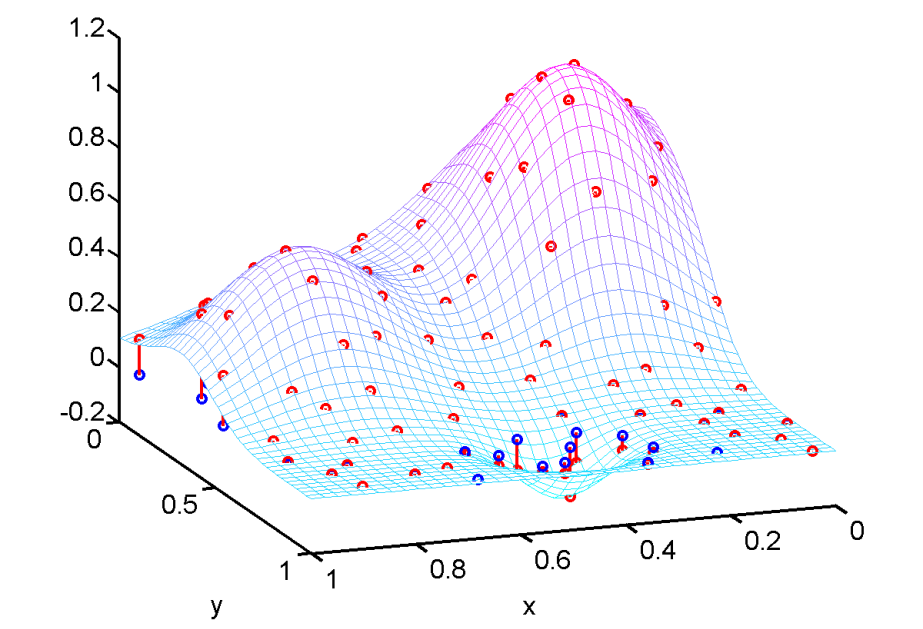
Lectures on Radial Basis Functions

by **Stefano De Marchi**
Department of Mathematics "Tullio Levi-Civita"
University of Padua (Italy)



February 13, 2018

These lecture notes were inspired mainly by two seminal books on the topic by Holger Wendland [74] and by Gregory E. Fasshauer [32]. The first presents more theoretical aspects while the second provides also useful Matlab functions for understanding better the theory and all the applictions discussed. The notes have then been used during a short teaching-visit of the author to the University of Antwerp, for the Erasmus Teaching Staff Mobility.

People interested on radial basis functions, can refer to the wide literature available that, especially in the last two decades, has grown very fast. The popularity of radial basis functions can be understood by means of the following "parallelism". In many cooking recepies the parsley is used to give flavour and colour to dishes. Radial basis functions can be considered as a *mathematical parsley* since they have been used in all mathematical problems requiring a powerful, i.e. efficient and stable, approximation tool.

These lectures were thoughts for students without a strong background on functional analysis, so in the presentation of the topics we deliberately avoid, when possible, to introduce functional analysis concepts. This is a great lack, but we hope that the people who will use these notes will be not too critical to us.

Moreover, these are only **introductory lectures** and some **examples** on the topic. Many important aspects and applications, for lack of time and because there are many other books (see e.g. [74, 32, **?**]), are not considered. Every lecture provides also a set of exercises solvable by using Matlab. This choice has been done with the aim of making the discussion more interesting from both the numerical and geometrical point of view.

We do hope that after this brief introduction, interested students will be encouraged and also interested in getting into this fascinating mathematical tool that has become more and more popular in used and popular in many fields.

Stefano De Marchi Padova, February 2018.

Contents

List of Figures

List of Tables

Introduction

These lectures focus on recent theoretical and computational aspects (implemented with the MATLAB software) in the field of *scattered data approximation*. More precisely, our attention is devoted to *meshfree* or *meshless* methods. Indeed, taking advantage of being independent of a mesh, they turn out to be truly performing in high dimensions. Unfortunately, not all of them allow to deal with a *large* number of points. Therefore, we also review the basic theory concerning local methods and in particular the Partition of Unity (PU) method. Furthermore, we give a general outline of collocation methods via RBF.

Historical remarks

Even if here we consider meshfree methods in the context of approximation theory, we have to point out that they have been used for the first time in statistics. In fact, their study began in 1870 with the work of W.S.B. Woolhouse [76]. It has been further developed few years later by E.L. De Forest [21, 22] and in 1883 by J.P. Gram [38]. Such works gave rise to *moving least squares* approximation (also known as *local regression* in statistic literature).

For what concerns approximation theory, the historical and theoretical foundation of meshless methods lies in the concept of *positive definite functions* or, more in general, *positive definite kernels*. Their development can be traced back to both the work of J. Mercer (1909) [57], a Fellow of Trinity College at Cambridge University, and the one of M. Mathias (1923) [50], who was a student at the University of Berlin. Their studies have been linked to Fourier transforms in the 1930s by S. Bochner [5], lecturer at the University of Munich at that time, and by I. Schoenberg [66], who was assistant professor at Colby College. Further studies about positive definite functions up to the mid 1970s have been collected by J. Stewart [71]. Later, C.A. Micchelli in 1986 made the connection between scattered data interpolation and positive definite functions [59]. At the same time, E.J. Kansa gave rise to the first methods for solving PDEs via meshless methods [48].

Many positive definite functions are nowadays classified as RBFs. Such term appeared for the first time in a publication made by N. Dyn and D. Levin in 1983 [30]. However, earlier in the 1970s, many researchers introduced functions that are now called RBFs. For instance, the term *multiquadrics* is due to R. Hardy, geodesist at Iowa State University [40].

Moreover, in 1972 R.L. Harder and R.N. Desmarais, two aerospace engineers, introduced the Thin Plate Splines (TPSs) [39]. Always at the same time, such functions appeared in several papers by J. Duchon, a mathematician at the Université Joseph Fourier in Grenoble [27, 28, 29]. Furthermore, we also have to mention J. Meingnet who, working at the Université Catholique de Luouvain, in 1970s introduced what are now known as polyharmonic splines [52, 53, 54, 55]. However, in the theory of RBFs the main step forward has been carried out by R. Franke in the 1980s. He proposed a comprehensive study about several methods for multivariate interpolation [34, 35].

Finally, concerning local techniques, we can affirm that the pioneer of the PU scheme has been D. Shepard. In fact, in the late 1960s he introduced, as an undergraduate student at Harvard University, what are now called the *Shepard weights* [69]. His studies have been motivated by the fields of geophysics and meteorology. Moreover, related to the same topics, we also have to mention few papers that contain sentences and traces which seem to indicate that the Shepard's scheme was known before his work (see the paper by I.K. Crain and B.K. Bhattacharyya (1967) [18] or the one by W.R. Goodin et al. [36]).

Aside from this short historical remark, in this dissertation we mainly refer to the recent books by M.D. Buhmann, G.E. Fasshauer, M.J. McCourt and H. Wendland [8, 32, 33, 74].

Lecture 1

Learning from splines

1.1 Motivations

In practical applications we have to face the problem of reconstructing an unknown function f from a set (usually small) of data. These data consist of two sets: the *data sites* $X = \{x_1, \dots, x_N\}$ and the *data values* $f_j = f(x_j),\ j = 1, \dots, N$. The reconstruction has to approximate the data values at the data sites. In practice we are looking for a function s that either *interpolates* the data, i.e. it satisfies the conditions $s(x_j) = f_j,\ 1 \leq j \leq N$ or *approximate* the data, i.e. $s(x_j) \approx f_j$. This latter is important, for example, when the data come from some measurement or contain noise.

In many cases the data are *scattered*, that is they have no special structure, and they are in a big amount (several millions). Moreover in several applications the data sites are considered in high dimension. Hence, for a unifying approach, methods have been developed in the last decades with the aim to meet all these (new) situations.

$\diamond\diamond$

We start from the univariate setting. We suppose that the data sites are ordered as follows

$$X : a < x_1 < x_2 < \cdots < x_N < b \tag{1.1}$$

and we have some data values $f_1, \dots, f_N$ to be interpolated at the data set X. What we want to do, mathematically speaking, is finding $s : [a, b] \to \mathbb{R}$ with the property $s(x_j) = f_j$ for all $j = 1, \dots, N$.

Notice, that the data values f_j is not necessary stem from a function f but we shall keep in mind this possibility for reasons that will become clearer later.

In the *univariate* setting, a simple solution of the above problem consists in taking s as polynomial p of degree at most $N - 1$. However, as we can see later, this solution is not working in higher dimensions. Remaining in the univariate case, **no one with**

experience in approximation theory would even try to interpolate a hundred thousand points with a polynomial. Indeed it is a well-established fact that a large data set is better dealt with *splines* than by polynomials. One aspect to notice in contrast to polynomials, the accuracy of the interpolation process using splines is not based on the polynomial degree but on the spacing of the data sites.

Let us review briefly the main properties of univariate splines, especially in the case of *cubic splines*. The set of cubic splines corresponding to the subdivision (1.1) is the space

$$\mathcal{S}_3(X) = \{s \in \mathcal{C}^2[a,b] : s_{|[x_i,x_{i+1}]} \in \mathbb{P}_3(\mathbb{R}),\ 0 \le i \le N\} \tag{1.2}$$

with $a = x_0$ and $x_{N+1} = b$. The space $S_3(X)$ has dimension $N+4$, so that the interpolation conditions $s(x_i) = f_i,\ 1 \le i \le N$ are not sufficient to guarantee a unique interpolant. To enforce uniqueness, in the case of *natural splines*, i.e. the set

$$\mathcal{N}_3(X) = \{s \in \mathcal{S}_3(X) : s_{|[a,x_1]}, s_{|[x_N,b]} \in \mathbb{P}_1(\mathbb{R})\} \tag{1.3}$$

that consists of all cubic splines that are linear polynomials on the outer intervals $[a, x_1]$ and $[x_N, b]$. It come easy to see that a cubic spline s is a natural spline if and only if it satisfies $s''(x_1) = s^{(3)}(x_1) = 0$ and $s''(x_N) = s^{(3)}(x_N) = 0$. With this choice we have imposed 4 additional conditions to the space, so it is natural to assume that the $\texttt{dim}(\mathcal{N}_3(X)) = N$. Even more, it can be shown that the initial interpolation problem has a unique solution in $\mathcal{N}_3(X)$.

Here some important properties of splines that are worth to be mentioned

1. They are piecewise polynomials.

2. An interpolating natural cubic spline satisfies a *minimal norm property*. Assume that f comes from the *Sobolev space* $\mathcal{H}^2[a,b]$, i.e. f is continuous in $[a,b]$ and has weak first and second order derivatives in $L_2[a,b]$ (a more precise definition will be done later or can be found in any books of functional analysis). Assume further that f is such that $f(x_j) = f_j, 1 \le j \le N$. If $s_{f,X}$ denotes the natural cubic spline interpolant (at the data set X) then

$$(f'' - s''_{f,X}, s''_{f,X})_{L_2[a,b]} = 0\,.$$

   This leads to the Pythagorean equation

$$\|f'' - s''_{f,X}\|^2_{L_2[a,b]} + \|s''_{f,X}\|^2_{L_2[a,b]} = \|f''\|^2_{L_2[a,b]}\,,$$

   indicating that the natural cubic splines interpolant is the function from $\mathcal{H}^2[a,b]$ that minimizes the semi-norm $\|f''\|^2_{L_2[a,b]}$ under the conditions $f(x_j) = f_j,\ 1 \le j \le N$.

3. They possess a local basis called *B-splines*. This basis, which is more stable than any other, can be defined by recursion, by divided differences of the *truncated cubic power* $p(x;t) = (x-t)^3_+$ or by convolution. Here x_+ takes the value of x for nonnegative x and zero otherwise.

Interested readers on splines and their many properties can refer to the following fundamental books by Schumaker [67] or de Boor [23].

Remarks

- Property 1. combined with the local basis, not only allows the efficient computation and evaluation of splines but also is the key ingredient for a simple error analysis. Hence, **the natural way of extending splines to the multivariate setting is based on this property**. To this end, a bounded region $\Omega \subset \mathbb{R}^d$ is partitioned into essentially disjoint regions $\{\Omega_j\}_{j=1}^N$ (patches). Then the spline space consists of those functions s that are piecewise polynomials on each Ω_j and that have smooth connections on the boundaries of two adjacent patches. In two dimensions the most popular subdivision of a polygonal region is by triangles. It is interesting to note that even in this simple case, the dimension of the spline space is in general unknown. When coming to higher dimensions it is not all clear what an appropriate replacement for the triangulation would be. Hence, even if great progresses have been made in the 2-dimensional setting, the method is not suited for general dimensions.
- Another possible generalization to the multivariate setting is based on the property 3. In particular a construction based on convolution led to the so called *Box-splines*. Again, even in the 2-dimensions the problem can be handle, for higher dimensions is still an open problem.

The property 2. is the motivation for a general framework in higher dimensions. This approach has allowed to develop a beautiful theory where *all space dimensions can be handled in the same way*. The resulting approximation spaces no longer consist of piecewise polynomials, so they can not be called splines. The new functions are known with the "fashionable words" of *Radial Basis Functions* (in what follows we refer to them simply as RBF).

1.2 From cubic splines to RBF

To get a better idea, let us remind that the set $\mathcal{S}_3(X)$ has the basis of truncated powers $(\cdot - x_j)_+^3$, $1 \le j \le N$ plus an arbitrary basis for $\mathbb{P}_3(\mathbb{R})$. Hence every $s \in \mathcal{N}_3(X)$ can be represented in the form

$$s(x) = \sum_{j=1}^{N} a_j (x - x_j)_+^3 + \sum_{j=0}^{3} b_j x^j, \quad x \in [a, b]\,. \tag{1.4}$$

Because s is a *natural spline* we have the additional information that s is linear on the two outer intervals. That is on $[a, x_1]$ the spline is simply $s(x) = b_0 + b_1 x$ (since $b_2 = b_3 = 0$). Thus (1.4) becomes

$$s(x) = \sum_{j=1}^{N} a_j (x - x_j)_+^3 + b_0 + b_1 x, \quad x \in [a, x_1]\,. \tag{1.5}$$

To derive the representation of s in $[x_N, b]$ we have simply to remove all subscripts + on the functions $(\cdot - x_j)_+^3$ in (1.5). Expanding these cubics and rearranging the sums we get

$$s(x) = \sum_{l=0}^{3} \binom{3}{l} (-1)^{3-l} \left(\sum_{j=1}^{N} a_j x_j^{3-l} \right) x^l + b_0 + b_1 x, \quad x \in [x_N, b]\,. \tag{1.6}$$

Thus, for s to be a natural spline, the coefficients of s have to satisfy

$$\sum_{j=1}^{N} a_j = \sum_{j=1}^{N} a_j x_j = 0\,. \tag{1.7}$$

This is a first characterization of natural cubic splines.

One more step. Using the identity $x_+^3 = \frac{(|x|^3 + x^3)}{2}$, and thanks to the relations (1.7) we get

$$\begin{aligned} s(x) &= \sum_{j=1}^{N} \frac{a_j}{2} |x - x_j|^3 + \sum_{j=1}^{N} \frac{a_j}{2} (x - x_j)^3 + b_0 + b_1 x \\ &= \sum_{j=1}^{N} \frac{a_j}{2} |x - x_j|^3 + \sum_{l=0}^{3} \frac{1}{2} \binom{3}{l} (-1)^{3-l} \left(\sum_{j=1}^{N} a_j x_j^{3-l} x^l \right) + b_0 + b_1 x \\ &= \sum_{j=1}^{N} \tilde{a}_j |x - x_j|^3 + \tilde{b}_0 + \tilde{b_1} x\,, \end{aligned}$$

where $\tilde{a}_j = a_j/2,\ 1 \le j \le N$, $\tilde{b}_0 = b_0 - \frac{1}{2}\sum_{j=1}^{N} a_j x_j^3$ and $\tilde{b}_1 = b_1 + \frac{3}{2}\sum_{j=1}^{N} a_j x_j^2$.

Proposition 1. *Every natural spline s has the representation*

$$s(x) = \sum_{j=1}^{N} a_j \phi(|x - x_j|) + p(x), \quad x \in \mathbb{R} \tag{1.8}$$

where $\phi(r) = r^3,\ r \ge 0$ and $p \in \mathbb{P}_1(\mathbb{R})$. The coefficients $\{a_j\}$ have to satisfy the relations (1.7). On the contrary, for every set $X = \{x_1, \dots, x_N\} \subset \mathbb{R}$ of pairwise distinct points and for every $f \in \mathbb{R}^N$ there exists a function s of the form (1.8), with (1.7), that interpolates the data, i.e. $s(x_j) = f(x_j),\ 1 \le j \le N$.

This is the starting point for understanding the origin of RBF. The resulting interpolant is, up to a low-degree polynomial, a linear combination of *shifts* of a radial function $\Phi = \phi(|\cdot|)$. The function is then called *radial* because is the composition of a univariate function with the Euclidean norm on $\mathbb{R}$.

The generalization to $\mathbb{R}^d$ is straightforward where the name "radial" becomes even more evident. In fact

$$s(x) = \sum_{i=1}^{N} a_j \phi(\|x - x_j\|_2) + p(x), \quad x \in \mathbb{R}^d, \tag{1.9}$$

where $\phi : [0, \infty) \to \mathbb{R}$ is a univariate fixed function and $p \in \mathbb{P}_{m-1}(\mathbb{R}^d)$ is a low degree d-variate polynomial. The additional conditions on the coefficients (corresponding to (1.7)) become

$$\sum_{i=1}^{N} a_j q(x_j) = 0, \quad \forall\, q \in \mathbb{P}_{m-1}(\mathbb{R}^d). \tag{1.10}$$

In many cases (see Lecture 2), we can avoid the *side conditions* on the coefficients (1.10). In these cases the interpolation problem has solution if the matrix

$$A_{\phi,X} := (\phi(\|x_i - x_j\|_2)_{1 \le i,j \le N},$$

is invertible. To be more precise we ask

Problem 1. *Does there exist a function $\phi : [0, \infty) \to \mathbb{R}$ such that for all $d, N \in \mathbb{N}$ and all pairwise distrinct $x_1, \ldots, x_n \in \mathbb{R}^d$ the matrix $A_{\phi,X}$ is nonsingular ?*

The answer is affirmative. Examples of functions that allow to build matrices nonsingular are: the *gaussians* $\phi(r) = e^{-\alpha r^2}$, $\alpha > 0$, the *inverse multiquadric* $\phi(r) = (c^2 + r^2)^{-1/2}$ and the *multiquadric* $\phi(r) = (c^2 + r^2)^{1/2}$, $c > 0$. In the two first cases it is even true that the matrix $A_{\phi,X}$ is always positive definite (and so invertible).

Remark. In what follows, in the context of RBFs, instead of $A_{\phi,X}$ we shall use simply A thinking to the interpolation matrix with radial basis functions.

1.3 The scattered data interpolation problem

In many disciplines one faces the following problem: *we are given a set of data (measurements, locations at which these measurements are taken,...) and we want to find a rule which allows to get information about the process we are studying also at locations different from those at which the measurements are taken (or provided).*

The main reasons why we are interested on such a problem in our setting are:

- Scattered data fitting is a fundamental problem in approximation theory and data modeling in general
- Mathematical challenge: we want a *well-posed problem formulation*
- This will naturally lead to *distance matrices*
- Later we generalize to *radial basis functions* or *positive definite kernels*

Problem 2. *Given data* $(\boldsymbol{x}_j, y_j)$, $j = 1, \ldots, N$, *with* $\boldsymbol{x}_j \in \mathbb{R}^d$, $y_j \in \mathbb{R}$, *find a (continuous) function* $\mathcal{P}_f$ *(depending on* f*) such that* $\mathcal{P}_f(\boldsymbol{x}_j) = y_j$, $j = 1, \ldots, N$.

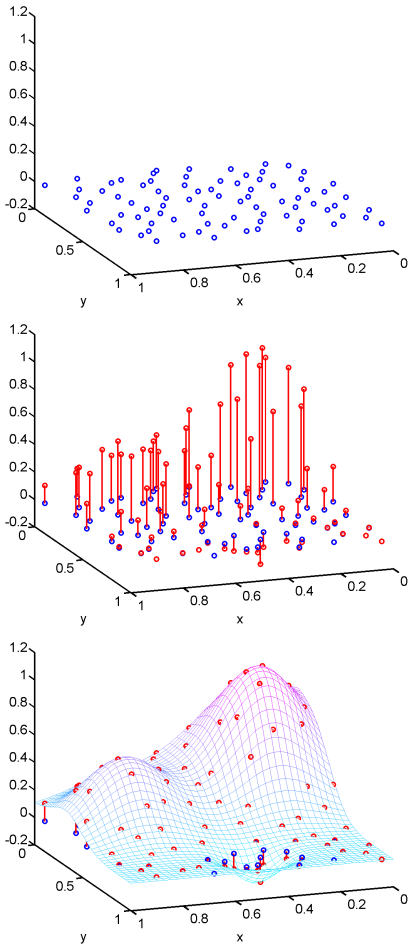


Figure 1.1: Data points, data values and data function

Now, assume $\mathcal{P}_f$ is a linear combination of certain *basis functions* B_k, that is

$$\mathcal{P}_f(\boldsymbol{x}) = \sum_{k=1}^{N} c_k B_k(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^d. \tag{1.11}$$

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 $\boldsymbol{y} = [y_1, \ldots, y_N]^T$.

The scattered data fitting problem will be *well-posed*, that is a solution to the problem will exist and be unique, *if and only if the matrix* A *is non-singular.*

1.3.1 The Haar-Mairhuber-Curtis theorem

We need to introduce the following definition

Definition 1. *Let the finite-dimensional linear function space $\mathcal{B} \subset \mathcal{C}(\Omega)$ have a basis $\{B_1, \dots, B_N\}$. Then $\mathcal{B}$ is a* **Haar space** *on Ω if*

$$\texttt{det}(A) \neq 0$$

for any set of distinct points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N \in \Omega$. Here, A is the matrix with entries $A_{i,j} = B_j(\boldsymbol{x}_i)$.

The existence of a Haar space guarantees the invertibility of the matrix A. 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 $N-1$. This means that the polynomials of degree $N-1$ form an N-dimensional Haar space for the set of distinct points $X = \{x_1, \dots, x_N\}$.

This is a counterexample useful to understand the necessity of a different approach than using polynomials.

Example 1. It is not possible to perform unique interpolation with (multivariate) polynomials of degree N to data given at arbitrary locations in $\mathbb{R}^2$.

The Haar-Mairhuber-Curtis theorem tells us that if we want to have a well-posed multivariate scattered data interpolation problem **we can no longer fix in advance the set of basis functions we plan to use for interpolation of arbitrary scattered data**. Instead, the *basis should depend on the data points.*

Theorem 1. (**Haar-Mairhuber-Curtis**)
If $\Omega \subset \mathbb{R}^d$, $d \geq 2$ contains an interior point, then there exist no Haar spaces of continuous functions except for the 1-dimensional case.

Proof. Let $d \geq 2$ and assume that $\mathcal{B}$ is a Haar space with basis $\{B_1, \dots, B_N\}$ with $N \geq 2$. We show that this leads to a contradiction. In fact, let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ be a set of N distinct points in $\Omega \subset \mathbb{R}^d$ and A the matrix such that $A_{j,k} = B_k(\boldsymbol{x}_j),\ j,k = 1, \dots, N$. By the above definition of Haar space $\texttt{det}(A) \neq 0$. Now, consider the closed path P in Ω connecting only $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. *This is possibile since Ω by assumption contains an interior point.* We then can exchange $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ by moving them continuosly along P (without interfering with other point $\boldsymbol{x}_j$). This means that the rows 1 and 2 of the matrix A have been changed 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 $\texttt{det}(A) = 0$ at some point along P. This contradicts the fact that $\texttt{det}(A) \neq 0$. ■

1.3.2 Distance matrices

We want to construct a (continuous) function $\mathcal{P}_f$ that interpolates samples obtained from a test function f_d at data sites $\boldsymbol{x}_j \in [0,1]^d$. that is we want

$$\mathcal{P}_f(\boldsymbol{x}_j) = f_d(\boldsymbol{x}_j), \qquad \boldsymbol{x}_j \in [0,1]^d .$$

For example $f_d(\boldsymbol{x}) = 4^d \prod_{k=1}^{d} x_k(1-x_k)$, $\boldsymbol{x} = (x_1, \ldots, x_d) \in [0,1]^d$, which is zeros on the boundary of the unit cube in $\mathbb{R}^d$ and has a maximum value of one at the center of the d-dimensional cube.

Assume for now that $\mathbf{d = 1}$. We have already seen that for small N one can use univariate polynomials; if N is relatively large it is better to use splines (the simplest approach is the C^0 piecewise linear splines "connect the dots")

A basis for the space of piecewise linear interpolating splines is

$$\{B_k = |\cdot - x_k| \;\; k = 1, \ldots, N\} .$$

Hence our spline interpolant can be written as

$$\mathcal{P}_f(x) = \sum_{k=1}^{N} c_k |x - x_k|, \qquad x \in [0,1]$$

and the coefficents c_k will be determined by the interpolation conditions

$$\mathcal{P}_f(x_j) = f_1(x_j), \qquad j = 1, \ldots, N$$

Some observations.

- The basis functions $B_k = |\cdot - x_k|$ are dependent on the data sites x_k as suggested by Haar-Mairhuber-Curtis.
- $B(x) = |x|$ is called *basic function*.
- The points x_k to which the basic function is shifted to form the basis functions, are usually referred to as *centers* or *knots*.
- Technically, one could choose these centers different from the data sites. However, usually *centers coincide with the data sites*. This simplifies the analysis of the method, and is sufficient for many applications. In fact, relatively little is known about the case when centers and data sites differ.
- B_k are (radially) symmetric about their centers x_k, $\longrightarrow$ **radial basis function**.

Now the coefficients c_k in the scattered data interpolation problem are found by solving the linear system

$$\begin{bmatrix} |x_1-x_1| & |x_1-x_2| & \dots & |x_1-x_N| \\ |x_2-x_1| & |x_2-x_2| & \dots & |x_2-x_N| \\ \vdots & \vdots & \ddots & \vdots \\ |x_N-x_1| & |x_N-x_2| & \dots & |x_N-x_N| \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} f_1(x_1) \\ f_1(x_2) \\ \vdots \\ f_1(x_N) \end{bmatrix} \tag{1.12}$$

- The matrix in (1.12) is an example of **distance matrix**.
- Distance matrices have been studied in geometry and analysis in the context of isometric embeddings of metric spaces for a long time.
- It is known that the distance matrix based on the Euclidean distance between a set of distinct points in $\mathbb{R}^d$ is always non-singular (see below).
- Therefore, **our scattered data interpolation problem is well-posed**.

Since distance matrices are non-singular for Euclidean distances *in any space dimension* d we have an immediate generalization: for the scattered data interpolation problem on $[0,1]^d$ we can take

$$\mathcal{P}_f(\boldsymbol{x}) = \sum_{k=1}^{N} c_k \|\boldsymbol{x}-\boldsymbol{x}_k\|_2, \qquad \boldsymbol{x} \in [0,1]^d, \tag{1.13}$$

and find the c_k by solving

$$\begin{bmatrix} \|\boldsymbol{x}_1-\boldsymbol{x}_1\|_2 & \|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2 & \dots & \|\boldsymbol{x}_1-\boldsymbol{x}_N\|_2 \\ \|\boldsymbol{x}_2-\boldsymbol{x}_1\|_2 & \|\boldsymbol{x}_2-\boldsymbol{x}_2\|_2 & \dots & \|\boldsymbol{x}_2-\boldsymbol{x}_N\|_2 \\ \vdots & \vdots & \ddots & \vdots \\ \|\boldsymbol{x}_N-\boldsymbol{x}_1\|_2 & \|\boldsymbol{x}_N-\boldsymbol{x}_2\|_2 & \dots & \|\boldsymbol{x}_N-\boldsymbol{x}_N\|_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} f_d(\boldsymbol{x}_1) \\ f_d(\boldsymbol{x}_2) \\ \vdots \\ f_d(\boldsymbol{x}_N) \end{bmatrix}.$$

- Note that the basis is again data dependent
- Piecewise linear splines in higher space dimensions are usually constructed differently (via a cardinal basis on an underlying computational mesh)
- For $d > 1$ the space $\texttt{span}\{\|\cdot-\boldsymbol{x}_k\|_2,\ k=1,\dots,N\}$ is **not** the same as piecewise linear splines

In order to show the non-singularity of our distance matrices we use the Courant-Fischer theorem (see for example the book by Meyer [58]):

Theorem 2. *Let A be a real symmetric $N \times N$ matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$, then*

$$\lambda_k = \max_{\dim \mathcal{V}=k} \min_{\substack{\boldsymbol{x}\in\mathcal{V} \\ \|\boldsymbol{x}\|=1}} \boldsymbol{x}^T A \boldsymbol{x} \quad \textit{and} \quad \lambda_k = \min_{\dim \mathcal{V}=N-k+1} \max_{\substack{\boldsymbol{x}\in\mathcal{V} \\ \|\boldsymbol{x}\|=1}} \boldsymbol{x}^T A \boldsymbol{x}.$$

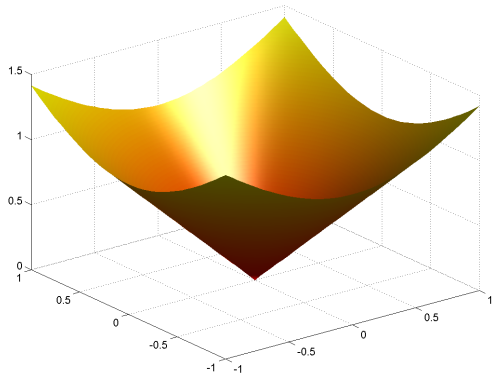


Figure 1.2: A typical basis function for the Euclidean distance matrix fit, $B_k(\boldsymbol{x}) = \|\boldsymbol{x}-\boldsymbol{x}_k\|_2$ with $\boldsymbol{x}_k = \mathbf{0}$ and $d = 2$.

Definition 2. *A real symmetric matrix A is called* **Conditionally Negative Definite (CND)** *of order one (or* almost negative definite*) if its associated quadratic form is negative, that is*

$$\sum_{j=1}^{N}\sum_{k=1}^{N} c_j c_k A_{jk} < 0 \tag{1.14}$$

for all $\boldsymbol{c} = [c_1, \dots, c_N]^T \neq \mathbf{0} \in \mathbb{R}^N$ that satisfy $\sum_{j=1}^{N} c_j = 0$.

A matrix which is CND of order 1, is CND on a subspace of dimension $N-1$, that is it has at least $N-1$ negative eigenvalues. Indeed holds the following result.

Theorem 3. *An $N \times N$ symmetric matrix A which is almost negative definite and has a non-negative trace possesses one positive and $N-1$ negative eigenvalues.*

Proof. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ denote the eigenvalues of A. From the Courant-Fischer theorem we get

$$\lambda_2 = \min_{\dim\mathcal{V}=N-1} \max_{\substack{\boldsymbol{x}\in\mathcal{V} \\ \|\boldsymbol{x}\|=1}} \boldsymbol{x}^T A\boldsymbol{x} \leq \max_{\boldsymbol{c}:\ \sum c_k=0 \atop \|\boldsymbol{c}\|=1} \boldsymbol{c}^T A\boldsymbol{c} < 0,$$

so that A has at least $N-1$ negative eigenvalues. But since $\mathrm{tr}(A) = \sum_{k=1}^{N} \lambda_k \geq 0$, A also must have at least one positive eigenvalue. ■

Example 2. It is known that $\phi(r) = r$ is a *strictly conditionally negative definite function* of order one, i.e., the matrix A with $A_{jk} = \|\boldsymbol{x}_j - \boldsymbol{x}_k\|_2$ is almost negative definite. Moreover, since $A_{jj} = \phi(0) = 0$, $j = 1, \dots, N$ then $\mathrm{tr}(A) = 0$. Therefore, our distance matrix is non-singular by the above theorem.

1.3.3 Data sets

Depending on the type of approximation problem we are given, we may or may not be able to select where the data is collected, i.e., the location of the data sites or *design*.

Standard choices in low space dimensions are depicted in Fig. 1.3. In *higher space dimensions* it is important to have space-filling (or low-discrepancy) quasi-random point sets. Examples include: **Halton points, Sobol points, lattice designs, Latin hypercube designs and quite a few others (digital nets, Faure, Niederreiter, etc.)**.

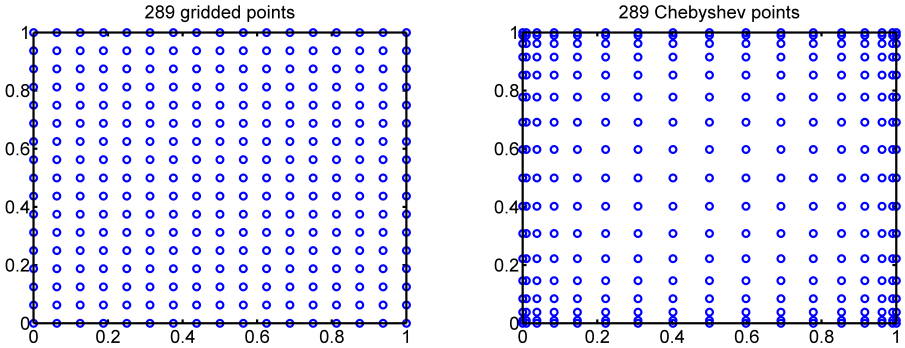


Figure 1.3: tensor products of equally spaced points and tensor products of Chebyshev points

Definition 3. *Given a sequence* $X = \{x_1, ..., x_N\}$ *its discrepancy is*

$$D_N(X) := \sup_{B \in J} \left| \frac{\#(X, B)}{N} - \lambda_d(B) \right| \tag{1.15}$$

where

- $J := \prod_{i=1}^{d} [a_i, b_i) = \{x \in \mathbb{R}^d : \ a_i \leq x_i \leq b_i, \ 0 \leq a_i < b_i < 1\}$ *(d-dimensional intervals),*
- $\#(X, B)$ *is the number of points of* X *in* B*.*
- λ_d *is Lebesgue measure*

When $D_N(X) \approx \texttt{mis}(B)$ then D_N is called **low discrepancy**.

Low-discrepancy sequences are known also as *quasi-random* sequences, because they are often used as uniformely distributed random numbers.

Example 3. A typical application of such sequences is `numerical quadrature (or cubature)`. For example in the one-dimensional case

$$\int_0^1 f(t)dt \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i)\,. \tag{1.16}$$

- If in (1.16) $x_i = i/N$ then we get the rectangle formula.
- If x_i are random numbers, then (1.16) is the Montecarlo method.
- If x_i are a low discrepancy sequences then (1.16) is a quasi-Montecarlo method.

1.3.4 Halton points

Here we show how one can generate `Halton points` in every spatial dimension. Halton points are uniformely distributed random point in $(0,1)^d$ generated from `Van der Corput` sequences. We start by generating Van der Corput sequences. Let $k \in \mathbb{N}$ be chosen.

(i) Every $n \in \mathbb{N}$ can be written as

$$n = \sum_{i=0}^{k} a_i p^i$$

where the coefficients a_i are integer such that $0 \le a_i < p$. For example taking $n = 10$ and $p = 3$

$$10 = 1 \cdot 3^0 + 0 \cdot 3^1 + 1 \cdot 3^2$$

giving $k = 2, a_0 = a_2 = 1,\ a_1 = 0$.

(ii) We define the function $h_p : \mathbb{N} \to [0,1)$ as $h_p(n) = \frac{\sum_{i=0}^{k} a_i}{p^{i+1}}$. For example

$$h_p(10) = \frac{1}{3} + \frac{1}{3^3} = \frac{10}{27}\,.$$

(iii) The Van der Corput sequence is then

$$h_{p,N} = \{h_p(n) \,:\, n = 0, 1, \ldots, N\}\,.$$

In our example

$$h_{3,10} = \{0, \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \frac{1}{27}, \frac{10}{27}\}\,.$$

Starting from the Van der Corput sequence, the Halton seqnuence is generated as follows: **take** d **distinct primes** $p_1, \ldots, p_d$ **and generated** $h_{p_1,N}, \ldots, h_{p_d,N}$ that we use as coordinates of d-dimensional points, so that

$$H_{d,N} = \{(h_{p_1,N}(n), \ldots, h_{p_d,N}(n) \,:\, n = 0, \ldots, N\}$$

is the set of $N+1$ Halton points in $[0,1)^d$.

Proposition 2. *Halton points form a nested sequence, that is if $M < N$ then $H_{d,M} \subset H_{d,N}$.*

These points can be constructed sequentially. Similar to these points are **Leja sequences** [24]. As a final observation, for Halton points we have

$$D_N(H_{d,N}) \leq \frac{C(\log N)^d}{N}.$$

$\diamond\diamond$

1. In Matlab the program `haltonseq.m` by Daniel Dougherty, dowloadable at the **Matlab Central File Exchange**, generates Halton points in every space dimension. The call is `haltonseq(numpts, dim)`. Notice that in this implementation the point $\mathbf{0} = (0, ..., 0)^T$ is not part of the point set, that is they are generated starting from $n = 1$ instead of $n = 0$ as described above. In recent Matlab versions, the function `P=haltonset(n,d)` computes `n` Halton points in dimension `d`.

2. Analogously, Sobol points can be generated in Matlab by `P=sobolset(d)`. This call constructs a d-dimensional point set `P` of the sobolset class, with default property settings. *For example*, if $d = 2$, `P` is an array $intmax \times 2$. If one wishes different properties, the call become `P=sobolset(d, p1, var1, p2,var2, ...)` that specifies property name/value pairs used to construct `P`.

3. Similarly, the Matlab function `X = lhsdesign(n,p)` returns an $n \times p$ matrix, `X`, containing a latin hypercube sample of n values on each of p variables. For each column of `X`, the n values are randomly distributed with one from each interval $(0, 1/n), (1/n, 2/n), ..., (1 - 1/n, 1)$, and they are randomly permuted.

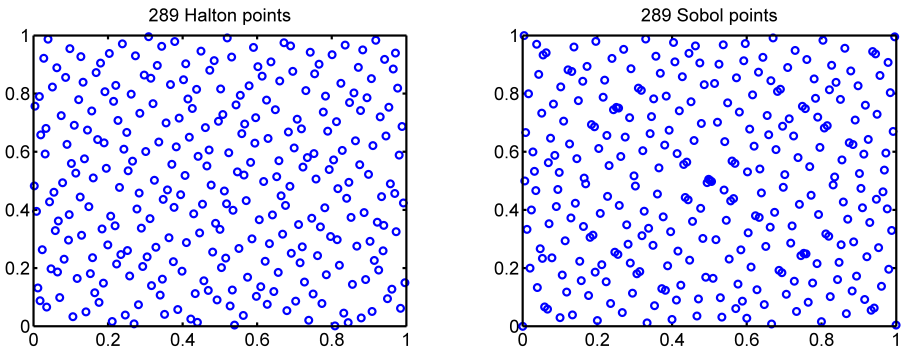


Figure 1.4: Halton and Sobol points

The difference between the standard (tensor product) designs and the quasi-random designs shows especially in higher space dimensions, as shown in Fig. 1.6

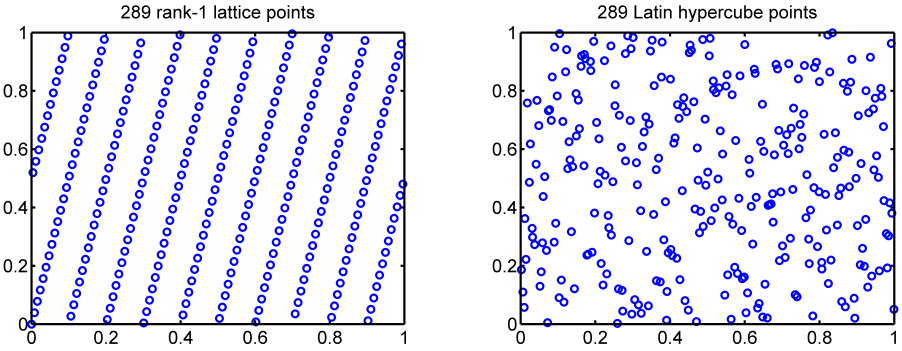


Figure 1.5: Lattice and Latin points

1.4 Exercises

The Haar-Maierhuber-Curtis theorem told us that is not possibile to interpolate by multivaraite polynomials of degree $N \geq 2$ scattered data in dimension $d \geq 2$. This suggested to take a basis of function localized at the so called *data-sites* .

The tools that we need for these exercises are

- The *Halton points* on the hypercube $[0,1]^d$, $d \geq 1$.
- The `fill-distance` (or *mesh size*) $h_{X,\Omega}$ of a set $X \subset \Omega$ con $\Omega \subseteq \mathbb{R}^d$ that is

$$h_{X,\Omega} = \sup_{x \in \Omega} \min_{\mathbf{x}_j \in X} \|x - x_j\|_2 \,, \tag{1.17}$$

  In Matlab, this distance can be easily computed by the command `hX=max(min(DME'))`, where `DME` is the distance matrix, generated by the function `DistanceMatrix.m` evaluated at a set of `target (or evaluation) points` (for example an equispaced grid finer than the of the data-sites `X`).

- The functions to be approximated are

$$f_s(\mathbf{x}) = 4^s \prod_{k=1}^{s} x_k(1-x_k), \quad \mathbf{x} = (x_1, \ldots, x_s) \in [0,1]^d \tag{1.18}$$

$$\texttt{sinc}(\mathbf{x}) = \prod_{k=1}^{s} \frac{\sin(\pi x_k)}{\pi x_k} \,. \tag{1.19}$$

1. By means of the Matlab function `haltonseq.m` compute $N = 5^d$ Halton points in dimensions $d = 1, 2, 3$. Compute for each set the corrisponding `fill-distance`, $h_{X,\Omega}$.
2. Verify graphically the *nested property* of Halton points, that is

$$H_{d,M} \subset H_{d,N}, \quad M < N \,. \tag{1.20}$$

In practice, using different colours plot the Halton points for different values of M and N.

3. Again, by using the function `DistanceMatrix.m` on different set of Halton points of dimension $d = 2$, verify that the corresponding *distance-matrix*, say `A`, is ill-conditioned, by computing its condtion number in the 2-norm (in Matlab `cond(A)`).

4. In dimension $d = 2$, using the function `DistanceMatrixFit.m`, build the RBF interpolant using the basis $\phi_k(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{x}_k\|_2$ (that is, the translates at $\boldsymbol{x}_k$ of the basic function $\phi(r) = r$) of the functions (1.18) and (1.19) by computing the Root Mean Square Error, RMSE (see for its definition Lecture 4). Verify that as N increases, the error decreases. Notice: the RMSE has to be evaluated in a finer grid of evaluation points.

5. Repeat the previous exercise by using the Gaussian radial basis function, $\Phi(\boldsymbol{x}) = \mathtt{e}^{-\epsilon^2\|\boldsymbol{x}\|^2}$, $\epsilon > 0$ again for $d = 2$. For accomplish this interpolation, use the function `RBFInterpolation2D.m` that generalizes the `DistanceMatrixFit.m`.

The Matlab files can be downloaded at the link

http://www.math.unipd.it/∼demarchi/TAA2010

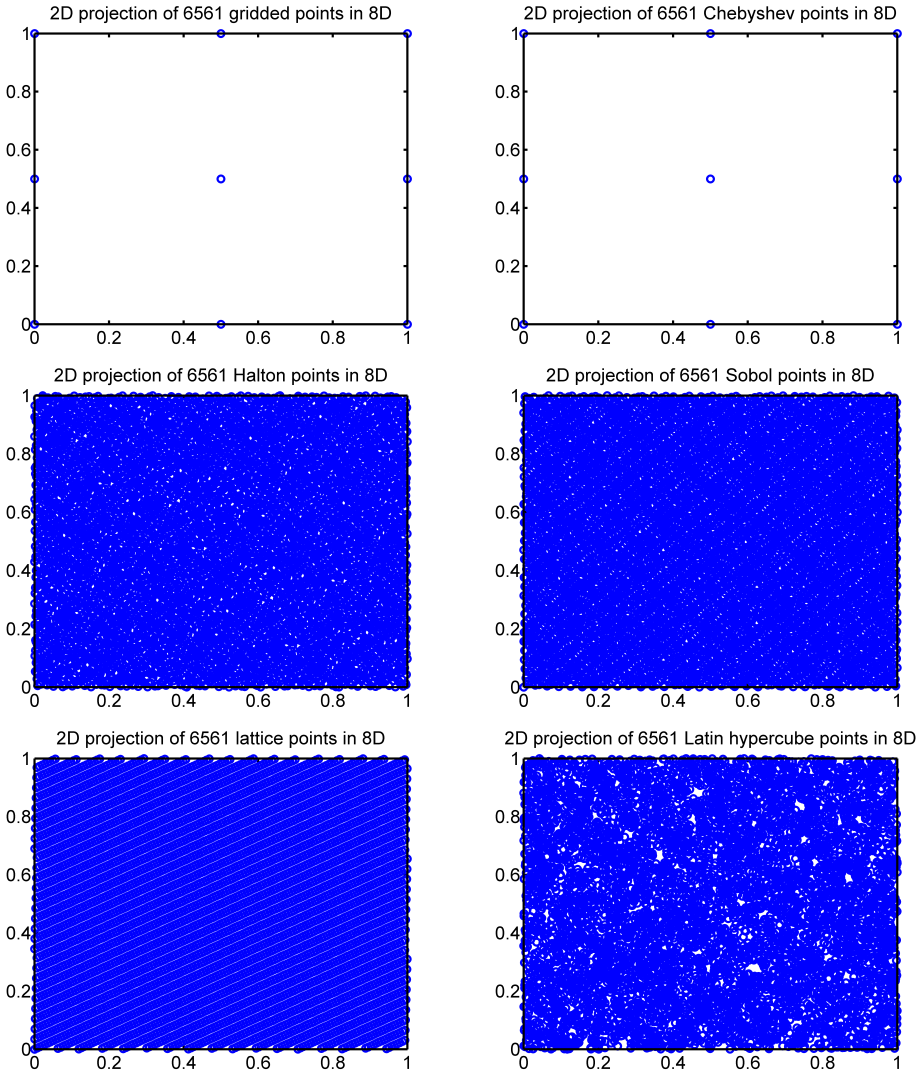


Figure 1.6: Projections in 2D of different point sets

Lecture 2

Positive definite functions

In the first lecture we saw that the scattered data interpolation problem with RBFs leads to the solution of a linear system

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

with $A_{i,j} = \phi(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2)$ and y_i the i-th data value. The solution of the system requires that the matrix A is non-singular. The situation is favourable if we know in advance that the matrix is *positive definite*. Moroever we would like to characterize the class of functions ϕ for which the matrix is positive definite.

2.1 Positive definite matrices and functions

We start by the definition of *positive definite matrices*.

Definition 4. *A real symmetric matrix A is called* `positive semi-definite` *if its associated quadratic form $\boldsymbol{c}^T A\boldsymbol{c} \geq 0$, that is*

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j A_{i,j} \geq 0 \tag{2.1}$$

for $\boldsymbol{c} \in \mathbb{R}^N$. If the quadratic form (2.1) is zero only for $\boldsymbol{c} = \boldsymbol{0}$ then A is called `positive definite` *(sometimes we will use the shorthand notation PD).*

The most important property of such matrices, is that their eigenvalues are positive and so is its determinant. The contrary is not true, that is a matrix with positive determinant is not necessarily positive definite. Just consider as a trivial example the matrix (in Matlab notation) `A=[-1 0 ; 0 -3]` whose determinant is positive, but `A` is not PD.

Hence, if in (1.11) the basis B_k generates a positive definite interpolation matrix that we would always have a well-defined interpolation problem. In order to get such property, we need to introduce the class of *positive definite functions*.

Definition 5. *A continuous complex valued function* $\Phi : \mathbb{R}^d \to \mathbb{C}$ *is called* `positive semi-definite` *if, for all* $N \in \mathbb{N}$, *all sets of pairwise distinct points* $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subset \mathbb{R}^d$ *and* $\boldsymbol{c} \in \mathbb{C}^N$ *the quadratic form*

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i \bar{c}_j \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j) \tag{2.2}$$

is nonnegative. The function Φ *is then called* `positive definite` *if the quadratic form above is positive for* $\boldsymbol{c} \in \mathbb{C}^N, \boldsymbol{c} \neq \boldsymbol{0}$.

Example 4. The function $\Phi(\boldsymbol{x}) = e^{i\boldsymbol{x}^T\boldsymbol{y}}$, for a fixed $\boldsymbol{y} \in \mathbb{R}^d$, is PD on $\mathbb{R}^d$. If fact,

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{N} c_i \bar{c}_j \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j) &= \sum_{i=1}^{N}\sum_{j=1}^{N} c_i \bar{c}_j e^{i(\boldsymbol{x}_i - \boldsymbol{x}_j)^T\boldsymbol{y}} \\
&= \left(\sum_{i=1}^{N} c_i e^{i\boldsymbol{x}_i^T\boldsymbol{y}}\right)\left(\sum_{j=1}^{N} \bar{c}_j e^{-i\boldsymbol{x}_j^T\boldsymbol{y}}\right) \\
&= \left|\sum_{i=1}^{N} c_i e^{i\boldsymbol{x}_i^T\boldsymbol{y}}\right|^2 \geq 0.
\end{aligned}$$

Remark. From the previous definition and the discussion done on Lecture 1, we should use PD functions as basis, i.e. $B_i(\boldsymbol{x}) = \Phi(\boldsymbol{x} - \boldsymbol{x}_i)$, that is an interpolant of the form $P_f(\boldsymbol{x}) = \sum_{i=1}^{N} c_i B_i(\boldsymbol{x})$. Moreover at this point, we do not need P_f be a *radial* function, but simply *translation invariant* (that is, P_f is the same as the translated interpolant to the original data). We will characterize PD and radial functions in $\mathbb{R}^d$ later in this lecture.

$\diamond\diamond$

These functions have some properties that we summarize in the following theorem.

Theorem 4. *Suppose* Φ *is a positive semi-definite function. Then*

1. $\Phi(\boldsymbol{0}) \geq 0$ *and* $\Phi(\boldsymbol{0}) = 0$ *iff* $\Phi \equiv 0$.
2. $\Phi(-\boldsymbol{x}) = \overline{\Phi(\boldsymbol{x})}$ *for all* $\boldsymbol{x} \in \mathbb{R}^d$.
3. $|\Phi(\boldsymbol{x})| \leq \Phi(\boldsymbol{0})$ *for all* $\boldsymbol{x} \in \mathbb{R}^d$ *(boundness)*

4. *If* $\{\Phi_i, i = 1, \dots, m\}$ *are positive semi-definite and* $\alpha_j \geq 0,\ j = 1, ..., m$, *then* $\Phi = \sum_{i=1}^{m} \alpha_j \Phi_j$ *is also positive semi-definite. If one of the* Φ_j *is positive definite and the corresponding coefficient* α_j *is positive, then* Φ *is also positive definite.*

5. *The product of two positive definite functions is positive definite.*

Proof. The proof can be found in [74, p. 65-66] or [32, 29-30]. ■

Remark. From property 2. of Theorem 4, it is clear that a positive semi-definite function is real valued iff is `even`. But we can also restrict to real coefficient vectors $\boldsymbol{c} \in \mathbb{R}^N$ in the quadratic form. In fact holds the following theorem.

Theorem 5. *Suppose* $\Phi : \mathbb{R}^d \to \mathbb{R}$ *is continuous. Then* Φ *is positive definite if and only if* Φ *is even and we have, for all* $N \in \mathbb{N}$ *and all* $\boldsymbol{c} \in \mathbb{R}\backslash\{\boldsymbol{0}\}$ *and all pairwise distinct* $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j) > 0\,.$$

Proof. If Φ is PD and real valued, then it is even by the previous theorem (by property 2.). Letting $c_k = a_k + i\, b_k$ then

$$\begin{aligned}\sum_{i,j=1}^{N} c_i \bar{c}_j \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j) &= \sum_{i,j=1}^{N} (a_i a_j + b_i b_j)\Phi(\boldsymbol{x}_i - \boldsymbol{x}_j) + \\ &+ i \sum_{i,j=1}^{N} a_j b_i [\Phi(\boldsymbol{x}_i - \boldsymbol{x}_j) - \Phi(\boldsymbol{x}_j - \boldsymbol{x}_i)]\,.\end{aligned}$$

As Φ is even, the second sum on the right hand side is zero. The first sum is nonnegative bacause of the assumption, vanishing only if $a_i = b_i = 0$. ■

Example 5. The cosine function is PD on $\mathbb{R}$. In fact, for all $x \in \mathbb{R}$, $\cos x = \frac{e^{ix}+e^{-ix}}{2}$. By property 4. of Theorem 4 and the fact that the exponential is PD (see Example 4), we conclude.

When we are dealing with `radial functions` i.e $\Phi(\boldsymbol{x}) = \phi(\|\boldsymbol{x}\|)$, then it will be convenient to refer to the *univariate* function ϕ as `positive definite radial function`. A consequence of this notational convention is the following Lemma.

Lemma 1. *If* $\Phi(\boldsymbol{x}) = \phi(\|\boldsymbol{x}\|)$ *is PD (or positive semi-definite) and radial on* $\mathbb{R}^d$, *then* Φ *is also PD (or positive semi-definite) and radial on* $\mathbb{R}^\delta$ *for any* $\delta \leq d$.

◇◇

The Schoenberg characterization of PD and radial functions

The classical way to characterize PD and radial functions is by this Theorem, due to Schoenberg [66].

Theorem 6. *A continuous function $\phi : [0, \infty) \to \mathbb{R}$ is PD and radial on $\mathbb{R}^d$ if and only if it is the Bessel transform of a finite nonnegative Borel measure μ on $[0, \infty)$, i.e.*

$$\phi(r) = \int_0^\infty \Omega_d(rt) d\mu(t)$$

where

$$\Omega_d(r) = \begin{cases} \cos(r) & d=1 \\ \Gamma(\frac{d}{2}) \left(\frac{2}{r}\right)^{(d-2)/2} J_{(d-2)/2}(r) & d \geq 2 \end{cases}$$

Here J_ν is the classical Bessel function of the first kind of order ν, that is the solution of the Bessel differential equation. For example, around $x = 0$, this function can be expressed as the Taylor series

$$J_\nu(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\,\Gamma(k+\nu+1)} \left(\frac{x}{2}\right)^{2k+\nu}.$$

In particular, as already seen, $\Phi(x) = \cos(x)$ can be seen as a fundamental PD and radial function on $\mathbb{R}$.

Observation. Since these lectures are intended for people with no preperation and mathematical background on measure theory and functional analysis, we prefer to avoid to introduce more results characterizing PD functions by using (Fourier) transforms. We prefer to use another and more comprehensible approach based on the definition of completely monotone and multiply monotone functions. Interested readers can satisfy their hunger for knowledge by looking at the books [8, 32, 74].

2.1.1 Completely monotone functions

In the framed Theorem 6 and successive remark, we observed that we avoid to characterize the positive definiteness of RBF using Fourier transforms, also because Fourier transforms are not always easy to compute.

Definition 6. *A function $\phi : [0, \infty) \to \mathbb{R}$ that is $\mathcal{C}[0, \infty) \cap \mathcal{C}^\infty(0, \infty)$ and satisfies*

$$(-1)^k \phi^{(k)}(r) \geq 0, \quad r > 0,\ k = 0, 1, 2, \ldots$$

is called `Completely Monotone` *(shortly CM).*

This definition allows to verify when a function ϕ is positive definite and radial for all dimensions d.

Here we enumerate some of the most important PD functions showing that they are CM.

1. The function $\phi(r) = c,\ c \geq 0$ is CM on $[0, \infty)$.
2. The function $\phi(r) = e^{-c\,r},\ c \geq 0$ is CM on $[0, \infty)$ since
$$(-1)^k \phi^{(k)}(r) = c^k e^{-cr} \geq 0,\ k = 0, 1, 2, ...$$
3. The function $\phi(r) = \frac{1}{(1+r)^\beta},\ \beta \geq 0$ is CM on $[0, \infty)$ since
$$(-1)^k \phi^{(k)}(r) = (-1)^{2k} \beta(\beta - 1) \cdots (\beta + k - 1)(1 + r)^{-\beta - k} \geq 0,\ k = 0, 1, 2, ...$$

There are now two Theorems to quote that give more informations

Theorem 7. *A function ϕ is CM on $[0, \infty)$ if and only of $\Phi = \phi(\|\cdot\|^2)$ is positive semi-definite and radial on $\mathbb{R}^d$ for all d.*

Notice that now Φ is defined via the *square* of the norm.

Theorem 8. *A function ϕ is CM on $[0, \infty)$ but not constant if and only of $\Phi = \phi(\|\cdot\|^2)$ is PD and radial on $\mathbb{R}^d$ for all d.*

Hence as a consequence of these two theorems, the functions $\phi(r) = e^{-cr}, c > 0$ and $\phi(r) = \frac{1}{(1+r)^\beta},\ \beta \geq 0$ are CM on $[0, \infty)$ and since they are not constant, in $\mathbb{R}^d$ for all d, we have

1. $\Phi(\boldsymbol{x}) = e^{-c^2\|\boldsymbol{x}\|^2},\ c > 0$ is positive definite and radial on $\mathbb{R}^d$ for all d. This is the family of `Gaussians`. The parameter c is a *shape parameter* that change the shape of the function, making it more spiky when $c \to \infty$ and flatter when $c \to 0$.
2. $\Phi(\boldsymbol{x}) = (1 + \|\boldsymbol{x}\|^2)^{-\beta},\ \beta \geq 0$ is positive definite and radial on $\mathbb{R}^d$ for all d. This is the family of `inverse multiquadrics`.

2.1.2 Multiply monotone functions

This characterization allows to check when a function is PD and radial on $\mathbb{R}^d$ for `some fixed` d.

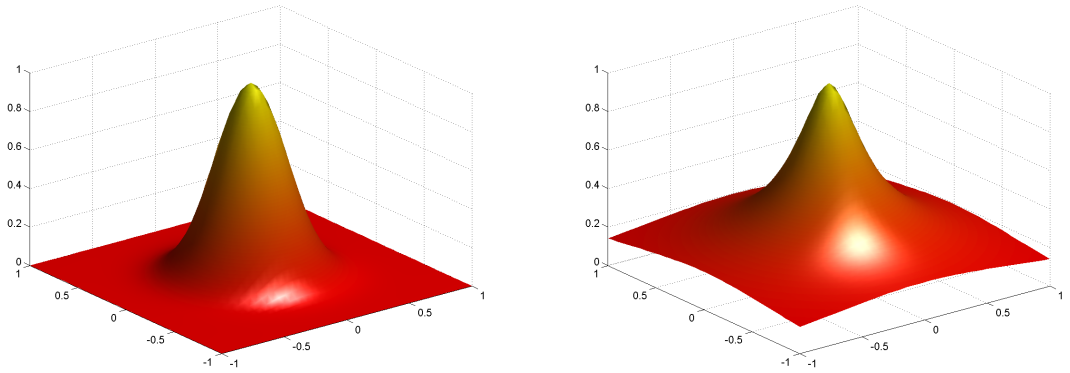


Figure 2.1: Left: Gaussian with $c = 3$. Right: inverse multiquadric with $\beta = 1$.

Definition 7. *A function* $\phi : (0, \infty) \to \mathbb{R}$ *which is* $\mathcal{C}^{s-2}(0, \infty)$, $s \geq 0$ *and for which* $(-1)^k \phi^{(k)}(r) \geq 0$, *non-increasing and convex for* $k = 0, 1, \ldots, s-2$ *is called* `s times monotone` *on* $(0, \infty)$. *In case* $s = 1$ *we only require* $\phi \in \mathcal{C}(0, \infty)$ *to be non-negative and non-increasing.*

In what follows sometimes we use the shorthand notation `MM` for multiply monotone functions.

Convexity can be expressed by the inequality $\phi'' \geq 0$ if ϕ'' exists. Hence, a MM function is essentially a CM function with *truncated monotonicity*.

Example 6. Here we introduce two `new` families of radial basis functions.

1. The `truncated power function`

$$\phi_k(r) = (1 - r)_+^k$$

is k-times monotone for any k. Indeed

$$(-1)^s \phi_k^{(s)} = k(k-1)\cdots(k-s+1)(1-r)_+^{k-s} \geq 0, \quad s = 0, 1, \ldots, k$$

These functions (we will see later) lead to radial functions that are PD on $\mathbb{R}^d$ provided $k \geq \lfloor d/2 \rfloor + 1$.

2. Consider the integral operator I defined as

$$(If)(r) = \int_r^\infty f(t)dt\,, \quad r \geq 0\,,$$

with an f which is k-times monotone. Then If is $(k+1)$-times monotone. This follows from the fundamental theorem of calculus. This operator I plays an important role in the construction of `compactly supported` radial basis functions (that for lack of time we will not discuss in these lectures).

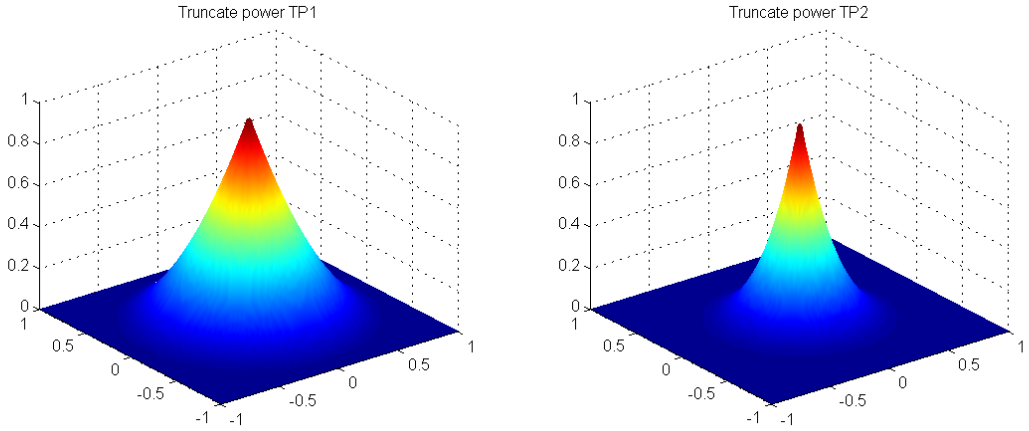


Figure 2.2: Left: truncated power with $k = 1$. Right: truncated power with $k = 2$.

2.1.3 Other positive definite radial functions

1. `Gaussians-Laguerre`. These are defined as

$$\Phi(\boldsymbol{x}) = e^{\|\boldsymbol{x}\|^2} L_n^{d/2}(\|\boldsymbol{x}\|^2)$$

where $L_n^{d/2}$ indicates the Laguerre polynomial of degree n and order $d/2$, that is

$$L_n^{d/2}(t) = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \binom{n + d/2}{n - k} t^k \,.$$

For example for $n = 1, 2$ and $d = 1, 2$ see Table 2.1.

d	$n = 1$	$n = 2$
1	$(3/2 - x^2)\mathrm{e}^{-x^2}$	$(15/8 - 5/2\, x^2 + 1/2x^4)\mathrm{e}^{-x^2}$
2	$(2 - \|\mathbf{x}\|^2)\mathrm{e}^{-\|\mathbf{x}\|^2}$	$(3 - 3\|\mathbf{x}\|^2 + 1/2\|\mathbf{x}\|^4)\mathrm{e}^{-\|\mathbf{x}\|^2}$

Table 2.1: Gaussians-Laguerre functions for different d and n

Notice, that the Gaussians-Laguerre, depends on the space dimension d. Therefore they are PD and radial on $\mathbb{R}^d$ (and therefore also on $\mathbb{R}^\delta$ for $\delta \leq d$).

2. `Poisson` functions. These functions has the definition

$$\Phi(\boldsymbol{x}) = \frac{J_{d/2-1}(\|\boldsymbol{x}\|)}{\|\boldsymbol{x}\|^{d/2-1}}, \quad d \geq 2\,,$$

where J_p is the *Bessel function of the first kind and order p*. While these functions are not defined at the origin, they can be extended to be $\mathcal{C}^\infty(\mathbb{R}^d)$ (see Table 2.2 for

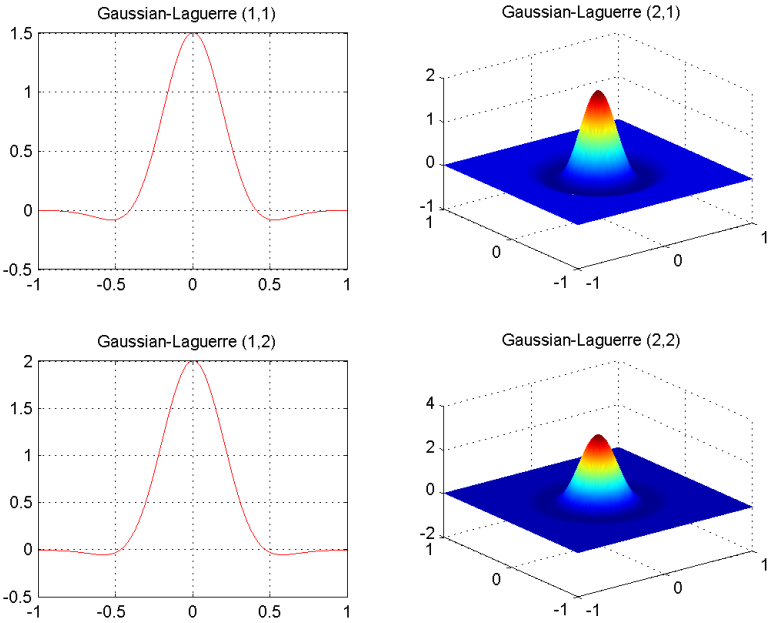


Figure 2.3: First row: G-L when $d = 1$ and $n = 1, 2$. Second row: G-L for $d = 2$ and $n = 1, 2$. In both cases we used the scaling factor $\epsilon = 3$. For the two dimensional case we also normalized so that $\Phi(\mathbf{0}) = 1$.

$d = 2$	$d = 3$	$d = 4$
$J_0(\|\mathbf{x}\|)$	$\sqrt{\frac{2}{\pi}}\frac{\sin(\|\mathbf{x}\|)}{\|\mathbf{x}\|}$	$\frac{J_1(\|\mathbf{x}\|)}{\|\mathbf{x}\|}$

Table 2.2: Poisson functions for various d

the cases $d = 2, 3, 4$.) **Notice:** J_p in Matlab can be computed using the function `besselj(p, z)` (where `z` is an array of evaluation points)

3. `Matérn` functions. They are defined as

$$\Phi(\boldsymbol{x}) = \frac{K_{\beta-d/2}(\|\boldsymbol{x}\|)\|\boldsymbol{x}\|^{\beta-d/2}}{2^{\beta-1}\Gamma(\beta)}, \quad d < 2\beta\,,$$

where K_p is the *modified Bessel function of the second kind* of order p, that can be defined as a function of the Bessel function of first kind as follows

$$K_p(x) = \frac{\pi}{2}\frac{J_{-p}(x) - J_p(x)}{\sin(\pi\, p)}$$

In Table 2.3 we present the Mátern functions for three values of β, indicated as

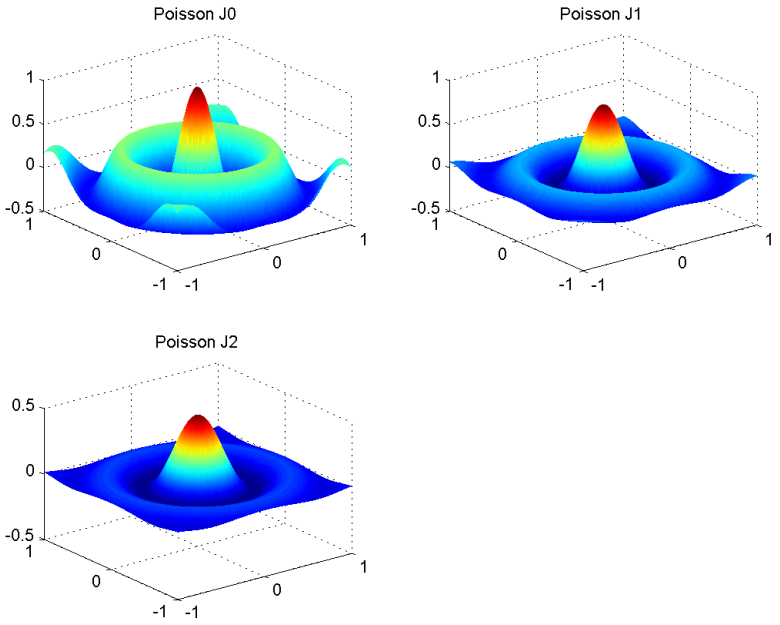


Figure 2.4: Poisson RBF for $d = 2, 3, 4$, respectively. Here the shape parameter is $\epsilon = 10$

β_i, $i = 1, 2, 3$. Notice that the *Matérn* function for β_1 is not differentiable at the origin; while for β_2 is $\mathcal{C}^2(\mathbb{R}^s)$ and for β_3 is $\mathcal{C}^4(\mathbb{R}^s)$.

$\beta_1 = \frac{d+1}{2}$	$\beta_2 = \frac{d+3}{2}$	$\beta_3 = \frac{d+5}{2}$
$\mathrm{e}^{-\|x\|}$	$(1 + \|x\|)\,\mathrm{e}^{-\|x\|}$	$\left(3 + 3\|x\| + \|x\|^2\right)\mathrm{e}^{-\|x\|}$

Table 2.3: Matérn functions for different values of the parameter β

4. The `generalized inverse multiquadrics`

$$\Phi(x) = (1 + \|x\|^2)^{-\beta}, \quad d < 2\beta$$

It is worth to mention that the requirement $d < 2\beta$ implies the dependence of the space dimension. This is required so that $\Phi(x) \in L_1(\mathbb{R}^d)$. As observed in [74, p. 77], this restriction can be relaxed and take $\beta > 0$ leading to a positive definite function in $\mathbb{R}^d$).

5. `Whittaker` functions. The idea of these functions is based on the following construction. Let $f \in \mathcal{C}[0, \infty)$ be a non-negative and not identically equal to zero, and define the function

$$\phi(r) = \int_0^\infty (1 - rt)_+^{k-1} f(t) dt\,. \tag{2.3}$$

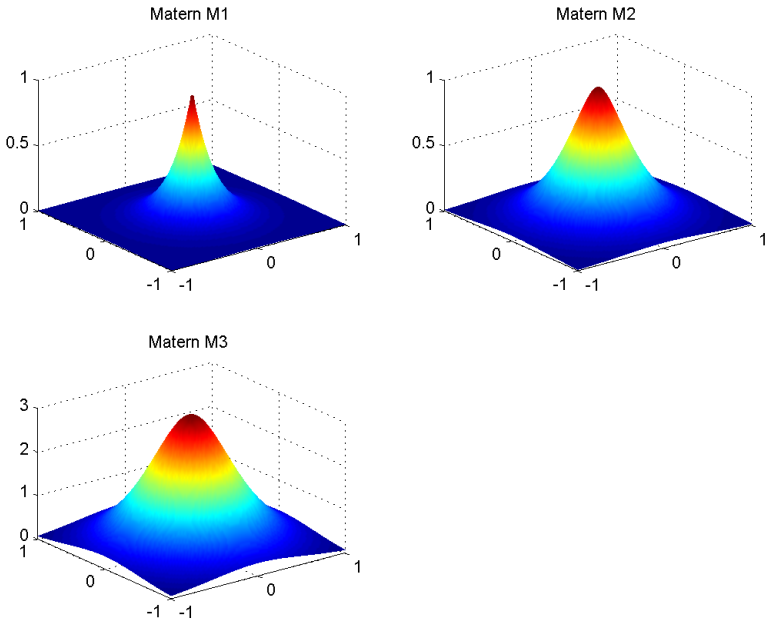


Figure 2.5: Matern RBF for different β, respectively. The shape parameter is $\epsilon = 5$

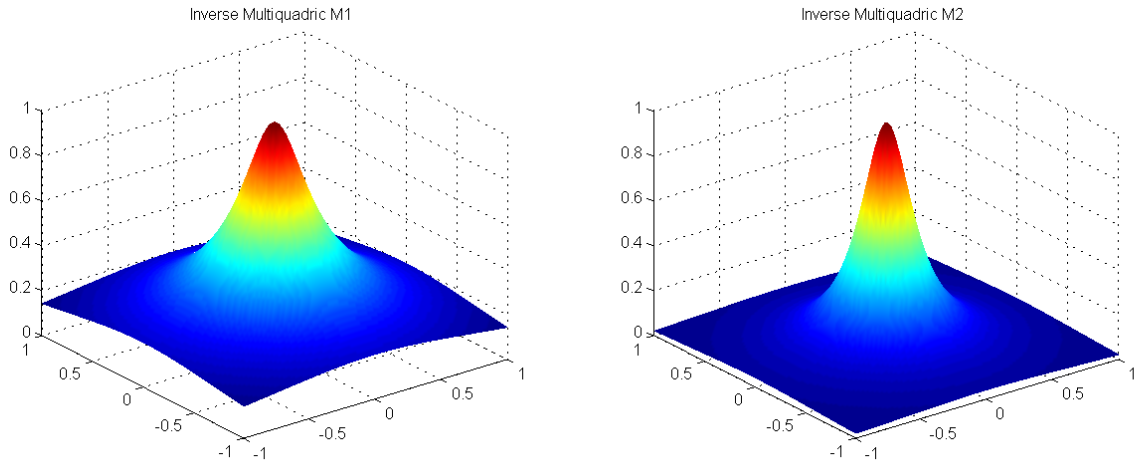


Figure 2.6: Inverse multiquadric with shape parameter $\epsilon = 3$. On the left $\beta = 1/2$ that corresponds to the *Hardy multiquadric*. On the right with $\beta = 1$ that corresponds to the *inverse quadric*.

Then the $\Phi(\boldsymbol{x}) = \phi(\|\boldsymbol{x}\|)$ is positive definite and radial on $\mathbb{R}^d$ provided $k \geq \lfloor \frac{d}{2} \rfloor + 2$.

In fact the quadratic form

$$\sum_{i,j=1}^{N} c_i c_j \phi(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|) = \int_0^\infty \sum_{i,j=1}^{N} c_i c_j \varphi_{k-1}(t\|\boldsymbol{x}_i - \boldsymbol{x}_j\|) f(t) dt$$

which is non-negative since the truncated power $\varphi_{k-1}(\|\cdot\|)$ is positive semi-definite (see above) and f is non-negative. It is easy to see that it is indeed positive definite, and so ϕ is PD itself.

Then the choice of f gives different Whittaker functions. For example, as described in [32, p. 44], taking $f(t) = t^\alpha e^{\beta t}, \alpha \geq 0, \beta > 0$ we get a function with a long expression (see [32, p.44, formula (4.10)]) that uses the so called *Whittaker -M* function.

Special cases when $\beta = 1$ are provided in Table 2.4.

α	$k = 2$	$k = 3$
0	$\frac{\beta - \|x\| + \|x\|\mathrm{e}^{-\beta/\|x\|}}{\beta^2}$	$\frac{\beta^2 - 2\beta\|x\| + 2\|x\|^2 - 2\|x\|^2\mathrm{e}^{-\beta/\|x\|}}{\beta^3}$
1	$\frac{\beta - 2\|x\| + (\beta + 2\|x\|)\mathrm{e}^{-\beta/\|x\|}}{\beta^3}$	$\frac{\beta^2 - 4\beta\|x\| + 6\|x\|^2 - (2\beta\|x\| + 6\|x\|^2)\mathrm{e}^{-\beta/\|x\|}}{\beta^4}$

Table 2.4: Whittaker functions for various choices of k and α

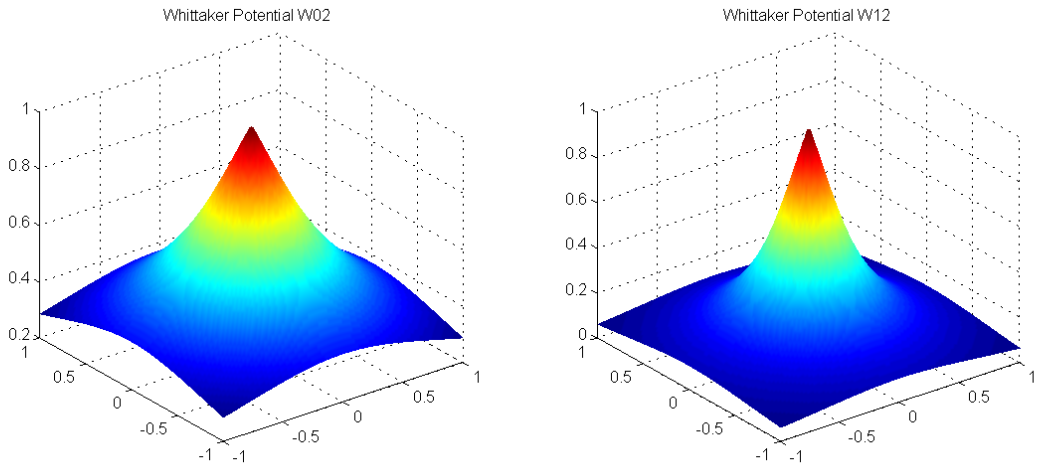


Figure 2.7: Whittaker functions with $\alpha = 0, k = 2$ and $\alpha = 1, k = 2$.

2.2 Exercises

1. Plot some of the radial function positive definite (centered in the origin). When $d = 1$ take $x \in [-1, 1]$ while for $d = 2$ consider $\mathbf{x} \in [-1, 1]^2$.

   - The *Gaussians -Laguerre* for $n = 1, 2$ and $d = 1, 2$. See Table 2.1 for the corresponding definitions.
   - The *Poisson* functions for $d = 2, 3, 4$ in $[-1, 1]^2$ using as shape parameter $\epsilon = 10$, see Table 2.2
   - The *Matérn* function in $[-1, 1]^2$, for three different values of β and shape parameter $\epsilon = 10$, as defined in Table 2.3.
   - The *generalized inverse multiquadrics* $\Phi(x) = (1 + \|x\|^2)^{-\beta}$, $s < 2\beta$, in $[-1, 1]^2$, in these two (simple) cases (with $\epsilon = 5$): $\beta = 1/2$ (which corresponds to the so-called *Hardy inverse multiquadrics*) and $\beta = 1$ (which is the *inverse quadrics*).
   - The *truncated powers* $\Phi(x) = (1 - \|x\|)_+^l$ for $l = 2, 4$ (in $[-1, 1]^2$).
   - The *Whittaker's potentials* in the square $[-1, 1]^2$ for different values of the parameters α, k and β, as defined in Table 2.4. For these last plots use $\beta = 1$.

2. Interpolate the *Franke function*

$$\begin{aligned} f(x_1, x_2) &= .75\exp[-((9x_1 - 2)^2 + (9x_2 - 2)^2)/4] + .75\exp[-(9x_1 + 1)^2/49 - (9x_2 + 1)/10] \\ &+ .5\exp[-((9x_1 - 7)^2 + (9x_2 - 3)^2)/4] - .2\exp[-(9x_1 - 4)^2 - (9x_2 - 7)^2]; \end{aligned}$$

   on a grid of 20×20 Chebyshev points in $[0, 1]^2$ with Poisson and Matérn functions. Compute also the corresponding RMSE.

Lecture 3

Conditionally positive definite functions

The multivariate polynomials are not suitable for solving the scattered data interpolation problem. Only data sites in special locations can guarantee well-posedness of the interpolation problem. In order to have a flavour of this setting, we introduce the notion of *unisolvency.*

Definition 8. *A set $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{R}^d$ is called m-*`unisolvent`*, if the only polynomial of total degree at most m interpolating the zero data on X is the zero polynomial.*

Similarly, X is unisolvent for $\mathbb{P}_m(\mathbb{R}^d)$ (i.e. polynomials in d variables of degree at most m) if there exists a unique polynomial in $\mathbb{P}_m(\mathbb{R}^d)$ of lowest possible degree which interpolates the data X.

Then, to get a unique solution of the polynomial interpolation problem with a d-variate polynomial of degree $\leq m$, to some given data, we need to find a subset of $X \subset \mathbb{R}^d$ with cardinality $M = \binom{m+d}{m} = \texttt{dim}(\mathbb{P}_m(\mathbb{R}^d)$ which is m-unisolvent.

Example 7. These are examples of points that form unisolvent sets

- A simple example is this one. Take $N = 5$ points in $\mathbb{R}^2$. There is no unique way to use bivariate linear interpolation o quadratic. In fact linear polynomials have dimension $M = 3$, bivariate $M = 6$.
- The `Padua points` on the square $[-1, 1]^2$, form the first complete example of unisolvent points whose *Lebesgue constant* has optimal growth (cf. [6]).
- In $\mathbb{R}^2$ the `Chung and Yao points`. The construction of these points is based on lattices. In practise, a lattice $\Lambda \subset \mathbb{R}^d$ has the form $\Lambda = \left\{ \sum_{i=1}^{d} h_i v_i,\ h_i \in \mathbb{Z} \right\}$ with $\{v_i, \ldots, v_d\}$ a basis for $\mathbb{R}^d$. For details see the paper [17].

The m-unisolvency of the set $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ is equivalent to the fact that the matrix P such that

$$P_{i,j} = p_j(x_i), \quad i = 1, \ldots, N, \quad j = 1, ..., M$$

for any polynomial basis p_j, has full (column)-rank. For $N = M$ this is the classical polynomial interpolation matrix.

This observation, can be easily checked when in $\mathbb{R}^2$ we take 3 collinear points: they are not 1-unisolvent, since a linear interpolant, that is a plane through three arbitrary heights at these three collinear points is not uniquely determined. Otherwise such a set is 1-unisolvent.

Remark. This problem arises when we want to construct interpolants with *polynomial precision*. That is, interpolants that reproduce polynomials.

Example 8. Take the function $f(x, y) = (x+y)/2$ on the unit square $[0, 1]^2$. Using Gaussian RBF (with $\epsilon = 6$) interpolate it on a grid of $N = 1089 = 33 \times 33$ uniformly distributed points. This will lead to an interpolant which is not a linear function as is f (i.e. the interpolation does not work).

Hence, instead of using an interpolant of the form $P_f(\boldsymbol{x}) = \sum_{k=1}^{N} c_k e^{-\epsilon^2 \|\boldsymbol{x}-\boldsymbol{x}_k\|^2}$, we use the following

$$P_f(\boldsymbol{x}) = \sum_{i=1}^{N} c_i e^{-\epsilon^2 \|\boldsymbol{x}-\boldsymbol{x}_i\|^2} + \underbrace{c_{N+1} + c_{N+2}x + c_{N+3}y}_{\text{polynomial part}} . \tag{3.1}$$

Having only N interpolation conditions, namely

$$P_f(\boldsymbol{x}_i) = (x_i + y_i)/2, \quad i = 1, \ldots, N\,, \tag{3.2}$$

how can we find the remaining three conditions so that the resulting system will be square? As we shall see later, the solution is

$$\sum_{i=1}^{N} c_i = 0, \ \sum_{i=1}^{N} c_i x_i = 0, \ \sum_{i=1}^{N} c_i y_i = 0\,. \tag{3.3}$$

Combining the interpolation conditions (3.2) with the *side conditions* (3.3), the resulting linear system becomes

$$\begin{bmatrix} A & P \\ P^T & O \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix} \tag{3.4}$$

where the matrix A and the vectors $\boldsymbol{c}$ and $\boldsymbol{y}$ are the same as in Lecture 1; while P is a $N \times 3$ matrix with entries $P_{i,j} = p_j(\boldsymbol{x}_i)$ with $p_1(\boldsymbol{x}) = 1$, $p_2(\boldsymbol{x}) = x$, $p_3(\boldsymbol{x}) = y$. Finally the matrix O is 3×3 of zeros.

The general form of the (3.2) is

$$P_f(\boldsymbol{x}) = \sum_{i=1}^{N} c_i \phi(\|\boldsymbol{x} - \boldsymbol{x}_i\|) + \sum_{k=1}^{M} d_k p_k(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^d. \tag{3.5}$$

where $p_1, \ldots, p_M$ is a basis for the d-variate polynomials of degree $\leq m-1$, whose dimension is $M = \binom{m-1+d}{m-1}$. The side-conditions become

$$\sum_{i=1}^{N} c_i p_j(\boldsymbol{x}_i) = 0, \quad j = 1, \ldots, M, \tag{3.6}$$

ensuring a unique solution for the system

$$\begin{bmatrix} A & P \\ P^T & O \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix} \tag{3.7}$$

where now the matrices and arrays have the following dimensions: A is $N \times N$, P is $N \times M$, O is $M \times M$; **c, y** are $N \times 1$ and **d, 0** are $M \times 1$.

Remark. It seems ill-adapted for the use to formulate the general setting for the polynomial space $\mathbb{P}_{m-1}(\mathbb{R}^d)$ instead, as before, of $\mathbb{P}_m(\mathbb{R}^d)$. The reason will be explained later and, as we will see, it will be quite natural.

3.0.1 Conditionally positive definite matrices and functions

In order to prove that the augmented system (3.7) is non-singular we start with the easiest case of $m = 1$ (in any dimension d). This is equivalent to the reproduction of the polynomials of degree $m - 1 = 0$, i.e. the constants.

Definition 9. *A real symmetric matrix A, $N \times N$, is called* `conditionally positive semi-definite of order` $m = 1$ *if its associated quadratic form $\boldsymbol{c}^T A \boldsymbol{c} \geq 0$ for all $\boldsymbol{c} = (c_1, \ldots, c_N)^T \in \mathbb{R}^N$ that satisfy*

$$\sum_{i=1}^{N} c_i = 0. \tag{3.8}$$

If $\boldsymbol{c} \neq \boldsymbol{0}$ implies strictly inequality, i.e. $\boldsymbol{c}^T A \boldsymbol{c} > 0$, than A is called `conditionally positive of order one`.

Notice: the conditions (3.8) can be viewed as a condition of "ortogonality" w.r..t. constant functions.

The following theorem shows that the augmented system is uniquely solvable

Theorem 9. *Let A be real symmetric of order N that is conditionally postive definite (CPD) of order one. Let $P = (1, \dots, 1)^T$ be an $N \times 1$ array. Then the system*

$$\begin{bmatrix} A & P \\ P^T & O \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ d \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ 0 \end{bmatrix} \tag{3.9}$$

is uniquely solvable

Proof.. Assume that the solution $(\boldsymbol{c}, d)^T \in \mathbb{R}^{N+1}$ is a solution of the homogeneous system, i.e. with $\boldsymbol{y} = \boldsymbol{0}$. We then prove that $\boldsymbol{c} = \boldsymbol{0}$ and $d = 0$. From the first block, multiplying by $\boldsymbol{c}^T$, we have

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

an using the second equation $P^T \boldsymbol{c} = \boldsymbol{c}^T P = 0$ we get that $\boldsymbol{c}^T A \boldsymbol{c} = 0$. Since A is CPD of order 1, by asumption we get $\boldsymbol{c} = \boldsymbol{0}$. Moreover, from the first block, $A\boldsymbol{c} + dP = \boldsymbol{0}$ that implies $d = 0$. ■

Corollary 1. *For reproducing constant functions in $\mathbb{R}^d$, the system to be solved has the form (3.9).*

We are now ready to introduce `conditionally positive definite functions of order` m.

Definition 10. *A continuous function $\Phi : \mathbb{R}^d \to \mathbb{C}$ is said to be* `conditionally positive semi-definite of order` *m in $\mathbb{R}^d$, if*

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_i \bar{c}_j \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j) \geq 0 \tag{3.10}$$

for any N set $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subset \mathbb{R}^d$ of pairwise distinct points, and $\boldsymbol{c} = (c_1, \dots, c_N)^T \subset \mathbb{C}^N$ such that

$$\sum_{k=1}^{N} c_k p(\boldsymbol{x}_k) = 0$$

for any complex-valued polynomial p of degree $\leq m - 1$. The function Φ is then called `conditionally positive definite` *of order m on $\mathbb{R}^d$ if the quadratic form (3.11) vanishes only when $\boldsymbol{c} \equiv \boldsymbol{0}$.*

The first important fact concerning conditionally positive (semi)-definite functions is their `order`. To this aim holds the following important result.

Proposition 3. *A function which is conditionally positive (semi)-definite of order m is also conditionally positive (semi)-definite of any order $s \geq m$. Moreover, a function that is conditionally positive (semi)-definite of order m in $\mathbb{R}^d$ is also conditionally positive (semi)-definite of order m on $\mathbb{R}^k$ with $k \leq d$.*

Hence the natural question is to look for the `smallest possibile order` m. When speak of the order of such class of functions we will always refer to the minimal possible m.

Examples of CPD functions

Here we list the most important conditionally positive definite radial functions

- `Generalized multiquadrics`.

$$\Phi(\boldsymbol{x}) = (1 + \|\boldsymbol{x}\|^2)^{\beta}, \quad \mathbf{x} \in \mathbb{R}^d, \ \beta \in \mathbb{R}\backslash\mathbb{N}_0\,,$$

  which are CPD of order $m = \lceil \beta \rceil$ (and higher).

  Notice that in the definition we have to exclude positive integer values for β, otherwise we are led to polynomials of even degree.

  Two cases: for $\beta = 1/2$ we obtain the well-known `Hardy multiquadric` that is CPD of order 1; for $\beta = 5/2$ we have a function which is CPD of order 3.

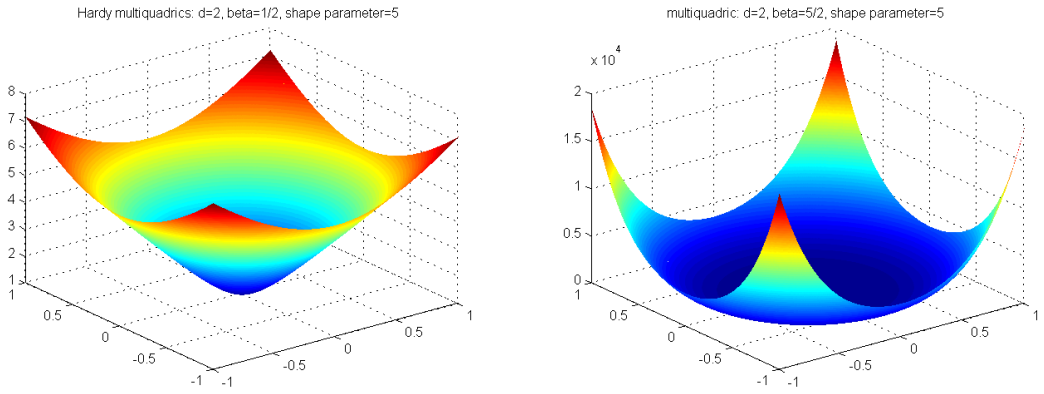


Figure 3.1: Multiquadrics. Left: Hardy multiquadric. Right: with $\beta = 5/2$.

- `Radial powers`

$$\Phi(\mathbf{x}) = \|\mathbf{x}\|^{\beta}, \quad \mathbf{x} \in \mathbb{R}^s, 0 < \beta \notin 2\mathbb{N}\,,$$

  which are CPD of order $m = \lceil \frac{\beta}{2} \rceil$ (and higher).

  For $\beta = 3$ we get a function which is CPD of order 2 and for $\beta = 5$ a function which is CPD of order 3. A property of these functions is that they are *shape parameter free*. This has the advantage that the user need not worry about finding a *good* (or the best) value for ϵ.

- `Thin-plate splines`

$$\Phi(\mathbf{x}) = \|\mathbf{x}\|^{2\beta} \log \|\mathbf{x}\|, \quad \mathbf{x} \in \mathbb{R}^s, \beta \in \mathbb{N}\,,$$

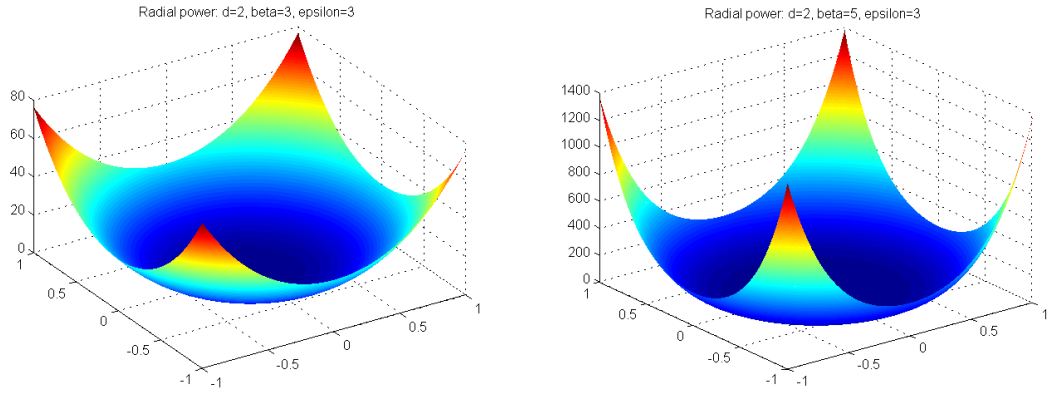


Figure 3.2: Radial powers. Left: with $\beta = 3$, $\epsilon = 3$. Right: with $\beta = 5$, $\epsilon = 3$.

which are CPD of order $m = \beta + 1$.

The classical TPS is for $\beta = 1$ that is CPD of order 2. For $\beta = 2$ we get a CPD function of order 3. Also TPS are *shape parameter free*.

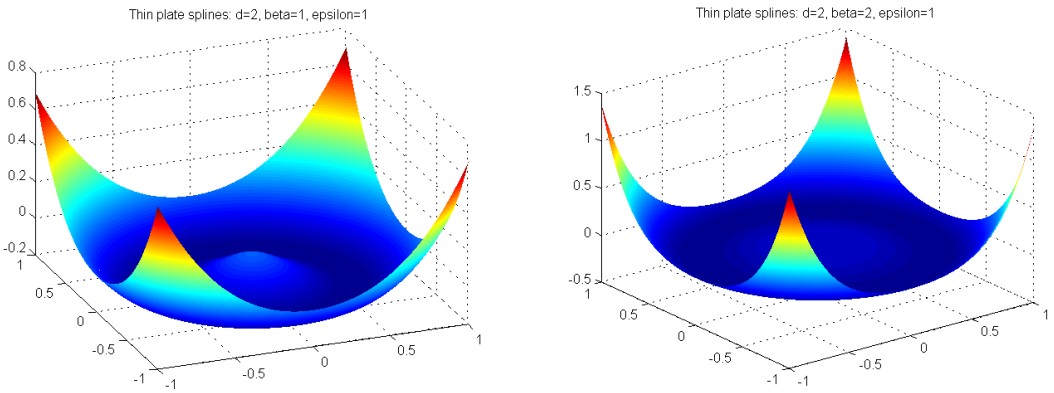


Figure 3.3: Thin plate splines. Left: for $\beta = 1$, $\epsilon = 1$. Right: for $\beta = 2$, $\epsilon = 1$.

◇◇

As for the PD case, the definition reduces to real coefficients and polynomials if the basis function is real-valued and even. This is the case when the function Φ is radial.

Theorem 10. *A continuous function* $\Phi : \mathbb{R}^d \to \mathbb{R}$ *is said to be* `conditionally positive definite of order` m *in* $\mathbb{R}^d$*, if*

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j) > 0 \tag{3.11}$$

for any N set $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subset \mathbb{R}^d$ of pairwise distinct points, and $\boldsymbol{c} = (c_1, \dots, c_N)^T \subset \mathbb{R}^N \setminus \{0\}$ such that

$$\sum_{k=1}^{N} c_k p(\boldsymbol{x}_k) = 0$$

for any real-valued polynomial p of degree $\leq m-1$.

The special case $m = 1$ appers already in the linear algebra literature and it is dealt with $\leq$ and is then referred as *conditionally negative definite*. The constraints are simply $\sum_{i=1}^{N} c_i = 0$. Since the matrix A is CPD of order 1, it is PD in a subspace of dimension $N-1$ (or in general $N-M$ where $M = \dim(\mathbb{P}_{m-1}(\mathbb{R}^d))$. That is only $N-1$ (or $N-M$) eigenvalues are positive.

For conditionally positive definite functions of order $m = 1$ the following important property holds (which is Example 2 of Lecture 1).

Theorem 11. *Suppose Φ is CPD of order 1 and that $\Phi(0) \leq 0$. Then the matrix $A \in \mathbb{R}^{N\times N}$, i.e. $A_{i,j} = \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j)$, has one negative and $N-1$ positive eigenvalues. In particular it is invertible.*

Proof. From the Courant-Fischer theorem, we conclude that it has $N-1$ positive eigenvalues. Now, since $0 \geq N\Phi(0) = \mathtt{tr}(A) = \sum_{i=1}^{N} \lambda_i$, then A must have at least one negative eigenvalue. ■

For example, the generalized multiquadrics $\Phi(\boldsymbol{x}) = (-1)^{\lceil\beta\rceil}(1 + \|\boldsymbol{x}\|^2)^\beta$, $0 < \beta < 1$ (which includes the Hardy's one, $\beta = 1/2$) satisfied the previous Theorem.

Finally we can prove a Theorem similar to Theorem 9.

Theorem 12. *Let A be real symmetric of order N that is conditionally positive definite (CPD) of order m on $\mathbb{R}^d$ and the points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$ form an $m-1$-unisolvent set. Then the system*

$$\begin{bmatrix} A & P \\ P^T & O \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix} \tag{3.12}$$

is uniquely solvable.

Proof.. Assume that the solution $(\boldsymbol{c}, \boldsymbol{d})^T \in \mathbb{R}^{N+M}$ is a solution of the homogeneous system, i.e. with $\boldsymbol{y} = \boldsymbol{0}$. We then prove that $\boldsymbol{c} = \boldsymbol{0}$ and $\boldsymbol{d} = \boldsymbol{0}$. From the first block, multiplying by $\boldsymbol{c}^T$, we have

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

an using the second equation $P^T\boldsymbol{c} = \boldsymbol{c}^T P = \boldsymbol{0}^T$ we get that $\boldsymbol{c}^T A\boldsymbol{c} = 0$. Since A is CPD of order m, by assumption we get $\boldsymbol{c} = \boldsymbol{0}$. The unisolvency of the data sites, i.e. P has

columns linearly independent, and the fact that $\boldsymbol{c} = \mathbf{0}$ guarantee $\boldsymbol{d} = \mathbf{0}$ from the top block $A\boldsymbol{c} + P\boldsymbol{d} = \mathbf{0}$. ■

3.1 Exercises

1. Plot the most important *conditionally positive definite functions* in the square $[-1, 1]^2$.

   (a) `generalized multiquadrics`

   $$\Phi(\mathbf{x}) = (1 + \|\mathbf{x}\|^2)^\beta, \quad \mathbf{x} \in \mathbb{R}^2, \ \beta \in \mathbb{R}\backslash\mathbb{N}_0$$

   in particular for $\beta = 1/2$ we have the Hardy multiquadric which is of (minimum) order 1 and that for $\beta = 5/2$ which is of (minimun) order 3.

   (b) `radial powers`

   $$\Phi(\mathbf{x}) = \|\mathbf{x}\|^\beta, \quad \mathbf{x} \in \mathbb{R}^2, 0 < \beta \notin 2\mathbb{N}$$

   For example take $\beta = 3$ (which is CPD of order 2) and $\beta = 5$ (CPD of order 3). Verify furthermore that the power functions are *shape parameter free*.

   (c) `thin-plate splines`

   $$\Phi(\mathbf{x}) = \|\mathbf{x}\|^{2\beta} \log \|\mathbf{x}\|, \quad \mathbf{x} \in \mathbb{R}^2, \beta \in \mathbb{N}$$

   The most important is the one for $\beta = 1$ which is CPD of order 2. For $\beta = 2$ we get a function which is CPD of order 3. Verify that also these functions are *shape parameter free*.

2. For a CPD function of order 1 (such as the Hardy multiquadric) check the Theorem 11.

Lecture 4

Error estimates

In evaluating the error between the interpolant P_f and the data values at some set $\Xi = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M\} \subset \mathbb{R}^d$ of *evaluation points* we can compute the `root-mean-square error`, that is

$$RMSE := \sqrt{\frac{1}{M}\sum_{j=1}^{M}(P_f(\boldsymbol{\xi}_j) - f(\boldsymbol{\xi}_j))^2} = \frac{1}{\sqrt{M}}\|P_f - f\|_2\,. \tag{4.1}$$

The root-mean-square error (RMSE) is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed. These individual differences are called *residuals* when the calculations are performed over the data sample that was used for estimation, and are called *prediction errors* when computed out-of-sample. The RMSE serves to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power. RMSE is a good measure of accuracy, but only to compare forecasting errors of different models for a particular variable and not between variables, as it is scale-dependent. In practice, (4.1) is simply a quantitative error estimator.

Our goal is to provide error estimates for scattered data interpolation with (conditionally) positive definite functions. We start by considering the PD case.

4.1 Fill distance

The measure that is always used in approximation theory is the *fill-distance* or *mesh size*

$$h = h_{X,\Omega} := \sup_{\boldsymbol{x}\in\Omega}\min_{\boldsymbol{x}_i\in X}\|\boldsymbol{x} - \boldsymbol{x}_j\|_2\,, \tag{4.2}$$

which represents how well the data in X fill out the domain Ω and corresponds to the radius of the largest empty ball that can be placed among the data sites inside Ω.

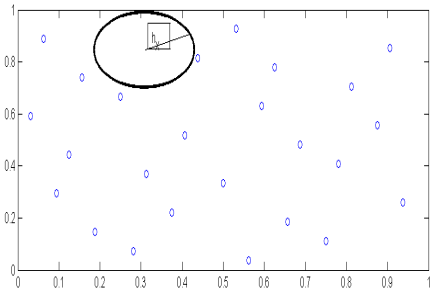


Figure 4.1: The fill distance of 25 Halton points $h \approx 0.2667$

In Matlab the fill distance can be determined by the line `h=max(min(DM_eval'))`, where `DM_eval` is the matrix consisting of the mutual distances between the evaluation points (for example a uniform grid in Ω) and the data set X. In Fig. 4.1 we show 25 Halton points and the corresponding fill distance computed on a grid of 11×11 evaluation points of $[0,1]^2$.

What we want to analysize, is whether the error $\|f - P_f^{(h)}\|_\infty \to 0$ as $h \to 0$, here $P_f^{(h)}$ indicates the interpolant depending on the fill-distance h. To understand the speed of convergence to zero, one has to understand the so-called *approximation order* of the interpolation process.

Definition 11. *We say that the process has* `approximation order` k *if*

$$\|f - P_f^{(h)}\|_p = \mathcal{O}(h^k), \quad k \to 0$$

here the norm is taken for $1 \leq p \leq \infty$.

For completeness, another quantity often used for stability analysis purposes (or finding "good" interpolation points) is the *separation distance* of the data sites X

$$q_X := \frac{1}{2} \min_{i \neq j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$$

which represents the radius of the largest ball that can be placed around every point of X such there are no overlaps. In Matlab `qX=min(min(DM_data+eye(size(DM_data))))/2`, where `DM_data` is the matrix of distances among the data sites. We added the identity matrix in order to avoid that this minimum will be 0, which is the separation distance of every points from itself.

Finally, the ratio

$$\rho_{X,\Omega} := \frac{q_X}{h_{X,\Omega}}$$

known as *uniformity*, which can be identified as $\rho_{X,\Omega} = \lim_{Y \in X_\Omega} \rho_{X,\Omega}$ among all point sets $Y \in X_\Omega$ with X_Ω consisting, in some cases, of the Voronoi vertices used to decompose $\mathbb{R}^d$ with Voronoi tiles. Therefore if $\rho_{X,\Omega} \approx 1$ the data points are nearly equispaced in the Euclidean norm. It has been proved in [25] that, in contrast with polynomial interpolation, radial basis interpolants behaves better when the points are nearly equispaced,

4.2 Lagrange form of the interpolant

This idea goes back to Wu and Schaback [78] and consists in expressing the interpolant by means of *cardinal functions* as in the polynomial case.

Instead of solving the system $A\boldsymbol{c} = \boldsymbol{y}$ as we did in the previous lectures, we consider the (new) system

$$A\boldsymbol{u}^*(\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x}) \tag{4.3}$$

where A is the $N \times N$ positive definite matrix (invertible!) $A_{i,j} = \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j)$, $i, j = 1, \dots, N$, $\boldsymbol{u}^* = (u_1^*, \dots, u_N^*)^T$, with the $u_j^*(\boldsymbol{x}_i) = \delta_{i,j}$ (i.e. cardinal functions), and $\boldsymbol{b} = (\Phi(\cdot - \boldsymbol{x}_1), \dots, \Phi(\cdot - \boldsymbol{x}_N))^T$.

Why this is possible? Thanks to the following theorem

Theorem 13. *If Φ is a positive definite kernel on $\mathbb{R}^d$. Then, for any set of distinct points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$, there exist functions $u_j^* \in \operatorname{span}\{\Phi(\cdot - \boldsymbol{x}_j),\ j = 1, \dots, N\}$ such that $u_j^*(\boldsymbol{x}_i) = \delta_{i,j}$.*

Here we enumerate a few facts in analogy with univariate fundamental Lagrange polynomials.

- The u_j^* can be determined as a ratio of determinants as for the fundamental Lagrange polynomials. Letting $V = \det(u_j^*(\boldsymbol{x}_i))$, and $V_{i,\boldsymbol{x}}$ the same determinant when $\boldsymbol{x}_i$ is substituted by a general point $\boldsymbol{x} \in \Omega$, then $u_i^*(\boldsymbol{x}) = V_{i,\boldsymbol{x}}/V$.
- The u_j^* do not depend on the data values (i.e. the f_j). In fact, once the data sites X and the kernel Φ are chosen, then they can be determined by solving the system (4.3). That is, they do depend on the data sites and the kernel.
- An important aspect, related to stability issues, is the choice of the data points. As proved in [25], the *quasi-uniform* points are always a good choice for RBF interpolation.

4.3 The power function

Another ingredient for understanding the error estimates is the *power function*. The starting point to define this function is the following quadratic form

$$Q(\boldsymbol{u}) = \Phi(\boldsymbol{0}) - 2\boldsymbol{u}^T\boldsymbol{b} + \boldsymbol{u}^T A\boldsymbol{u} \tag{4.4}$$

where the vector $\boldsymbol{b} \in \mathbb{R}^N$ is defined as in the previous section and $\boldsymbol{u}$ is any N dimensional vector.

Definition 12. *On $\mathbb{R}^d$ let us consider a subset Ω and a continuous kernel Φ which we assume PD. For any set $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subset \Omega$ the* `power function` *is defined as follows*

$$P_{\Phi,X}(\boldsymbol{x}) = \sqrt{Q(\boldsymbol{u}^*(\boldsymbol{x}))}\,, \tag{4.5}$$

where $\boldsymbol{u}^*$ is the vector of the cardinal functions in Theorem 13.

Using formula (4.4), combined with the system that defines the cardinal functions (4.3), we get two alternatives ways to compute the power function

$$P_{\Phi,X}(\boldsymbol{x}) = \sqrt{\Phi(\mathbf{0}) - (\boldsymbol{u}^*(\boldsymbol{x}))^T \boldsymbol{b}(\boldsymbol{x})} = \sqrt{\Phi(\mathbf{0}) - (\boldsymbol{u}^*(\boldsymbol{x}))^T A \boldsymbol{u}^*(\boldsymbol{x})}\,, \quad (\texttt{first}) \tag{4.6}$$

$$P_{\Phi,X}(\boldsymbol{x}) = \sqrt{\Phi(\mathbf{0}) - (\boldsymbol{b}(\boldsymbol{x}))^T A^{-1} \boldsymbol{b}(\boldsymbol{x})}\,, \quad (\texttt{second}). \tag{4.7}$$

Notice that when Φ is a PD kernel then A is, therefore we get immediately the following bounds:

$$0 \le P_{\Phi,X}(\boldsymbol{x}) \le \sqrt{\Phi(\mathbf{0})}\,.$$

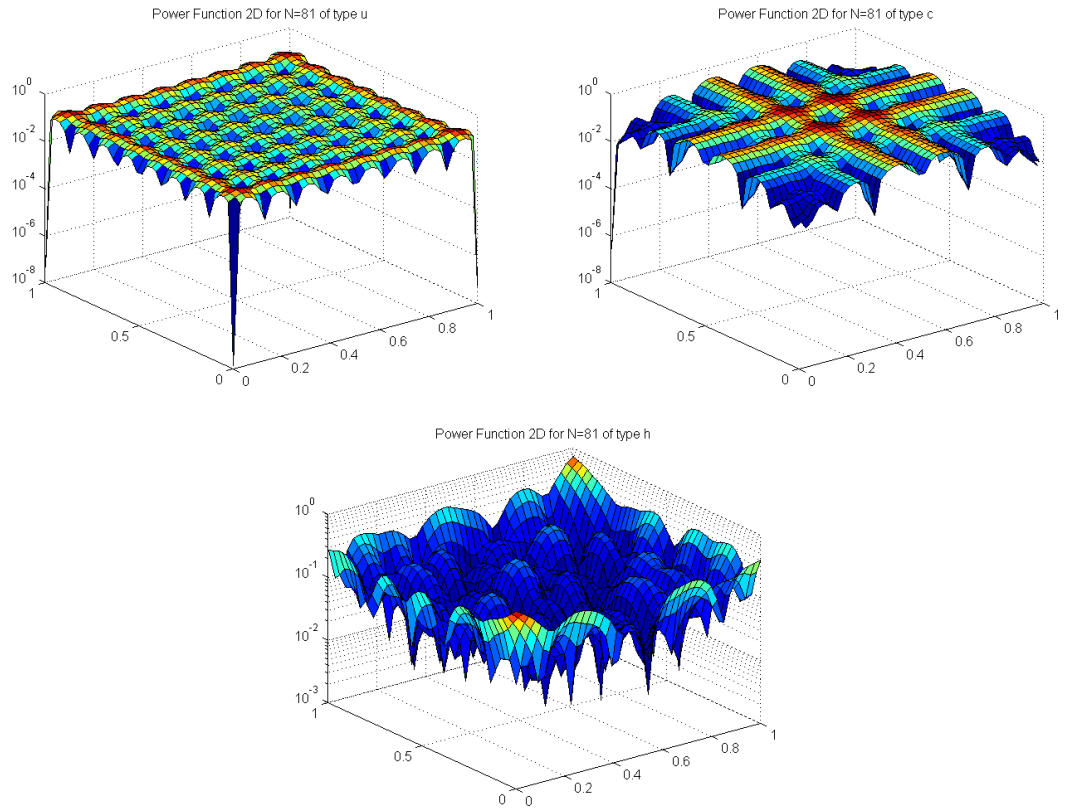


Figure 4.2: Power function for the Gaussian kernel with $\epsilon = 6$ on a grid of 81 uniform, Chebyshev and Halton points, respectively.

An interesting characterization of the power function is given in the next Theorem.

Theorem 14. *Let $\Omega \subset \mathbb{R}^d$ and Φ a PD kernel on $\mathbb{R}^d$. Let X as usual a set of N pairwise distinct points in Ω. The minimum of the quadratic form $Q(\boldsymbol{u})$ is when $\boldsymbol{u} = \boldsymbol{u}^*(\boldsymbol{x})$, that is*

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

Proof. Consider the formula (4.4), the minimum of this quadratic form is given by the solution of the linear system $A\boldsymbol{u} = \boldsymbol{b}(\boldsymbol{x})$ which, however yields the cardinal functions $\boldsymbol{u} = \boldsymbol{u}^*(\boldsymbol{x})$. ■

The power function, by definition, is a positive function, vanishing at the data sites, decreasing to zero as the number of data points increases. Therefore, if we take two data sets such that $Y \subset X$ then $P_{Y,\Omega} \geq P_{X,\Omega}$. This is referred as the *maximality property* of the power function.

As a final remark, the power function is defined similarly for conditionally positive definite functions.

4.4 Native space

The error bounds come rather naturally once we associate with each radial basic function a certain space of functions called `native space`. This space in connected to the so called *Reproducing Kernel Hilbert Space (RKHS)*. The theory of RKHS is beyond our aims, but for understanding a little better the error estimates that we will present, it is necessary to introduce some very basic notions of RKHS.

Definition 13. *A space of functions is called an* `Hilbert space` *if it is a real or complex inner product space that is also a complete metric space w.r.t. the distance induced by the inner product.*

Here the `inner product` between two functions f and g is thought as $(f, g) = \displaystyle\int_a^b f(x)g(x)dx$, in the real case or $(f, g) = \displaystyle\int_a^b f(x)\overline{g(x)}dx$, in the complex case, which has many of the familiar properties of the Euclidean (discrete) dot product.

Examples of Hilbert spaces are: any finite dimensional inner product space (for example $\mathbb{R}^n, \mathbb{C}^n$ equipped with the dot product of two vectors); the *Lebesgue spaces L^p*, *Sobolev spaces*. The space $\mathcal{C}([a, b])$ is an incomplete product space dense in $L^2([a, b])$ which is complete.

Definition 14. *Let H be a real Hilbert space of functions $f : \Omega \to \mathbb{R}$ with inner product $(\cdot, \cdot)_H$. A function $K : \Omega \times \Omega \to \mathbb{R}$ is called a RKHS for H if*

(i) $K(\cdot, \boldsymbol{x}) \in H$ for all $\boldsymbol{x} \in \Omega$

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

The second is the *reproducing property*. In particular, if $f = K$ then we get the kernel K since $K(\boldsymbol{x}, \boldsymbol{y}) = (K(\cdot, \boldsymbol{y}), K(\cdot, \boldsymbol{x}))_H$ for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$.

Notice: the RKHS is known to be unique and that the kernel K is positive definite.

We now show that *"every positive definite radial basis function can be associated with a RKHS: its* `native space`*."*

From Definition 14, a space H should contains functions of the form

$$f(\cdot) = \sum_{i=1}^{N} c_i K(\cdot, \boldsymbol{x}_i)$$

provided $\boldsymbol{x}_j \in \Omega$. Moreover

$$\begin{aligned} \|f\|_H^2 &= (f, f)_H = (\sum_{i=1}^{N} c_i K(\cdot, \boldsymbol{x}_i), \sum_{j=1}^{N} c_j K(\cdot, \boldsymbol{x}_j))_H \\ &= \sum_{i,j=1}^{N} c_i c_j (K(\cdot, \boldsymbol{x}_i), K(\cdot, \boldsymbol{x}_j))_H \\ &= \sum_{i,j=1}^{N} c_i c_j K(\boldsymbol{x}_i, \boldsymbol{x}_j)\,. \end{aligned}$$

Hence we define this (infinite dimensional) space

$$H_K(\Omega) := \operatorname{span}\{K(\cdot, \boldsymbol{y}) : \ \boldsymbol{y} \in \Omega\} \tag{4.8}$$

with associated a *bilinear form*

$$(\sum_{i=1}^{N_K} c_i K(\cdot, \boldsymbol{x}_i), \sum_{j=1}^{N_K} d_j K(\cdot, \boldsymbol{y}_j))_K = \sum_{i,j=1}^{N_K} c_i d_j K(\boldsymbol{x}_i, \boldsymbol{y}_j)$$

where $N_k = \infty$ is also possible.

The last observation is that, this bilinear form defines an inner product on $H_K(\Omega)$, so that $H_k(\Omega)$ is a *pre-Hilbert space*, that means that it is not complete.

The *native space* for the kernel K, indicated as $\mathcal{N}_K(\Omega)$ (or if not confusion arises, simply $\mathcal{N}$), is the completition of $H_K(\Omega)$ w.r.t. the K-norm $\|\cdot\|_K$, so that $\|f\|_K = \|f\|_{\mathcal{N}}$ for all $f \in H_K(\Omega)$. For details, please refer to the book by Wendland [74].

Remark. In dealing with positive definite (translation invariat) functions Φ, we will write

$$\Phi(\boldsymbol{x}-\boldsymbol{y}) = K(\boldsymbol{x},\boldsymbol{y})\,,$$

so that K is taken in $\Omega\times\Omega$ instead of simply Ω.

4.5 Generic error estimates

We quote here two results that gives a flavour of the topic. The first theorem gives a pointwise estimate based on the power function, whose proof can be found in [32, p. 117-118] while the second uses the fill-distance (the proof is again in [32, p. 121-122]).

Theorem 15. *Let $\Omega\subset\mathbb{R}^d$ and $\Phi\in\mathcal{C}(\Omega\times\Omega)$ be PD on $\mathbb{R}^d$. Let $X=\{\boldsymbol{x}_1,\dots,,\boldsymbol{x}_n\}$ be a set of distinct points. Take a function $f\in\mathcal{N}_\Phi(\Omega)$ and denote with P_f its interpolant on X. Then, for every $\boldsymbol{x}\in\Omega$*

$$|f(\boldsymbol{x})-P_f(\boldsymbol{x})| \le P_{\Phi,X}(\boldsymbol{x})\|f\|_{\mathcal{N}_\Phi(\Omega)}\,. \tag{4.9}$$

where the norm of f is the native space norm.

We can express such error by means of the *fill distance*.

Theorem 16. *Let $\Omega\subset\mathbb{R}^d$ and $\Phi\in\mathcal{C}^{2k}(\Omega\times\Omega)$ be symmetric and positive definite. Let $X=\{\boldsymbol{x}_1,\dots,,\boldsymbol{x}_n\}$ be a set of distinct points. Take a function $f\in\mathcal{N}_\Phi(\Omega)$ and its interpolant P_f on X. Then, there exist positive constants h_0 and C (independent of $\boldsymbol{x}$, f and Φ), with $h_{X,\Omega}\le h_0$, such that*

$$|f(\boldsymbol{x})-P_f(\boldsymbol{x})| \le Ch_{X,\Omega}^k\sqrt{C_\Phi(\boldsymbol{x})}\|f\|_{\mathcal{N}_\Phi(\Omega)}\,. \tag{4.10}$$

and

$$C_\Phi(\boldsymbol{x}) = \max_{|\beta|=2k}\ \max_{\boldsymbol{w},\boldsymbol{z}\in\Omega\cup B(\boldsymbol{x},c_2h_{X,\Omega})} |D_2^\beta\Phi(\boldsymbol{w},\boldsymbol{z})|\,.$$

Comparing (4.9) and (4.10) we get a bound of the power function in terms of the fill-distance

$$P_{\Phi,X}(\boldsymbol{x}) \le Ch_{X,\Omega}^k\sqrt{C_\Phi(\boldsymbol{x})}\,.$$

Moreover the Theorem 16 says that interpolation with a $\mathcal{C}^{2k}$ kernel Φ has approximation order k. This means that for kernels infinitely smooth, such as Gaussians, Laguerre-Gaussians, Poisson and generalized inverse multiquadrics, the bound above is arbitrarely high. On the contrary, Matérn, Whittaker radial functions have approximation order limited by the smoothness of the basic function ϕ.

One more observation is that the above estimates consider $f\in\mathcal{N}_{\Phi(\Omega)}$. There exist similar estimates for $f\notin\mathcal{N}_{\Phi(\Omega)}$ (see e.g. [32, §15.3]).

4.6 Strategies in reducing the interpolation error

This last section aims to give some insights to the problem of the choice of shape parameter ϵ in order to get the smallest (possible) interpolation error. In the recent literature there have been exploited various strategies. Here, we present only three of them, which turn out to be indeed the most used by practictioners.

In all that follows, we assume to use the same kernel Φ, we use *only one value* ϵ to scale all basis functions uniformly. The number of data points could be changed by comparison of results.

4.6.1 Trial and Error

It is the simplest approach. It consists in performing various interpolation experiments with different values of the shape parameter. The "best" parameter, say ϵ^* will be the one that minimize the interpolation error. In Figure 4.3 we plot the interpolation max-error varying ϵ for different data points, using the Gaussian kernel in the univariate case. The minimum of every curve gives the "optimal" value.

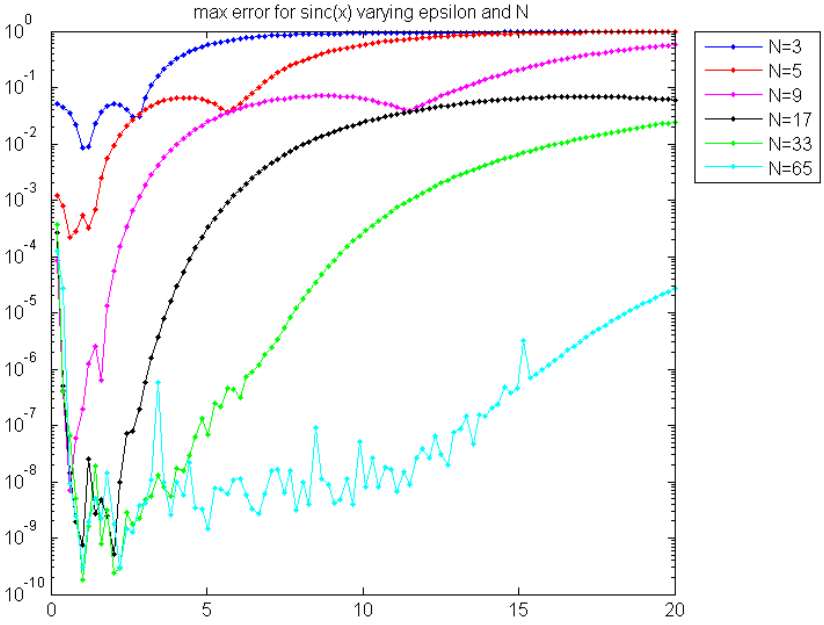


Figure 4.3: Trial and error strategy for the interpolation of the 1-dimensional sinc function by the Gaussian for $\epsilon \in [0, 20]$, taking 100 values of ϵ and for different equispaced data points.

4.6.2 Power function

This is simply connected to the error analysis presented in the previous section (cf. formula (4.9)). Once we have decided which Φ and data set X to use, we calculate the power function on scaled version of the kernel in order to optimize the error component that is independent of f. This approach is similar to the Trial and Error strategy and has the limit to forget the second part of the error, i.e. the one that depends on the basis function via the native space norm of f. In Figure 4.5 we plot the sup-norm of the power function for 500 values of $\epsilon \in [0, 20]$ for the Gaussian kernel and the set of uniform data points X with $N = 9, 25, 81, 289$. This strategy is implemented in the M-file `Powerfunction2D.m` in the Matlab codes provided in [32].

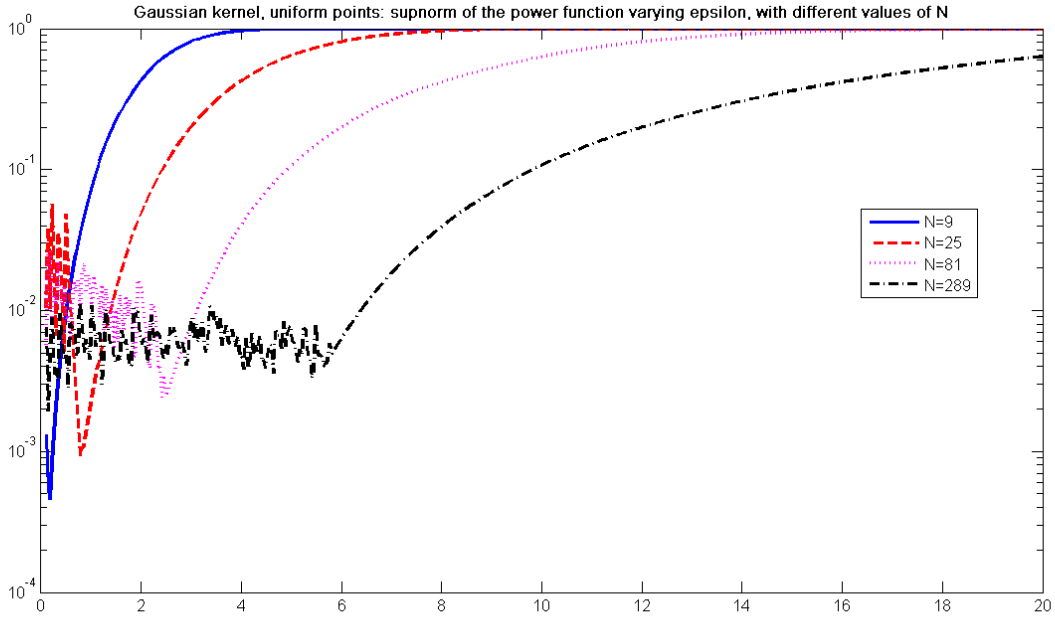


Figure 4.4: The sup-norm of the 2D power function on uniform points for the Gaussian kernel, varying $\epsilon \in [0, 20]$ for different values of N

4.6.3 Cross validation

This method is popular in the statistics literature, known in the case the 2-norm is used as *PRESS* (Predictive REsidual Sum of Squares). The "optimal" value ϵ^* is obtained minimizing the (least-squares) error for a fit of the data values based on an interpolant for which one of the centers is *left out*. In details, we start by constructing $P_{f,k}$, the radial basis

function interpolant to the data $\{f_1, \dots, f_{k-1}, f_{k+1}, \dots, f_N\}$, that is

$$P_{f,k}(\boldsymbol{x}) = \sum_{i=1, i\neq k}^{N} c_{i,k}\Phi(\boldsymbol{x} - \boldsymbol{x}_i)$$

such that

$$P_{f,k}(\boldsymbol{x}_i) = f_i, \quad i = 1, \dots, k-1, k+1, \dots, N\,.$$

Then we compute the error at the point $\boldsymbol{x}_k$, the one not used by $P_{f,k}$

$$E_k = f_k - P_{f,k}(\boldsymbol{x}_k)\,.$$

Then the "quality" of the fit is determined by the (sup) norm of the vector $E = [E_1, \dots, E_N]^T$.

In the experiments, people add a loop on ϵ in order to compare the error norms for different values of the shape parameter, choosing for ϵ^* the one that yields the minimum error norm. This method in general is quite expensive from the computation point of view (it has a complexity of $\mathcal{O}(N^4)$ flops). There is a way to "accelerate" the method, by computing E_k as

$$E_k = \frac{c_k}{A^{-1}_{k,k}}$$

where c_k is the k-th coefficient in the interpolant P_f based on *all* the data points and $A^{-1}_{k,k}$ is the k-th diagonal element of the inverse of the corresponding collocation matrix. Since both c_k and A^{-1} will be computed once for each value of ϵ, this results in $\mathcal{O}(N^3)$ flops. This strategy is implemented in the M-file `LOOCV.m` in the Matlab codes provided in [32].

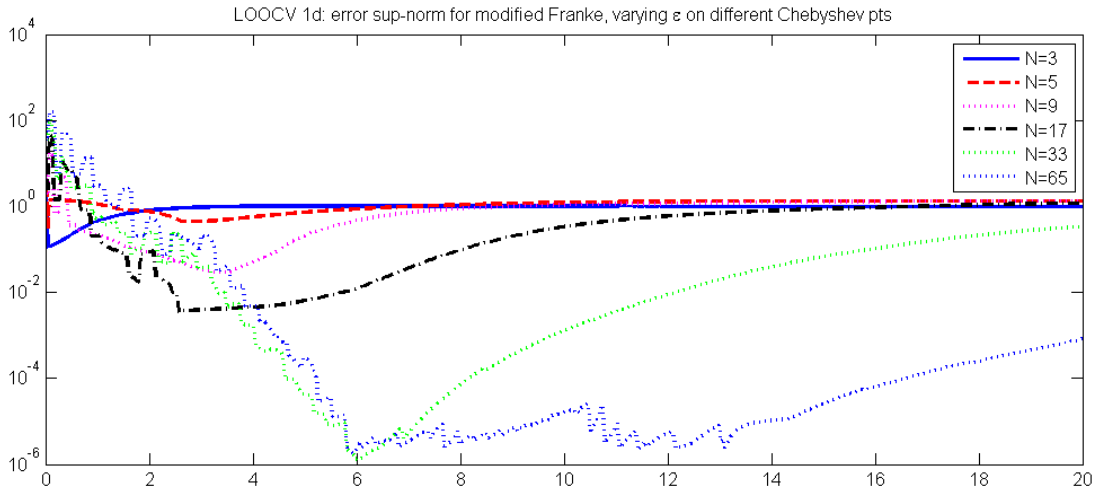


Figure 4.5: LOOCV 1d: sup norm of the error on Chebyshev points for the Gaussian kernel, varying $\epsilon \in [0, 20]$ for different values of N

4.7 Exercises

1. Find the *optimal shape parameter*, ϵ_{opt}, by means of the *trial & error* strategy for the following univariate functions:

   (a)
   $$f_1(x) = \texttt{sinc}(x) = \frac{\sin \pi x}{\pi x}\,.$$

   (b) *variant of the Franke function*
   $$f_2(x) = \frac{3}{4}\left(\texttt{e}^{-(9x-2)^2/4} + \texttt{e}^{-(9x+1)^2/49}\right) + \frac{1}{2}\texttt{e}^{-(9x-7)^2/4} - \frac{1}{10}\texttt{e}^{-(9x-4)^2}\,,$$

   For each of the f_i, $i = 1, 2$ produce a table of the form

N	$\|P_{f_i} - f_i\|_\infty$	ϵ_{opt}
3		
5		
9		
17		
33		
65		

   where, for each N, ϵ_{opt} corresponds to the minimum of the error curve in the sup-norm, computed by varying the shape parameter $\epsilon \in [0, 20]$. As radial basis function for the P_{f_i} take the Gaussian.

2. Plot the *power function* in 2-dimension for the *Gaussian kernel* with $\epsilon = 6$ on a grid of $N = 9^2 = 81$ equispaced, Chebyshev and Halton points in $[-1, 1]^2$. We'll see that the power function will depend on the chosen points.

   Verify that $P_{X,\Omega}(\boldsymbol{x}_i) = 0$ for all $\boldsymbol{x}_i \in X$. Show how varies the maximum value of the power function as N increases.

   Use the M-file `Powerfunction2D.m`.

3. Plot $\|P_{\Phi,X}\|_\infty$ for the Gaussian kernel in 2-dimensions, by using for the power function the formula
   $$P_{\Phi,X}(\boldsymbol{x}) = \sqrt{\Phi(\boldsymbol{0}) - (\boldsymbol{b}(\boldsymbol{x}))^T A^{-1}\,\boldsymbol{b}(\boldsymbol{x})}$$
   with A representing the interpolation matrix and $\boldsymbol{b}(\boldsymbol{x}) = [\Phi(\boldsymbol{x}-\boldsymbol{x}_1), \cdots, \Phi(\boldsymbol{x}-\boldsymbol{x}_N)]^T$, by varying $\epsilon \in [0, 20]$, and $N = 9$, 25, 81, 289. Take equispaced points both as centers and evaluation points of the power function.

   Make a table similar to the one of the previous exercise adding one column for the condition number of A corresponding to the "optimal" shape parameter. Use the function `Powerfunction2D.m` for computing the 2-dimensional power function.

Bibliography

[1] G. Allasia, R. Besenghi, R. Cavoretto, A. De Rossi, *Scattered and track data interpolation using an efficient strip searching procedure*, Appl. Math. Comput. **217** (2011), pp. 5949–5966.

[2] S. Arya, D.M. Mount, N.S. Netanyahu, R. Silverman, A.Y. Wu, *An optimal algorithm for approximate nearest neighbor searching in fixed dimensions*, J. ACM **45** (1998), pp. 891–923.

[3] M. Behzad, G. Chartrand, L. Lesniak-Foster, *Graphs and Digraphs*, Prindle, Weber and Schmidt, Boston, 1979.

[4] D. Belton, *Improving and extending the information on principal component analysis for local neighborhoods in 3D point clouds*, in: C. Jun et al. (Eds.), The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-Part B5, vol. 37, 2008, pp. 477–484.

[5] S. Bochner, *Vorlesungen über Fouriersche Integrale*, Akademische Verlagsgesellschaft, Leipzig, 1932.

[6] L. Bos, M. Caliari, S. De Marchi, M. Vianello, Y. Xu, *Bivariate Lagrange interpolation at the Padua points: the generating curve approach*, J. Approx. Theory **143** (2006), pp. 15–25.

[7] I. Babuška, J.M. Melenk, *The partition of unity method*, Int. J. Numer. Meth. Eng. **40** (1997), pp. 727–758.

[8] M.D. Buhmann, *Radial Basis Functions: Theory and Implementation*, Cambridge Monogr. Appl. Comput. Math., vol. 12, Cambridge Univ. Press, Cambridge, 2003.

[9] J.C. Carr, R.K. Beatson, J.B. Cherrie, T.J. Mitchell, W.R. Fright, B.C. Mccallum, T.R. Evans, *Reconstruction and representation of 3D objects with radial basis functions*, in: Proceedings of the 28-th Annual Conference on Computer Graphics and Interactive Techniques, ACM press, New York, 2001, pp. 67–76.

[10] J.C. Carr, W.R. Fright, R.K. Beatson, *Surface interpolation with radial basis functions for medical imaging*, IEEE Trans. Med. Imaging **16** (1997), pp. 96–107.

[11] R. Cavoretto, *A numerical algorithm for multidimensional modeling of scattered data points*, Comput. Appl. Math. **34** (2015), pp. 65–80.

[12] R. Cavoretto, A. De Rossi, *Achieving accuracy and efficiency in spherical modelling of real data*, Math. Method. Appl. Sci. **37** (2014), pp. 1449–1459.

[13] R. Cavoretto, A. De Rossi, *Spherical interpolation using the partition of unity method: An efficient and flexible algorithm*, Appl. Math. Lett. **25** (2012), pp. 1251–1256.

[14] R. Cavoretto, A. De Rossi, *Fast and accurate interpolation of large scattered data sets on the sphere*, J. Comput. Appl. Math. **234** (2010), pp. 1505–1521.

[15] R. Cavoretto, A. De Rossi, E. Perracchione, *Efficient computation of partition of unity interpolants through a block-based searching technique*, Comput. Math. Appl. **71** (2016), pp. 2568–2584.

[16] Y.L. Chen, S.H. Lai, *A partition of unity based algorithm for implicit surface reconstruction using belief propagation*, in: IEEE International Conference on Shape Modeling and Applications 2007, IEEE Press, Los Alamitos, 2007, pp. 147–155.

[17] K.C. Chung, T.H. Yao, *On lattices adimmitting unique Lagrange interpolations*, SIAM J. Numer. Anal. **14** (1977), pp. 735–743.

[18] I.K. Crain, B.K. Bhattacharyya, *Treatment of non-equispaced two-dimensional data with a digital computer*, Geoexploration **5** (1967), pp. 173–194.

[19] S. Cuomo, A. Galletti, G. Giunta, A. Starace, *Surface reconstruction from scattered point via RBF interpolation on GPU*, in: M. Ganzha et al. (Eds.), 2013 Federated Conference on Computer Science and Information Systems, IEEE Press, Los Alamitos, 2013, pp. 433–440.

[20] M. De Berg, M. Van Kreveld, M. Overmars, O. Schwarzkopf, *Computational Geometry*, Springer, Berlin, 1997.

[21] 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, pp. 319–353, 1874.

[22] 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, pp. 275–339, 1873.

[23] C. de Boor, *A Practical Guide to Splines*, revised edition, Springer, New York 2001.

[24] S. De Marchi, On Leja sequences: some results and applications, Appl. Math. Comput. **152** (2004), pp. 621–647.

[25] S. De Marchi, R. Schaback, H. Wendland, *Near-Optimal Data-independent Point Locations for Radial Basis Function Interpolation* , Adv. Comput. Math. **23** (2005), pp. 317–330

[26] Stefano De Marchi, Maddalena Mandarà and Anna Viero, *Meshfree approximation of multi-asset European and American option problems*, Aracne ed. Roma, pp. 76, ISBN: 9788854851511 (2012).

[27] 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), pp. 325–334.

[28] J. Duchon, *Splines minimizing rotation-invariant semi-norms in Sobolev spaces*, in: W. Schempp at al. (Eds.), Constructive Theory of Functions of Several Variables, Springer Lecture Notes in Mathematics, Springer-Verlag, Berlin, vol. 571, 1977, pp. 85–100.

[29] 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), pp. 5–12.

[30] N. Dyn, D. Levin, *Iterative solution of systems originating from integral equations and surface interpolation*, SIAM J. Numer. Anal. **2** (1981), pp. 377–390.

[31] G.E. Fasshauer, A.Q.M Khaliq and D.A. Voss, *Using Meshfree Approximations for multi-asset American Option Problems* , J. Chin. Inst. Eng. 27 (2004), pp. 563–571.

[32] G.E. Fasshauer, *Meshfree Approximations Methods with* Matlab, World Scientific, Singapore, 2007.

[33] G.E. Fasshauer, M.J. McCourt, *Kernel-based Approximation Methods Using* Matlab, World Scientific, Singapore, 2015.

[34] R. Franke, *Scattered data interpolation: Tests of some methods*, Math. Comp. **38** (1982), pp. 181–200.

[35] R. Franke, *A critical comparison of some methods for interpolation of scattered data*, Report No. NPS-53-79-003, Naval Postgraduate School, Monterey, 1982.

[36] W.R. Goodin, G.J. McRae, J.H. Seinfeld, *A comparison of interpolation methods for sparse data: Application to wind and concentration fields*, J. Appl. Meteor. **18** (1979), pp. 761–771.

[37] R.J. Gould, *Graph Theory*, Dover Publications, Mineola, 2012.

[38] J.P. Gram, *Über Entwicklung reeler Functionen in Reihen mittelst der Methode der kleinsten Quadrate*, J. Math. **94** (1883), pp. 41–73.

[39] R.L. Harder, R.N. Desmarais, *Interpolation using surface splines*, J. Aircraft **9** (1972), pp. 189–191.

[40] R.L. Hardy, *Multiquadric equations of topography and other irregular surfaces*, J. Geophys. Res. **76** (1971), pp. 1905–1915.

[41] A. Heryudono, E. Larsson, A. Ramage, L. Von Sydow, *Preconditioning for radial basis function partition of unity methods*, J. Sci. Comput. **67** (2016), pp. 1089–1109.

[42] Y.C. Hon, R. Schaback, *On unsymmetric collocation by radial basis functions*, Appl. Math. Comput. **119** (2001), pp. 177–186.

[43] H. Hoppe, *Surface Reconstruction from Unorganized Points*, Ph.D. Thesis, University of Washington, Washington, 1994.

[44] H. Hoppe, T. Derose, T. Duchamp, J. Mcdonald, W. Stuetzle, *Surface reconstruction from unorganized points*, in: J.J. Thomas (Ed.), Proceedings of the 19-th annual conference on Computer graphics and interactive techniques, ACM press, New York, vol. 26, 1992, pp. 71–78.

[45] H. Hotelling, *Analysis of a complex of statistical variables into principal components*, J. Educ. Psych. **24** (1993), pp. 417–441.

[46] A. Iske, *Multiresolution Methods in Scattered Data Modelling*, Lecture Notes in Computational Science and Engineering Vol. 37, Springer (2004).

[47] I.T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.

[48] E.J. Kansa, *Application of Hardy's multiquadric interpolation to hydrodynamics*, in: Proc. 1986 Simul. Conf. **4**, 1986, pp. 111–117.

[49] L. Ling, R. Opfer, R. Schaback, *Results on meshless collocation techniques*, Eng. Anal. Bound. Elem. **30** (2006), pp. 247–253.

[50] M. Mathias, *Über positive Fourier-Integrale*, Math. Z. **16** (1923), pp. 103–125.

[51] D.H. McLain, *Two dimensional interpolation from random data*, Comput. J. **19** (1976), pp. 178–181.

[52] J. Meinguet, *Surface spline interpolation: Basic theory and computational aspects*, in: S.P. Singh et al. (Eds.), Approximation Theory and Spline Functions, Reidel, Dordrecht, 1984, pp. 127–142.

[53] J. Meinguet, *Multivariate interpolation at arbitrary points made simple*, Z. Angew. Math. Phys. **30** (1979), pp. 292–304.

[54] J. Meinguet, *An intrinsic approach to multivariate spline interpolation at arbitrary points*, in: N.B. Sahney (Ed.), Polynomial and Spline Approximations, Reidel, Dordrecht, 1979, pp. 163–190.

[55] J. Meinguet, *Basic mathematical aspects of surface spline interpolation*, in: G. Hammerlin (Ed.), Numerische Integration, Birkhauser, Basel, 1979, pp. 211–220.

[56] J.M. Melenk, I. Babuška, *The partition of unity finite element method: Basic theory and applications*, Comput. Meth. Appl. Mech. Eng. **139** (1996), pp. 289–314.

[57] J. Mercer, *Functions of positive and negative type and their connection with the theory of integral equations*, Phil. Trans. Royal Society **209** (1909), pp. 415–446.

[58] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM (Philadelphia), 2000.

[59] C.A. Micchelli, *Interpolation of scattered data: Distance matrices and conditionally positive definite functions*, Constr. Approx. **2** (1986), pp. 11–22.

[60] B. Morse, T.S. Yoo, P. Rheingans, D.T. Chen, K.R. Subramanian, *Interpolating implicit surfaces from scattered surface data using compactly supported radial basis functions*, in: IEEE International Conference on Shape Modeling and Applications 2001, IEEE Press, Los Alamitos, 2001, pp. 89–98.

[61] Y. Ohtake, A. Belyaev, H.P. Seidel, *Sparse surface reconstruction with adaptive partition of unity and radial basis functions*, Graph. Models **68** (2006), pp. 15–24.

[62] Jonas Persson and Lina von Sydow , *Pricing European Multi-asset Options Using a Space-time Adaptive FD-method*, Technical report 2003-059 (2003), Department of Information Technology Uppsala University.

[63] Ulrika Pettersson, Elisabeth Larsson, Gunnar Marcusson and Jonas Persson, *Improved basis function methods for mult-dimensional option pricing*, J. Comput. Appl. Math. **222** (2008), pp. 82–93.

[64] A. Safdari-Vaighani, A. Heryudono, E. Larsson, *A radial basis function partition of unity collocation method for convection-diffusion equations arising in financial applications*, J. Sci. Comput. **64** (2015), pp. 341–367.

[65] R. Schaback, *Convergence of unsymmetric kernel-based meshless collocation methods*, SIAM J. Numer. Anal. **45** (2007), pp. 333–351.

[66] I.J. Schoenberg, *Metric spaces and completely monotone functions*, Ann. Math. **39** (1938), pp. 811–841.

[67] L.L. Schumaker, *Spline Functions - Basic Theory*, Wiley-Interscience, New York 1981.

[68] V. Shcherbakov, E. Larsson, *Radial basis function partition of unity methods for pricing vanilla basket options*, Comput. Math. Appl. **71** (2016), pp. 185–200.

[69] D. Shepard, *A two-dimensional interpolation function for irregularly spaced data*, in: Proceedings of 23-rd National Conference, Brandon/Systems Press, Princeton, 1968, pp. 517–524.

[70] H Stehfest, *Algorithm 368: Numerical inversion of Laplace transforms*, Commun. ACM **13** (1970), pp. 47–49.

[71] J. Stewart, *Positive definite function and generalizations, an historical survey*, Rocky Mountain J. Math. **6** (1976), pp. 409–434.

[72] D.Tavella, C.Randall, *Pricing finantial instuments: the Finite Difference Method*, John Wiey & Son, Ed. New York (2000)

[73] G. Turk, J.F. O'Brien, *Modelling with implicit surfaces that interpolate*, ACM Trans. Graph. **21** (2002), pp. 855–873.

[74] H. Wendland, *Scattered Data Approximation*, Cambridge Monogr. Appl. Comput. Math., vol. 17, Cambridge Univ. Press, Cambridge, 2005.

[75] H. Wendland, *Fast evaluation of radial basis functions: Methods based on partition of unity*, in: C.K. Chui et al. (Eds.), Approximation Theory X: Wavelets, Splines, and Applications, Vanderbilt Univ. Press, Nashville, 2002, pp. 473–483.

[76] 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), pp. 389–410.

[77] G. Yao, J. Duo, C.S. Chen, L.H. Shen, *Implicit local radial basis function interpolations based on function values*, Appl. Math. Comput. **265** (2015), pp. 91–102.

[78] Z. Wu, R. Schaback, *Local error estimates for radial basis function interpolation of scattered data*, IMA J. Numer. Anal. **13** (1993), pp. 13–27.

[79] S. Zhu, A.J. Wathen, *Convexity and solvability for compactly supported radial basis functions with different shapes*, J. Sci. Comput. **63** (2015), pp. 862–884.