and Interpolation, C. Chui, and L. Schumaker (eds.), World Scientific Publishing, Singapore, 1995, 537–540.

[602] X. Sun and E. W. Cheney, *Fundamental sets of continuous functions on spheres*, Southwest Missouri State University, 1994.

[603] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, *Weighted least squares support vector machines: robustness and sparse approximation*, Neurocomputing, to appear.

[604] G. Szegő, *Orthogonal Polynomials*, Amer. Math. Soc. Coll. Publ. Vol. XXIII, Providence, 1959.

[605] S. Tan, J. Hao, and J. Vandewalle, *Multivariable nonlinear system identification by radial basis function neural networks*, Arch. Control Sci., 4 (1995), 55–67.

[606] A. E. Tarwater, *A parameter study of Hardy's multiquadric method for scattered data interpolation*, Lawrence Livermore National Laboratory, TR UCRL-563670, 1985.

[607] V. N. Temlyakov, *The best m-term approximation and greedy algorithms*, Adv. in Comp. Math. 8 (1998), 249–265.

[608] L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*, SIAM, Philadelphia, PA, 1997.

[609] I. Tsukanov and V. Shapiro, *Meshfree modeling and analysis of physical fields in heterogeneous media*, preprint, University of Wisconsin, 2003.

[610] M. Unser, *Sampling – 50 years after Shannon*, Proc. IEEE 88 (2000), 569–587.

[611] F. I. Utreras, *Positive thin plate splines*, Approx. Theory Appl., 1 (1985), 77–108.

[612] F. I. Utreras, *Multiresolution and pre-wavelets using radial basis functions*, in Multivariate Approximation: From CAGD to Wavelets, Kurt Jetter and Florencio Utreras (eds.), World Scientific Publishing, Singapore, 1993, 321–333.

[613] F. I. Utreras and M. L. Vargas, *Monotone interpolations of scattered data in* $\mathbb{R}^2$, Constr. Approx., 7 (1991), 49–68.

[614] M. A. Villalobos and G. Wahba, *Multivariate thin plate spline estimates for the posterior probabilities in the classification problem*, Comm. Statist. A—Theory Methods, 12 (1983), 1449–1479.

[615] G. Wahba, *Convergence rate of "thin plate" smoothing splines when the data are noisy (preliminary report)*, Springer Lecture Notes in Math., 757, 1979, 233–245.

[616] G. Wahba, *Spline Interpolation and smoothing on the sphere*, SIAM J. Sci. Statist. Comput., 2 (1981), 5–16.

[617] G. Wahba, *Erratum: Spline interpolation and smoothing on the sphere*, SIAM J. Sci. Statist. Comput., 3 (1982), 385–386.

[618] G. Wahba, *Multivariate thin plate spline smoothing with positivity and other linear inequality constraints*, Statistical image processing and graphics (Luray, Va., 1983), Statist.: Textbooks Monographs, 72, Dekker, New York, 1986, 275–289.

[619] G. Wahba, *Multivariate model building with additive interaction and tensor product thin plate splines*, in Curves and Surfaces, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), Academic Press, New York, 1991, 491–504.

[620] G. Wahba, *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics 59, SIAM, Philadelphia, 1990.

[621] G. Wahba and Z. Luo, *Smoothing spline ANOVA fits for very large, nearly regular data sets, with application to historical global climate data*, in The Heritage of P. L. Chebyshev: a Festschrift in honor of the 70th birthday of T. J. Rivlin, Ann. Numer. Math. 4 (1997), no. 1-4, 579–597.

[622] A. Wald, *Statistical Decision Functions* John Wiley & Sons, New York, 1950 (reprinted by Chelsea Publishing 1971).

[623] G. S. Watson, *Smooth regression analysis*, Sankhya, Ser. A 26 (1964), 359–372.

[624] A. R. Webb, *Multidimensional scaling by iterative majorisation using radial basis functions*, Pattern Recognition, 28 (1995), 753–759.

[625] J. H. Wells and R. L. Williams, *Embeddings and Extensions in Analysis*, Springer, Berlin, 1975.

[626] H. Wendland, *Ein Beitrag zur Interpolation mit radialen Basisfunktionen*, Diplomarbeit, Universität Göttingen, 1994.

[627] H. Wendland, *Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree*, Adv. in Comput. Math., 4 (1995), 389–396.

[628] H. Wendland, *Error estimates for interpolation by compactly supported radial basis functions of minimal degree*, J. Approx. Theory (93) 1998, 258–272.

[629] H. Wendland, *Sobolev-type error estimates for interpolation by radial basis functions*, in Surface Fitting and Multiresolution Methods, A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press, Nashville TN, 1997, 337–344.

[630] H. Wendland, *Meshless Galerkin approximation using radial basis functions*, Math. Comp. 68 (1999), 1521–1531.

[631] H. Wendland, *Numerical solution of variational problems by radial basis functions*, in Approximation Theory IX, Vol.II: Computational Aspects, Charles K. Chui, and L. L. Schumaker (eds.), Vanderbilt University Press, 1999, 361–368.

[632] H. Wendland, *Local polynomial reproduction and moving least squares approximation*, IMA J. Numer. Anal. 21 (2001), no. 1, 285–300.

[633] H. Wendland, *Moving least squares approximation on the sphere*, in Mathematical Methods for Curves and Surfaces: Oslo 2000, T. Lyche and L. L. Schumaker (eds.), Vanderbilt University Press, 2001, 517–526.

[634] H. Wendland, *Scattered Data Modelling by Radial and Related Functions*, Habilitation Thesis, Universität Göttingen, 2002.

[635] H. Wendland, *Fast evaluation of radial basis functions: Methods based on partition of unity*, Approximation Theory X: Wavelets, Splines, and Applications, C. K. Chui, L. L. Schumaker, and J. Stöckler (eds.), Vanderbilt University Press, Nashville, 2002, 473-483.

[636] H. Wendland, *Surface reconstruction from unorganized points*, preprint, Universität Göttingen, 2002.

[637] H. Wendland, *Solving large generalized interpolation problems efficiently*, preprint, Universität Göttingen, 2003.

[638] M. Weinrich, *Charakterisierung von Funktionenräumen bei der Interpolationnn mit radialen Basisfunktionen*, Universität Göttingen, 1994.

[639] D. Wetterschereck and T. Dietterich, *Improving the performance of radial basis function networks by learning center locations*, in Advances in Neural Information Processing Systems 4. Proceedings of the 1991 Conference, J. E. Moody, S. J. Hanson, and R. P. Lippmann (eds.), Morgan Kaufmann, San Mateo, CA, 1992, 1133–1140.

[640] J. M. Whittaker, *On the functions which are represented by expansions of the interpolation theory*, Proc. Roy. Soc. Edinburgh 35 (1915), 181–194.

[641] E. T. Whittaker, *On a new method of graduation*, Proc. Edinburgh Math. Soc. 41 (1923), 63–75.

[642] B. A. Whitehead, *Genetic evolution of radial basis function coverage using orthogonal niches*, IEEE Trans. Neural Networks, 7 (1996), 1525–1528.

[643] B. A. Whitehead and T. D. Choate, *Cooperative-competitive genetic evolution of radial basis function centers and widths for time series prediction*, IEEE Trans. Neural Networks, 7 (1996), 869–880.

[644] D. V. Widder, *The Laplace Transform*, Princeton University Press, Princeton, 1941.

[645] R. E. Williamson, *Multiply monotone functions and their Laplace transform*, Duke Math. J., 23 (1956), 189–207.

[646] A. S. M. Wong, Y. C. Hon, S. L. Chung, and K. C. Ho, *A monitoring model on eutrophication for growth dynamics of phytoplankton in the Tolo Harbour of Hong Kong*, City University of Hong Kong, 1996.

[647] S. M. Wong, Y. C. Hon and T. S. Li, *Radial basis functions with compactly support and multizone decomposition: applications to environmental modelling*, in Boundary Element Technology XIII, C. S. Chen, C. A. Brebbia, and D. W. Pepper (eds.), WIT Press, 1999, 355–364.

[648] W. S. B. Woolhouse, *Explanation of a new method of adjusting mortality tables, with some observations upon Mr. Makeham's modification of Gompertz's theory*, J. Inst. Act. 15 (1870), 389–410.

[649] Z. Wu, *Die Kriging-Methode zur Lösung mehrdimensionaler Interpolationsprobleme*, Universität Göttingen, 1986.

[650] Z. Wu, *A class of convexity-preserving bases for radial basis interpolation* (Chinese), Math. Appl., 3 (1990), 33–37.

[651] Z. Wu, *Hermite-Birkhoff interpolation of scattered data by radial basis functions*, Approx. Theory Appl., 8 (1992), 1–10.

[652] Z. Wu, *A convergence analysis for a class of radial basis interpolations* (Chinese), Gaoxiao Yingyong Shuxue Xuebao, 6 (1991), 331–336.

[653] Z. Wu, *On the convergence of the interpolation with radial basis function*, Chinese J. Contemp. Math., 14 (1993), 269–277.

[654] Z. Wu, *Convergence of interpolation by radial basis functions* (Chinese), Chinese Ann. Math. Ser. A, 14 (1993), 480–486.

[655] Z. Wu, *Characterization of positive definite radial functions*, in Mathematical Methods for Curves and Surfaces, M. Dæhlen, T. Lyche, and L. Schumaker (eds.), Vanderbilt University Press, Nashville, 1995, 573–578.

[656] Z. Wu, *Compactly supported positive definite radial functions*, Adv. in Comput. Math., 4 (1995), 283–292.

[657] Z. Wu, *Generalized Strang-Fix condition for scattered data quasi-interpolation*, preprint, Fudan University, 2002.

[658] Z. Wu and R. Schaback, *Local error estimates for radial basis function interpolation of scattered data*, IMA J. Numer. Anal., 13 (1993), 13–27.

[659] Z. Wu and R. Schaback, *Shape preserving properties and convergence of univariate multiquadric quasi-interpolation*, Acta Math. Appl. Sinica (English Ser.), 10 (1994), 441–446.

[660] L. Xu, A. Kryzak, and A. Yuille, *On radial basis function nets and kernel regression statistical consistency, convergence rates, and receptive field sizes*, Neural Networks, 7 (1994), 609–628.

[661] Y. Xu and E. W. Cheney, *Interpolation by periodic radial functions*, Comput. Math. Appl., 24 (1992), 201–215.

[662] Y. Xu and E. W. Cheney, *Strictly positive definite functions on spheres*, Proc. Amer. Math. Soc., 116 (1992), 977–981.

[663] Y. Xu, W. A. Light, and E. W. Cheney, *Constructive methods of approximation by ridge functions and radial functions*, Numer. Algorithms, 4 (1993), 205–223.

[664] T. Yamada and L. C. Wrobel, *Properties of Gaussian radial basis functions in the dual reciprocity boundary element method*, Z. Angew. Math. Phys., 44 (1993), 1054–1067.

[665] Y. Yang and A. Barron, *Information-theoretic determination of minimax rates of convergence*, Ann. Statist. 27 (1999), 1564–1599.

[666] P. V. Yee and S. Haykin, *Regularized Radial Basis Function Networks: Theory and Applications*, Wiley-Interscience, 2001.

[667] J. Yoon, *Interpolation by radial basis functions on Sobolev space*, J. Approx. Theory 112 (2001), 1–15.

[668] J. Yoon, *L_p-error estimates for "shifted" surface spline interpolation on Sobolev space*, Math. Comp. 72 (2003), 1349–1367.

[669] D. Zeckzer, *Dreiecks-basierte lokale Scattered Data Interpolation unter Verwendung radialer Basismethoden*, Diplomarbeit, Universität Kaiserslautern, 1992.