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 Connecticut), and S. A. Nelson (Iowa State Univ.), Multivariate interpolation: A variational theory, unpublished manuscript, 1983 (proved Franke's conjecture).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.