Positive Definite Kernels: Past, Present and Future*

Gregory E. Fasshauer
Department of Applied Mathematics
Illinois Institute of Technology
fasshauer@iit.edu

Happy 65th Birthday, Robert!

Abstract

Positive definite kernels play an increasingly prominent role in many applications such as scattered data fitting, numerical solution of PDEs, computer experiments, machine learning, rapid prototyping and computer graphics. We discuss some of the historical and current developments of the theory and applications of positive definite kernels – always with an eye toward the mathematics of Göttingen in general and Robert Schaback in particular. A few comments concerning the future of the field are also provided.

1 Introduction

This article represents an expanded version of a colloquium presentation made during the *Workshop on Kernel Functions and Meshless Methods honoring the 65th birthday of Robert Schaback* in Göttingen on January 14, 2011. As such, the following discussion of the current state-of-the-art of approximation methods based on the use of positive definite kernels is intentionally biased toward favoring the work of Robert Schaback, his many students and collaborators, and the University of Göttingen, in general.

1.1 Working with Positive Definite Kernels

Let us begin by roughly outlining three general areas of study that involve positive definite kernels.

Theory: The foundation for this work lies mostly in *functional, numerical and stochastic analysis* and deals with concepts such as reproducing kernel Hilbert spaces, positive definite and completely monotone functions, convergence analysis, alternate basis representations, and (Gaussian) random fields.

Computation: This work reaches into *numerical linear algebra, computational statistics and computer science* and is concerned with issues such as parameter selection, stable, fast and efficient algorithms, regularization techniques, appropriate data structures, and the implementation in a high-performance computing environment.

Applications: This area is arguably the largest of these three and covers problems such as basic data fitting (in both the deterministic and stochastic settings), the numerical solution of

*Part of the work reported here was supported by the National Science Foundation under Grant No. DMS-0713848.

PDEs (both deterministic and stochastic), statistical or machine learning and classification, multivariate integration, multivariate optimization, engineering design (both in the sense of geometric design as well as in the sense of design of experiments), computer graphics, and many more.

Even though I have decided to describe the field of positive definite kernels using these three categories, the boundaries separating the areas are rather soft. In order for someone to make significant progress on any of the topics listed above, that person will almost certainly require at least some expertise in all three categories. As will become apparent from the discussion that follows, the contributions of Robert Schaback have been numerous and significant in all three of these general areas.

An indication of the important role positive definite kernels play in many different fields is provided by the following rapidly growing, and probably incomplete, list of monographs. All of these books contain at least a significant portion that is concerned with positive definite kernels.

Analysis: Berg, Christensen and Ressel [8], Bergman [9], Bochner [11], Meschkowski [103], Saitoh [123, 124], Wells and Williams [152]

Approximation Theory: Buhmann [21], Cheney and Light [24], Fasshauer [38], Freeden, Gervens and Schreiner [55], Iske [74], Wendland [155]

Engineering Applications: Atluri [3], Atluri and Shen [4], Belytschko and Chen [7], Forrester, Sobester and Keane [52], Li and Liu [88], Liu [90], Liu and Liu [91]

Geostatistics: Cressie [28], Kitanidis [80], Matérn [95], Matheron [97], Stein [146]

Probability Theory and Statistics: Berlinet and Thomas-Agnan [10], Wahba [151]

Statistical/Machine Learning: Catoni [23], Cristianini and Shawe-Taylor [29], Cucker and Zhou [30], Hastie, Tibshirani and Friedman [65], Joachims [76], Rasmussen and Williams [117], Schölkopf and Smola [141], Shawe-Taylor and Cristianini [144], Steinwart and Christmann [147], Vapnik [150]

In addition to these monographs, the historical papers by Aronszajn [2], Mercer [101] and Schmidt [140] as well as a few more recent survey articles such as those by Stewart [148], Martin Buhmann [20] and by Robert Schaback with Holger Wendland [135] and with Michael Scheuerer and Martin Schlather [139] should not be forgotten.

1.2 History

While the serious mathematical development of positive definite functions and reproducing kernel Hilbert spaces did not begin until well into the 20$^{\text{th}}$ century (see Section 2.2), one of the best known positive definite kernels, the *Gaussian kernel* or normal distribution function (see Example 1)

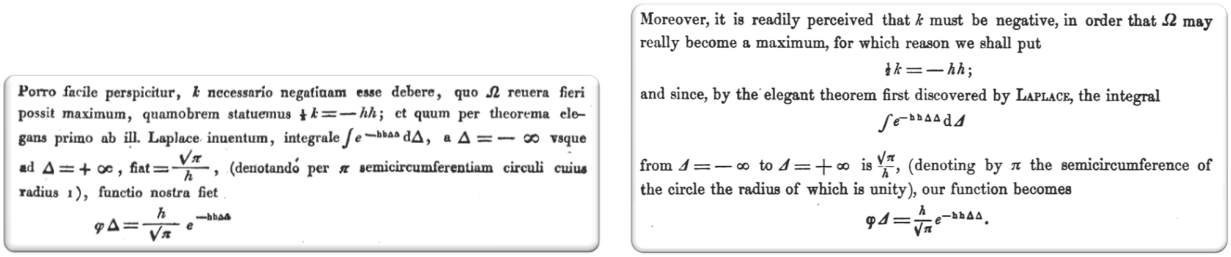
$$K(x,y) = \mathrm{e}^{-\varepsilon^2|x-y|^2}, \qquad x,y \in \mathbb{R}, \varepsilon > 0, \tag{1}$$

is closely associated with Carl Friedrich Gauß, who was director of the observatory in Göttingen from 1807 until his death in 1855.

Figure 1: Carl Friedrich Gauß in 1840.

The first occurrence of the Gaussian kernel in the writings of Gauß is shown in Figure 2. Gauß mentioned the function that now so often

carries his name in 1809 in his second book *Theoria motus corporum coelestium in sectionibus conicis solem ambientium* (Theory of the motion of the heavenly bodies moving about the sun in conic sections) [57] in the context of *least squares* calculations of the *maximum likelihood* (two other concepts introduced by Gauß in this book) of his astronomical measurements. As one can see, Gauß gives credit to Laplace for having first computed the relevant integral (in 1783). Figure 3 shows the corresponding excerpt from Pierre-Simon de Laplace's work [81, Sect.XXXVII].



Porro facile perspicitur, k necessario negatiuam esse debere, quo Ω reuera fieri possit maximum, quamobrem statuemus $\frac{1}{2}k = -hh$; et quum per theorema elegans primo ab ill. Laplace inuentum, integrale $\int e^{-hh\Delta\Delta} d\Delta$, a $\Delta = -\infty$ vsque ad $\Delta = +\infty$, fiat $= \frac{\sqrt{\pi}}{h}$, (denotando per π semicircumferentiam circuli cuius radius 1), functio nostra fiet

$$\varphi\Delta = \frac{h}{\sqrt{\pi}} e^{-hh\Delta\Delta}$$

(a) Original source in Latin.

Moreover, it is readily perceived that k must be negative, in order that Ω may really become a maximum, for which reason we shall put

$$\tfrac{1}{2}k = -hh;$$

and since, by the elegant theorem first discovered by Laplace, the integral

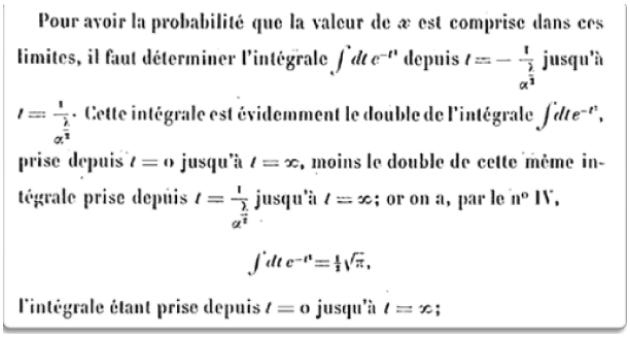
$$\int e^{-hh\Delta\Delta} d\Delta$$

from $\Delta = -\infty$ to $\Delta = +\infty$ is $\frac{\sqrt{\pi}}{h}$, (denoting by π the semicircumference of the circle the radius of which is unity), our function becomes

$$\varphi\Delta = \frac{h}{\sqrt{\pi}} e^{-hh\Delta\Delta}.$$

(b) English translation.

Figure 2: Excerpt from [57, Sect.177] showing the first occurrence of the Gaussian kernel in the work of Gauß.



Pour avoir la probabilité que la valeur de x est comprise dans ces limites, il faut déterminer l'intégrale $\int dt\, e^{-t}$ depuis $t = -\frac{1}{2\alpha^2}$ jusqu'à $t = \frac{1}{2\alpha^2}$. Cette intégrale est évidemment le double de l'intégrale $\int dt e^{-t}$, prise depuis $t = 0$ jusqu'à $t = \infty$, moins le double de cette même intégrale prise depuis $t = \frac{1}{2\alpha^2}$ jusqu'à $t = \infty$; or on a, par le n° IV,

$$\int dt\, e^{-t} = \tfrac{1}{2}\sqrt{\pi},$$

l'intégrale étant prise depuis $t = 0$ jusqu'à $t = \infty$;

Figure 3: Laplace's earlier [81, Sect.XXXVII] discussion of the normal distribution.

In order to provide a smooth and direct connection from Carl Friedrich Gauß to Robert Schaback we display a mathematical ancestry tree of Robert Schaback compiled based on data available at [96] in Figure 4. A few things about this ancestry chart are worth noting.

1. Each of the mathematicians listed in an orange-yellowish box had his primary appointment in Göttingen.

2. Many of the names listed have made significant contributions to the foundations of the present-day field of positive definite kernels (certainly Gauß, Bessel, Hilbert, Schmidt, Bochner, Karlin and Schaback).

3. The lower-left branch was added because it features the names of Schmidt, Bochner and Karlin.

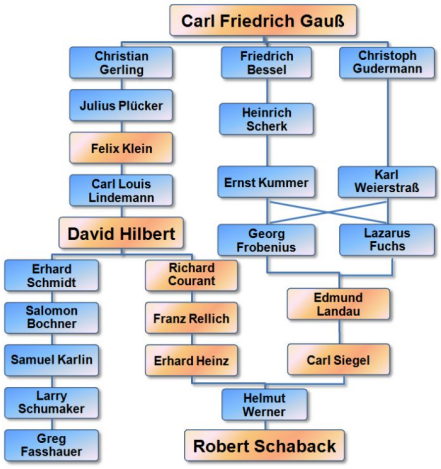


Figure 4: Robert Schaback's mathematical ancestry traced back to Gauß.

4. The top portion of the tree is so significant that is was chosen as the image on the main page of [96].

The mathematical heritage of Gauß, Hilbert and Schaback is continued by Robert Schaback's many Ph.D. students listed in Table 1. Robert mentored the first four students on the list during his time in Bonn from 1973-1976. The names of those students whose thesis was related to positive definite kernels was italicized. This provides further evidence for the important role the University of Göttingen plays in the field of positive definite kernels.

Ludwig Cromme (1975)	Hartmut Förster (1977)	Werner Ehm (1977)
Sabine Kamprowski (1978)	Immo Diener (1983)	Walter Brübach (1984)
Elizabeth Gillet (1985)	Klaus Nottbohm (1986)	*Zongmin Wu* (1986)
Heiko Bürger (1992)	*Armin Iske* (1994)	*Marko Weinrich* (1994)
Thomas Früh (1995)	*Holger Wendland* (1996)	Martina Domeyer (1997)
Carsten Franke (1998)	*Lin-Tian Luh* (1998)	*Sandro Hartmann* (1998)
Jens Trapp (1998)	Uwe Heinz Bünting (1999)	Robert Zores (1999)
Anja Schreiber (2000)	*Roland Opfer* (2004)	*Mohammed Mouattamid* (2005)
Karsten Scheibe (2006)	*Tobias Block* (2008)	*Christian Rieger* (2008)
Stefan Müller (2009)		

Table 1: Robert Schaback's Ph.D. students at the Universities of Bonn (1975-1978) and Göttingen. Italicized names indicate a thesis topic related to positive definite kernels.

2 Positive Definite Kernels and Reproducing Kernel Hilbert Spaces

2.1 Positive Definite Kernels

The study of positive definite functions, or – slightly more generally – positive definite kernels, in the field of analysis began either with the work of Maximilian Mathias [98], a student of Erhard Schmidt's or with that of James Mercer [101]. As pointed out in Stewart's survey from 1976 [148],

> *Mathias and the other early workers with p.d. functions of a real variable were chiefly concerned with Fourier transforms and apparently did not realize that more than a decade previously Mercer and others had considered the more general concept of positive definite kernels $K(x, y)$ [...] in research on integral equations. I have likewise found that present-day mathematicians working with some of the manifestations of p.d. functions are unaware of other closely related ideas. Thus one of the purposes of this survey is to correlate some of the more important generalizations of p.d. functions with the hope of making them better known.*

Perhaps the most fundamental contributions, namely characterizations of positive definite functions in terms of Fourier transforms, were made a few years later by Salomon Bochner [11] and Iso Schoenberg [142]. These contributions were used by Micchelli [104] as the starting point of his proofs of the non-singularity of the system matrices associated with radial basis function interpolation (more on interpolation later). Also in the 1930s, Aleksandr Khinchin [79] used Bochner's theorem to establish the foundation for the study of stationary stochastic processes in probability theory.

Figure 5: Left to right: Erhard Schmidt, James Mercer, Salomon Bochner, Iso Schoenberg and Aleksandr Khinchin.

We now present the well-known definition of a positive definite matrix K as it can be found in just about any book on linear algebra (see, e.g., [72]) and relate the concept of positive definite functions/kernels to this idea.

Definition 1. *A real symmetric* $N \times N$ matrix K *is called* positive definite *if its associated quadratic form is positive for any nonzero coefficient vector* $\boldsymbol{c} = [c_1, \ldots, c_N]^T \in \mathbb{R}^N$*, i.e.,*

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j \mathsf{K}_{ij} > 0.$$

For the purposes of this paper a *kernel* K is nothing but a real-valued *function of two variables*, i.e.,

$$K : \Omega \times \Omega \to \mathbb{R}, \qquad K : (\boldsymbol{x}, \boldsymbol{z}) \mapsto K(\boldsymbol{x}, \boldsymbol{z}).$$

Here Ω is usually a subset of $\mathbb{R}^d$, but it may also be a rather general set as happens frequently in statistical learning applications. Other possible domains Ω include spheres or other Riemannian manifolds (see, e.g., Example 3 or [56]), or even locally compact groups [62]. More examples of such general settings are reported in [148].

A positive definite kernel K can be viewed as an infinite-dimensional positive definite matrix K. In fact, this can be done in two different ways. First, in the sense of Mathias, where we assume that the matrix K is generated by the kernel K in the following sense.

Definition 2. *A symmetric* kernel K *is called* positive definite on Ω *if its associated kernel matrix* $\mathsf{K} = (K(\boldsymbol{x}_i, \boldsymbol{x}_j))_{i,j=1}^{N}$ *is positive definite* for any $N \in \mathbb{N}$ *and* for any *set of distinct points* $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$.

Many generalizations of this basic definition exist. For example, the coefficient vectors $\boldsymbol{c} = [c_1, \ldots, c_N]^T$ as well as the kernel K may be allowed to be complex (as already assumed in Bochner's work), or the kernel may be matrix-valued as is desired for recent applications in fluid dynamics, where one may want to ensure that the kernel is divergence-free by construction (see, e.g., [143] and the earlier fundamental work by Narcowich and Ward [109] as well as [92]).

Alternatively, coming more from the work of Mercer, who was concerned with integral operators, we can replace the finite-dimensional discrete quadratic form by an infinite-dimensional one and arrive at a definition such as

Definition 3. *A symmetric* kernel K *is called* integrally positive definite on Ω *if*

$$\int_{\Omega} \int_{\Omega} K(\boldsymbol{x}, \boldsymbol{z}) u(\boldsymbol{x}) u(\boldsymbol{z}) \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{z} > 0$$

Figure 6: Left to right: David Hilbert, Nachman Aronszajn, Stefan Bergman and E. H. Moore.

for all $u \in L_1(\Omega)$.

Bochner showed that the notions of positive definiteness as stated in Definitions 2 and 3 are equivalent for continuous kernels. We will return to the work of Mercer and Schmidt and its connection to integral operators in Section 6 when we discuss eigenfunction expansions of positive definite kernels.

2.2 Reproducing Kernel Hilbert Spaces

The concept of a *reproducing kernel Hilbert space* was introduced in 1950 independently by Nachman Aronszajn [2] and Stefan Bergman [9], and even earlier by E. H. Moore [106] who referred to reproducing kernels as *positive Hermitian matrices*, and to whom, along with Mercer, Aronszajn gives much credit in [2].

The *reproducing property* satisfied by a symmetric reproducing kernel K is as follows

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

i.e., the reproducing kernel acts as a point evaluation functional for all functions $f \in \mathcal{H}(K, \Omega)$. Here $\mathcal{H}(K, \Omega)$ is a *Hilbert space* of functions on Ω (the reproducing kernel Hilbert space of K) and $\langle \cdot, \cdot \rangle$ denotes the associated inner product. The term *native space of* K was introduced by Robert Schaback [130, 131] in the context of radial basis functions.

Many more details on positive definite functions, positive definite kernels and reproducing kernel Hilbert spaces can be found in most of the books listed in Section 1.

2.3 Examples of Positive Definite Kernels

We now briefly present a kind of dictionary of different types of positive definite kernels. The aim is by no means to be exhaustive, but to provide a general impression of the various types of kernels one might consider for different applications. Again, most of the books listed in Section 1 can serve as a source for many alternate examples (often specialized for certain kinds of applications, such as radial basis functions, covariance kernels for kriging methods, etc.).

Most of the examples we have chosen to present are closely related to the well-known Gaussian kernel already mentioned in (1).

Example 1 (Radial kernels)**.** Much of the work on positive definite kernels in approximation theory (and also much of the work of Robert Schaback) has focused on so-called *radial basis functions*. In other fields such as probability theory and statistics such kernels are usually referred to as *isotropic*. These kernels are rotationally (and also translationally) invariant, i.e., their level curves (or more

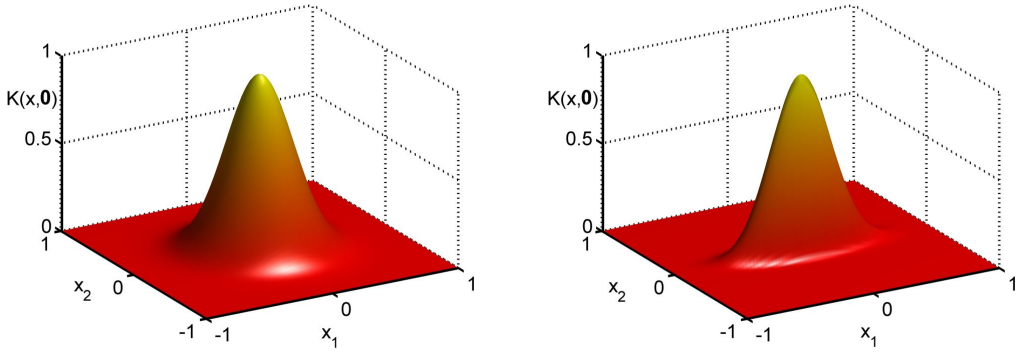
generally level hyper-surfaces) are circles (hyper-spheres) and they can be expressed in terms of a univariate function $\kappa : \mathbb{R}_0^+ \to \mathbb{R}$ of a non-negative scalar variable, i.e.,

$$K(\boldsymbol{x}, \boldsymbol{z}) = \kappa(\|\boldsymbol{x} - \boldsymbol{z}\|), \quad \boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^d,$$

where $\|\cdot\|$ is some norm on $\mathbb{R}^d$ – usually the Euclidean norm. A typical example of a radial (or isotropic) kernel is the *multivariate Gaussian*

$$K(\boldsymbol{x}, \boldsymbol{z}) = \kappa(\|\boldsymbol{x} - \boldsymbol{z}\|) = \mathrm{e}^{-\varepsilon^2 \|\boldsymbol{x}-\boldsymbol{z}\|^2}, \qquad \kappa(r) = \mathrm{e}^{-\varepsilon^2 r^2}, \tag{2}$$

whose graph is shown in Figure 7a for the case $d = 2$ with *center* $\boldsymbol{z}$ placed at the origin. The parameter ε appearing in the definition of the kernel goes by many different names. We will refer to it as the *shape parameter* as is customary in the radial basis function literature, but other terms such as *scale parameter*, *width*, or (reciprocal of the) standard deviation (such that $\frac{1}{\varepsilon^2} = \sigma^2$, the *variance*) are frequently encountered in the literature.



(a) Isotropic Gaussian kernel with shape parameter $\varepsilon = 3$.

(b) Anisotropic Gaussian with shape parameters $\varepsilon_1 = 3, \varepsilon_2 = 9$.

Figure 7: Two-dimensional, $\boldsymbol{x} = (x_1, x_2)$, Gaussian kernel centered at the origin $\boldsymbol{z} = (0, 0)$.

Example 2 (Translation invariant kernels)**.** It is quite common to relax the symmetry requirements on the kernel somewhat and demand that it still provides translational (but no longer rotational) invariance. Such kernels are frequently called *stationary kernels*, especially in the statistics literature. This nomenclature is a little unfortunate (or the definition of stationarity chosen by approximation theorists is) since the word stationary has its own specific meaning in the study of convergence rates of approximation schemes using scaled approximation spaces (see, e.g., the discussion in [38]). A translation invariant kernel can be expressed in terms of a multivariate function $\widetilde{K} : \mathbb{R}^d \to \mathbb{R}$ of a single d-dimensional variable, i.e.,

$$K(\boldsymbol{x}, \boldsymbol{z}) = \widetilde{K}(\boldsymbol{x} - \boldsymbol{z}), \quad \boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^d.$$

A standard example of a multivariate anisotropic translation invariant kernel is provided by the tensor product of univariate translation invariant kernels:

$$K(\boldsymbol{x}, \boldsymbol{z}) = \prod_{\ell=1}^{d} \widetilde{K}_\ell(x_\ell - y_\ell), \quad x_\ell, y_\ell \in \mathbb{R},$$

where the one-dimensional kernels $\widetilde{K}_\ell : \mathbb{R} \to \mathbb{R}$, $\ell = 1, \dots, d$, could all have different shape parameters. A typical example is given by the *anisotropic Gaussian kernel*

$$K(\boldsymbol{x}, \boldsymbol{z}) = \prod_{\ell=1}^{d} \mathrm{e}^{-\varepsilon_\ell^2 (x_\ell - y_\ell)^2} = \mathrm{e}^{-\sum_{\ell=1}^{d} \varepsilon_\ell^2 (x_\ell - y_\ell)^2}, \quad \widetilde{K}(\boldsymbol{z}) = \mathrm{e}^{-\boldsymbol{z}^T \mathsf{E} \boldsymbol{z}}.$$

The plot for a two-dimensional anisotropic Gaussian with shape parameter $\varepsilon_1 = 3$ in the x_1-direction and $\varepsilon_2 = 9$ in the x_2-direction is shown in Figure 7b.

Note that one could also interpret the anisotropic Gaussian as a radial kernel provided the norm $\|\cdot\|$ is not the basic Euclidean norm, but instead a weighted 2-norm of the form $\|\boldsymbol{z}\|_{\mathsf{E}} = \sqrt{\boldsymbol{z}^T \mathsf{E} \boldsymbol{z}}$ with diagonal matrix $\mathsf{E} = \mathrm{diag}(\varepsilon_1, \dots, \varepsilon_d)$. In fact, any symmetric positive definite matrix E could be used to define an admissible inner product/norm. Clearly, the anisotropic Gaussian becomes isotropic (2-norm radial) if $\varepsilon_\ell = \varepsilon$ for all $\ell = 1, \dots, d$.

Example 3 (Zonal kernels)**.** As mentioned in Section 2.1, the domain Ω need not be a Euclidean space $\mathbb{R}^d$. The unit sphere $\mathbb{S}^2$ in $\mathbb{R}^3$ is a domain that is important for many geophysical applications. In this case, the analog of a radial kernel is given by a so-called *zonal kernel* which can be described in terms of a scalar univariate function $\tilde{\kappa} : [-1, 1] \to \mathbb{R}$ of the dot product of two points on the surface of the sphere, i.e.,

$$K(\boldsymbol{x}, \boldsymbol{z}) = \tilde{\kappa}(\boldsymbol{x} \cdot \boldsymbol{z}), \quad \boldsymbol{x}, \boldsymbol{z} \in \mathbb{S}^2.$$

The analogy to the radial formulation is provided by the fact that the geodesic distance δ, i.e., the shortest distance measured along a great circle arc, between two points $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{S}^2$ is given by $\delta(\boldsymbol{x}, \boldsymbol{z}) = \arccos \boldsymbol{x} \cdot \boldsymbol{z}$.

If we express the Gaussian kernel in terms of the geodesic distance, we obtain a *spherical Gaussian*

$$K(\boldsymbol{x}, \boldsymbol{z}) = \mathrm{e}^{-2\varepsilon(1 - \boldsymbol{x} \cdot \boldsymbol{z})},$$

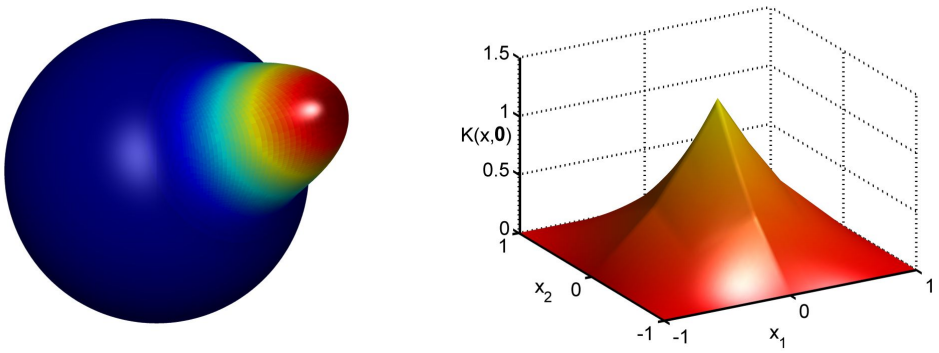
i.e., $\tilde{\kappa}(t) = \mathrm{e}^{-2\varepsilon(1-t)}$. The graph of such a spherical Gaussian kernel with shape parameter $\varepsilon = 9$ centered at the point $\boldsymbol{z} = \frac{1}{\sqrt{3}}(1, 1, 1) \in \mathbb{S}^2$ is shown in Figure 8a.

We also remark that the Euclidean distance (which cuts straight through the sphere) and the geodesic distance δ (which measures distance along the surface of the sphere) are linked by $\|\boldsymbol{x} - \boldsymbol{z}\| = \sqrt{2 - 2\boldsymbol{x} \cdot \boldsymbol{z}} = 2 \sin \frac{\delta(\boldsymbol{x}, \boldsymbol{z})}{2}$. This relation can be used to define restricted Euclidean kernels that were referred to as "poor man's" spherical kernels in [38]. This setting has recently been carefully studied in [56] and applied to various two-dimensional manifolds embedded in $\mathbb{R}^3$. It should be noted that this kind of restricted kernel will in general no longer be zonal in terms of the underlying Riemannian metric of the manifold.

Example 4 (Multiscale kernels)**.** Roland Opfer developed the idea of so-called *multiscale kernels* in his Ph.D. thesis [112] under the direction of Robert Schaback (see also [113]). Such kernels are given by linear combinations of shifts and scales of a single compactly supported kernel φ, i.e.,

$$K(\boldsymbol{x}, \boldsymbol{z}) = \sum_{j \geq 0} \lambda_j K_j(\boldsymbol{x}, \boldsymbol{z}) = \sum_{j \geq 0} \lambda_j \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \varphi(2^j \boldsymbol{x} - \boldsymbol{k}) \varphi(2^j \boldsymbol{z} - \boldsymbol{k}).$$

As a consequence, these kernels have wavelet-like properties and are well-suited for applications such as image compression with the advantage that these new multiscale kernels need not be placed at points $\boldsymbol{x}, \boldsymbol{z}$ of a regular grid.



(a) Spherical Gaussian with shape parameter $\varepsilon = 9$ and $\boldsymbol{z} = \frac{1}{\sqrt{3}}(1,1,1)$.

(b) Piecewise linear multiscale kernel centered at $\boldsymbol{z} = (0,0)$.

Figure 8: Spherical Gaussian and piecewise linear multiscale kernel.

A typical multiscale kernel can be formed by starting, e.g., with the piecewise linear kernel

$$\varphi(\boldsymbol{x}) = \prod_{\ell=1}^{d} (1 - x_\ell)_+,$$

where

$$(x)_+ = \max(x, 0) = \begin{cases} x, & \text{if } x > 0, \\ 0, & \text{otherwise} \end{cases}$$

is the cut-off function. Using four levels ($j = 0, 1, 2, 3$) and dyadic weights $\lambda_j = 2^{-2j}$ we get

$$K(\boldsymbol{x}, \boldsymbol{z}) = \sum_{j=0}^{3} 2^{-2j} \sum_{\boldsymbol{k} \in \mathbb{Z}^2} \varphi(2^j \boldsymbol{x} - \boldsymbol{k}) \varphi(2^j \boldsymbol{z} - \boldsymbol{k})$$

whose graph is displayed in Figure 8b for $\boldsymbol{z} = (0,0)$.

Example 5 (Power series kernels)**.** Barbara Zwicknagl's Diplom thesis [159] under the guidance of Robert Schaback led to the development of so-called *power series kernels* on $\Omega = (-1,1)^d$ (see also her presentation [160] at the Göttingen workshop). They are of the form

$$K(\boldsymbol{x}, \boldsymbol{z}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}_0^d} w_{\boldsymbol{\alpha}} \frac{\boldsymbol{x}^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} \frac{\boldsymbol{z}^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!},$$

where the coefficients $w_{\boldsymbol{\alpha}}$ need to satisfy the summability condition $\sum_{\boldsymbol{\alpha} \in \mathbb{N}_0^d} \frac{w_{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!^2} < \infty$.

A simple example is provided by the *exponential kernel*

$$K(\boldsymbol{x}, \boldsymbol{z}) = \mathrm{e}^{\boldsymbol{x} \cdot \boldsymbol{z}} = \sum_{n=0}^{\infty} \frac{1}{n!} (\boldsymbol{x} \cdot \boldsymbol{z})^n = \sum_{\boldsymbol{\alpha} \in \mathbb{Z}^d} \frac{1}{|\boldsymbol{\alpha}|!} \binom{|\boldsymbol{\alpha}|}{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} \boldsymbol{z}^{\boldsymbol{\alpha}}$$

of which two instances are plotted in Figure 9. In Figure 9a K is centered at the point $\boldsymbol{z} = (\frac{1}{2}, \frac{1}{2})$, while in Figure 9b the center is given by $(\frac{3}{4}, \frac{3}{4})$. Note that this kernel is *not* translation invariant.

Similar (finite) polynomial or infinite power series kernels of the dot product type frequently arise in the machine learning literature.

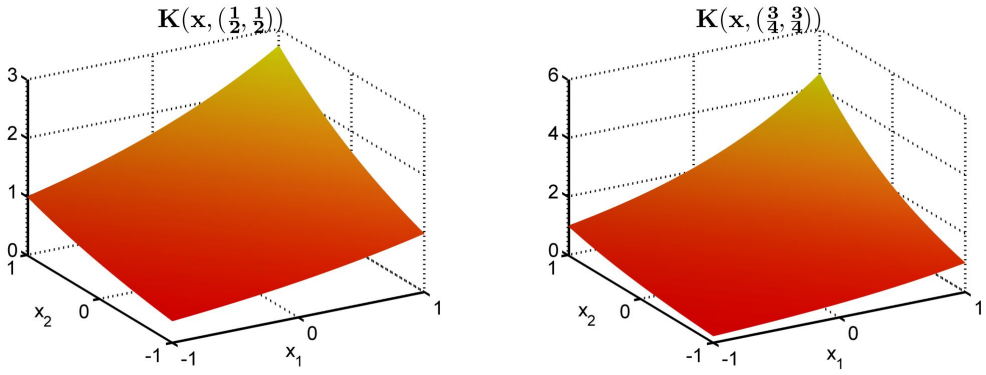
In [159] the power series kernel

$$K(\boldsymbol{x}, \boldsymbol{z}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}_0^d} \frac{(2\varepsilon^2)^{|\boldsymbol{\alpha}|}}{\boldsymbol{\alpha}!} \boldsymbol{x}^{\boldsymbol{\alpha}} \boldsymbol{z}^{\boldsymbol{\alpha}} \tag{3}$$

was used to derive an alternative proof of the spectral convergence order of the Gaussian kernel (see Section 5.1) by using a decomposition of the Gaussian of the form

$$\mathrm{e}^{-\varepsilon^2 \|\boldsymbol{x}-\boldsymbol{z}\|^2} = \mathrm{e}^{-\varepsilon^2 \|\boldsymbol{x}\|^2} K(\boldsymbol{x}, \boldsymbol{z}) \mathrm{e}^{-\varepsilon^2 \|\boldsymbol{z}\|^2},$$

where K is the special kernel of (3).



(a) Kernel centered at $\boldsymbol{z} = (\frac{1}{2}, \frac{1}{2})$.

(b) Kernel centered at $\boldsymbol{z} = (\frac{3}{4}, \frac{3}{4})$.

Figure 9: Exponential power series kernels.

We will return to series expansions of positive definite kernels in Section 6 when we expand a kernel in terms of its orthogonal eigenfunctions.

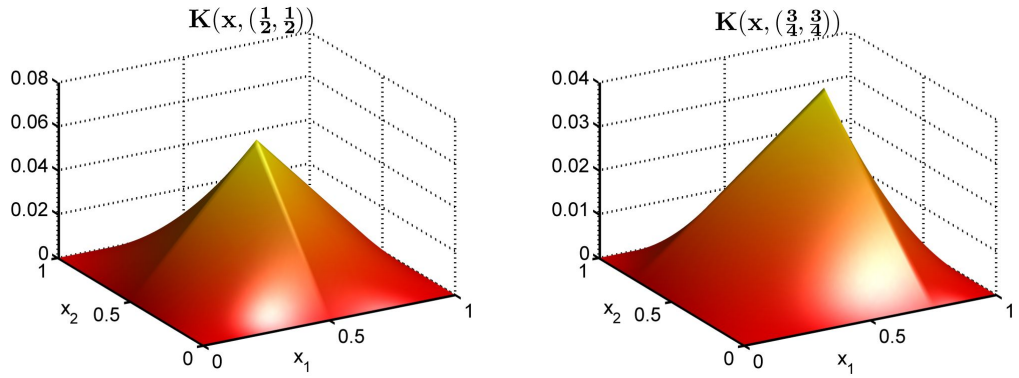
Example 6 (Other general kernels)**.** Kernels $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ of the most general type, i.e.,

$$K(\boldsymbol{x}, \boldsymbol{z}), \quad \boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^d,$$

are also common in the literature. One of the best-known examples may be the *Brownian bridge product kernel*

$$K(\boldsymbol{x}, \boldsymbol{z}) = \prod_{\ell=1}^{d} \left(\min\{x_\ell, y_\ell\} - x_\ell y_\ell \right).$$

Two instances of this kernel are plotted in Figure 10; K is centered at $(\frac{1}{2}, \frac{1}{2})$ in Figure 10a and at $(\frac{3}{4}, \frac{3}{4})$ in Figure 10b. Clearly, this kernel is *not* translation invariant. In contrast to the other examples listed above, the Brownian bridge kernel satisfies zero boundary conditions on the boundary of the unit cube. This kernel is sometimes used to simulate stochastic processes driven by Brownian motion with zero boundary conditions arising in finance applications (see, e.g., [111]).



(a) Kernel centered at $\boldsymbol{z} = (\frac{1}{2}, \frac{1}{2})$.

(b) Kernel centered at $\boldsymbol{z} = (\frac{3}{4}, \frac{3}{4})$.

Figure 10: Brownian bridge tensor product kernels.

2.4 Misconception I

The fundamental work of Bochner [11], and subsequently Schoenberg [142], on the characterization of positive definite radial functions in terms of completely monotone functions, i.e., functions whose derivatives satisfy the alternating sign property

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

has played an important role in the development of both the theory and application of radial basis functions. However, as sometimes happens in mathematics – especially when one does not pay careful attention to the hypotheses of a theory, and remembers facts only as sound bites – unfortunate *misconceptions* may arise. In the case of positive definite radial functions many people may have associated the Bochner/Schoenberg characterization with something like

Sound bite 1. *"Positive definite radial functions are equivalent to completely monotone functions."*

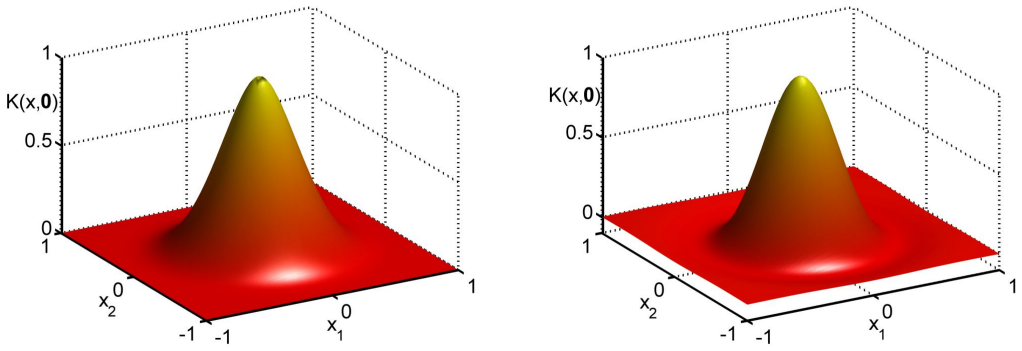
This is very unfortunate since it probably prevented researchers for quite some time from thinking about oscillatory radial basis functions or even compactly supported radial functions since both phenomena are ruled out by the alternating sign property associated with complete monotonicity.

So when *compactly supported radial basis functions* were introduced by Robert Schaback [126] this created quite a stir, and a more careful reading of Schoenberg's work shows that the above misconception should be corrected to something more like

Sound bite 2 (Improvement of Sound bite 1)**.** *"Functions that are positive definite radial* on $\mathbb{R}^d$ for all d *are equivalent to completely monotone functions."*

This latter statement is in fact still not completely precise since one also needs to account for a variable transformation $r \mapsto r^2$ (for complete details see [38] or [155]).

Once this important detail was noticed, a number of constructions for compactly supported radial kernels were suggested (see, e.g., [19, 136, 153] and also the presentation of Armin Iske at the Göttingen workshop [75]). One typical compactly supported radial kernel, a so-called Wendland function, is displayed in Figure 11a. We would also like to point out that the d-fold tensor product Brownian bridge kernel of Example 6 is compactly supported and positive definite in $\mathbb{R}^d$, but not radial.



(a) C^2 Wendland function $\kappa(r) = (1-r)_+^4(4r+1)$ centered at the origin in $\mathbb{R}^2$.

(b) $\kappa(r) = (1 - \varepsilon^2 r^2 + \frac{\varepsilon^4 r^4}{6})e^{-\varepsilon^2 r^2}$, a Laguerre-Gaussian kernel with $\varepsilon = 3$ centered at the origin in $\mathbb{R}^2$.

Figure 11: Compactly supported and oscillatory radial kernels.

Subsequently, *oscillatory radial basis functions* were also suggested by various people (see, e.g., [45, 48]), and Figure 11b shows a Laguerre-Gaussian function from [45].

We end this section with a quote from the recent survey by Stefano DeMarchi and Robert Schaback [33] which appeared in this journal:

> *"new nonstandard kernels [...] should get more attention."*

We could not agree more.

3 Kernel Interpolation

3.1 Scattered Data Fitting

In order to get into the more practical aspects of working with positive definite kernels we present

Fundamental Application (The Scattered Data Fitting Problem)**.** *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* s_f *such that* $s_f(\boldsymbol{x}_j) = y_j$, $j = 1, \ldots, N$.

Traditionally, this kind of problem arises in geophysical applications such as the terrain model illustrated in Figure 12. The data for this example can be found as the `volcano` data set in the statistical software package R [116] and represents real measurements of 5307 elevation measurements obtained from Maunga Whau (Mt. Eden) in Auckland, NZ, obtained on a $10m \times 10m$ grid. In order to be able to distinguish individual measurements in the plots displayed in Figure 12 we selected a random (scattered) subset of 10% of the original data points.

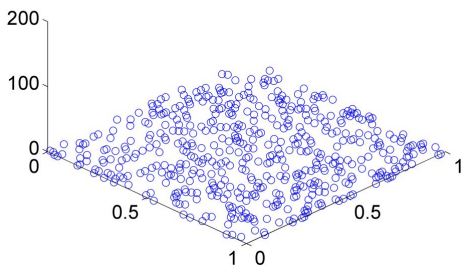
The fitting surface is obtained using a linear combination of Gaussian kernels of the form

$$s_f(\boldsymbol{x}) = \sum_{j=1}^{N} c_j e^{-\varepsilon^2 \|\boldsymbol{x} - \boldsymbol{x}_j\|^2}, \tag{4}$$

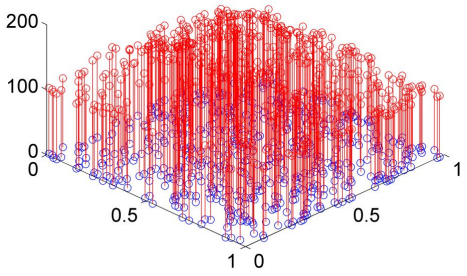
where the $\boldsymbol{x}_j$ correspond to the $N = 530$ scattered blue points shown in Figure 12a (normalized to the unit square), and the unknown coefficients c_j are obtained by solving the interpolation equations

$$s_f(\boldsymbol{x}_i) = y_i, \quad i = 1, \ldots, N,$$

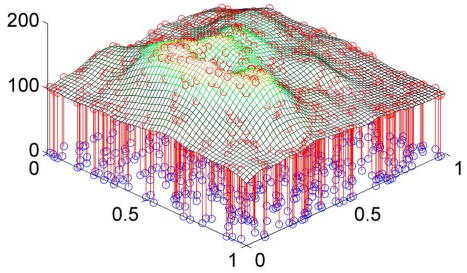
where the y_i correspond to the red elevation measurements shown in Figure 12b together with their associated data sites $\boldsymbol{x}_i$. A value of $\varepsilon = 15.605278$ was used for this example and was obtained by leave-one-out cross-validation (LOOCV). A more detailed discussion of algorithms for the selection of the shape parameter appears below in Section 5.2.



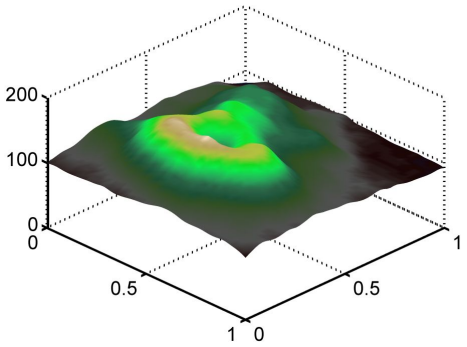
(a) Scattered measurement locations for volcano data (normalized to the unit square).



(b) Elevation measurements associated with the locations shown in (a).



(c) Volcano data fitted by a Gaussian kernel interpolant.



(d) Surface of (c) rendered with more realistic colors.

Figure 12: Illustration of the scattered data fitting problem.

It is probably worth noting that geostatisticians most certainly will prefer to fit this kind of data using a *kriging method* based on a kernel which is not nearly as smooth as the Gaussian. Moreover, such a model will also be designed to account for possible measurement errors. However, this is not the focus of the current article, so in keeping with our focus on the Gaussian kernel and in the interest of simplicity we have interpolated the given data using infinitely smooth Gaussians. The rather large shape parameter obtained for this example by LOOCV indicates that the basis functions used in (4) are very localized and thus the choice of the Gaussian as kernel for this application is most likely not an appropriate one. In the recent paper [139] Michael Scheuerer, Robert Schaback and Martin Schlather investigated this issue and asked the question "Interpolation of spatial data – a stochastic or a deterministic problem?". In particular, they made some progress on the issue of how to decide whether a specific kernel is "correctly chosen" for a given application.

The curious reader may want to compare our interpolated surface to a photograph of Mt. Eden such as the one that can be found at `http://www.teara.govt.nz/en/volcanoes/3/5/2` as well

as to some of the surfaces presented in [139]. The latter comparison is not really a fair one since we used 10% of the data, while [139] used only $N = 300$ points for the interpolation and reserved the remainder of the data to compute the root mean square error of the interpolant. It may be surprising that even with the full data set a Gaussian kernel fit (using even narrower basis functions with an LOOCV-optimal shape parameter of $\varepsilon = 64.360659$) does not match the photograph as well as the fit obtained in [139] with a more flexible (and less smooth) Matérn kernel whose variance, smoothness and shape parameters were chosen using either maximum likelihood or LOOCV estimation.

Problems in geophysical or spatial statistics are only one type of situations to which the scattered data fitting formulation applies. Another important application arises in the design and analysis of computer experiments. The book [52] is a good introduction to this topic since its authors emphasize the use of positive definite kernels in the context of kriging and support vector machine algorithms. As we will see in Section 8.1, the scattered data fitting method can also be nicely generalized to the numerical solution of partial differential equations via collocation. Perhaps the most significant point of the computer experiment and PDE collocation applications for numerical analysts is the fact that in these cases the data can generally be assumed to be free of measurement noise, so the above formulation does indeed directly apply.

3.2 The Mairhuber-Curtis Theorem

In [129] Robert Schaback provided the book [16] by Dietrich Braess as a reference for the so-called *Mairhuber-Curtis theorem*. Such a theorem, based on the work [94, 31] of John Mairhuber (a student of Iso Schoenberg's) and Philip Curtis – both shown in Figure 13 – has been precisely formulated for the purposes of multivariate scattered data interpolation and proved in, e.g., the book by Holger Wendland [155]. The implications of the Mairhuber-Curtis theorem can be succinctly summarized in

Sound bite 3. *"The linear function space used for multivariate interpolation should be data-dependent."*

Figure 13: John Mairhuber (left) and Philip Curtis (right).

We provide an illustration of the proof and implications of this theorem in the movies MairhuberCurtis1.avi and MairhuberCurtis2.avi that are included with this paper. These movies are screen captures of some explorations performed with a Mathematica notebook that will has been posted online [40]. A screen shot of the Mathematica module is shown in Figure 14.

In MairhuberCurtis1.avi, the left part of the screen shows nine data sites, $\boldsymbol{x}_i$, $i = 1, \ldots, 9$, located in the unit square and the right part of the screen shows nine corresponding, but *fixed*, basis functions B_j, $j = 1, \ldots, 9$, that happen to be Gaussian kernels. One then sees the mouse pick

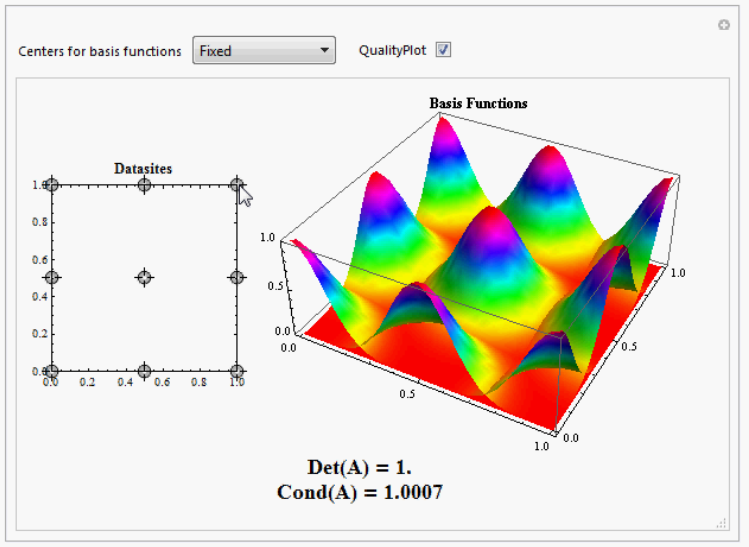


Figure 14: Screen shot of the Mathematica module used to illustrate the proof and implications of the Mairhuber-Curtis theorem.

up and interchange the positions of two of the data sites (while the basis functions, which are *not* connected with the data sites, remain at their fixed positions). As a consequence, the determinant of the interpolation matrix with entries $B_j(\boldsymbol{x}_i)$ changes sign since we have interchanged two of its rows. By continuity arguments we can therefore conclude that a configuration of data sites $\boldsymbol{x}_i$ exists for which – for the given and fixed basis – the interpolation matrix is singular, and therefore the multivariate scattered data problem is not well-posed. This is in contrast to the well-known univariate situation, where one can always fix a basis (such as polynomials of degree N) to uniquely interpolate any data specified at a set of $N+1$ distinct points on the real line.

The message for multivariate interpolation is therefore the one given by Sound bite 3, and the fact that a data-dependent basis does indeed "work" is illustrated in MairhuberCurtis2.avi. One can see that the mode of operation is switched to using a data-dependent basis, and that the basis functions are no longer fixed since they are directly tied to the data sites via $B_j = K(\cdot, \boldsymbol{x}_j)$ with a Gaussian kernel K. This means that as soon as the mouse moves one of the data sites, the basis function that is now centered at this point follows along. Furthermore, we can observe that the determinant of the interpolation remains positive throughout the entire switching of positions, thus it seems plausible that the matrix K is always non-singular (a fact guaranteed by Definition 2 as long as K is a positive definite kernel).

Therefore, we will always approach the multivariate kernel-based scattered data interpolation problem with a *data-dependent linear function space* that provides an interpolant of the form

$$s_f(\boldsymbol{x}) = \sum_{j=1}^{N} c_j K(\boldsymbol{x}, \boldsymbol{x}_j), \qquad \boldsymbol{x} \in \Omega \subseteq \mathbb{R}^d, \tag{5}$$

where $K : \Omega \times \Omega \to \mathbb{R}$ is a positive definite reproducing kernel. As already explained above, we find the coefficients c_j by solving the interpolation equations

$$s_f(\boldsymbol{x}_i) = y_i, \quad i = 1, \ldots, N.$$

This leads to a linear system $\mathsf{K}\boldsymbol{c} = \boldsymbol{y}$ with symmetric positive definite *kernel matrix* (or interpolation matrix) K whose entries are determined by the kernel K, i.e.,

$$\mathsf{K}_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j), \qquad i, j = 1, \ldots, N.$$

In addition to the well-posedness of positive definite kernel-based multivariate scattered data interpolation just discussed, the theory of reproducing kernels, or more generally that of *optimal approximation* or *optimal recovery* (see, e.g., [59, 129]), provides a number of *optimality properties* of the kernel-based interpolant. For example, it has minimum norm in the native Hilbert space $\mathcal{H}(K, \Omega)$ among all interpolants from $\mathcal{H}(K, \Omega)$. Moreover, assuming that the data y_i are sampled from a function $f \in \mathcal{H}(K, \Omega)$, i.e., $y_i = f(\boldsymbol{x}_i)$, the kernel-based interpolant is the best approximation of f from $\mathcal{H}(K, \Omega)$ measured in the corresponding Hilbert space norm. We will not discuss these properties any further, but instead refer the reader to the original literature or the books [38, 155]. The recent paper [139] of Scheuerer, Schaback and Schlather places these optimality properties side-by-side with a list of corresponding properties obtained when one views the scattered data fitting problem from a stochastic Gaussian process point of view, i.e., when one applies kriging methods.

4 Data-dependent Basis Functions

Up until now we have mostly focused on the positive definite kernel K and the generally infinite-dimensional Hilbert space $\mathcal{H}(K, \Omega)$ associated with it. However, once we focus on a specific scattered data fitting problem, i.e., once we fix a finite set (of data sites) $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{R}^d$ and an associated data-dependent linear function space $H(K, \mathcal{X}) = \operatorname{span}\{K(\cdot, \boldsymbol{x}_1), \ldots, K(\cdot, \boldsymbol{x}_N)\}$ as outlined in the previous section, then it also makes sense to consider different bases for the *finite-dimensional kernel space* $H(K, \mathcal{X})$.

Remark. Throughout this paper we always assume that the data sites have been chosen for us. This assumption is by no means a natural one since in applications such as the design of experiments an important part of the problem usually lies in determining a "good" *design*, i.e., a good choice of data sites. The error bounds for kernel interpolation given later also depend on the specific choice of data sites, and therefore a choice that might minimize the error bound is certainly desirable. If one, however, includes the data sites as variables in the data fitting problem, then one ends up with a *nonlinear* problem and we want to avoid that discussion here. In fact, to our knowledge, a satisfactory theoretical approach to this problem does not yet exist. This is in contrast to the multivariate integration problem, where so-called *low discrepancy* point sets have been studied for a long time. The paper [35] of Robert Schaback with Stefano DeMarchi and Holger Wendland has provided some initial progress in this direction for kernel-based interpolation.

4.1 Standard Basis Functions

The most common approach used for the solution of the scattered data interpolation problem is to employ the standard basis

$$\{K(\cdot, \boldsymbol{x}_1), \ldots, K(\cdot, \boldsymbol{x}_N)\}.$$

In particular, this has led to the widely used radial basis functions of the form $K(\|\cdot - \boldsymbol{x}_j\|)$. The main criticism of radial basis functions has been rooted in the facts that (1) the kernel matrices K are often *dense matrices*, and (2) these matrices are often *ill-conditioned*. We do not want to focus on item (1) is this paper, although this is a very important issue and there exist a number of

approaches (most of them iterative in nature) that address this problem. Instead we want to look at a few options for dealing with item (2). The recent preprint by Maryam Pazouki and Robert Schaback [114] is an excellent reference that provides a very nice discussion of this topic.

Figure 15 shows three standard Gaussian basis functions, all translates of the same basic kernel $K(\boldsymbol{x}-\boldsymbol{z}) = \mathrm{e}^{-\varepsilon^2\|\boldsymbol{x}-\boldsymbol{z}\|^2}$ to different centers $\boldsymbol{z}$ chosen to lie at a corner, center and edge midpoint of the unit square, respectively. A value of $\varepsilon = 3$ was used for the plots. The fact that the kernel is cut off at the edge of the square distorts the fact that all three basis functions are simply translates of the same kernel.

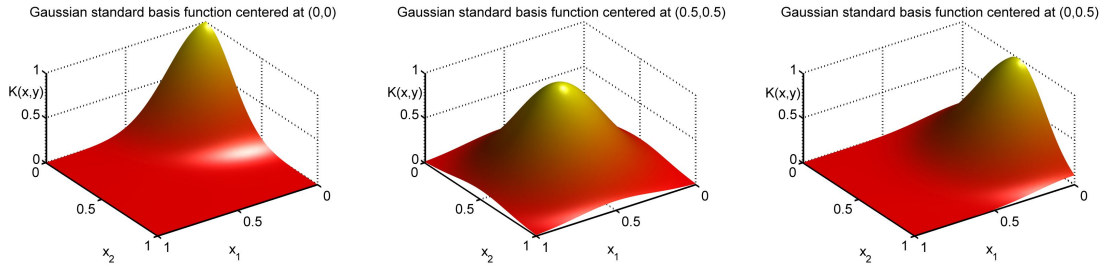


Figure 15: Standard Gaussian basis functions centered at different points in the unit square.

4.2 Cardinal Basis Functions

In many ways, the *ideal* basis is given by a *Lagrange* or *cardinal basis*. Such functions satisfy the *Lagrange property*

$$\boldsymbol{u}_j^*(\boldsymbol{x}_i) = \delta_{ij}, \qquad i,j = 1,\ldots,N, \tag{6}$$

so that we can find them (in a pointwise sense) by solving the linear system

$$\mathsf{K}\boldsymbol{u}^*(\boldsymbol{x}) = \boldsymbol{k}(\boldsymbol{x}), \tag{7}$$

for any given evaluation point $\boldsymbol{x} \in \mathbb{R}^d$. Here we again use the kernel matrix K with entries $\mathsf{K}_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and the right-hand side vector is given by the standard basis functions, i.e., $\boldsymbol{k} = [K(\cdot,\boldsymbol{x}_1),\ldots,K(\cdot,\boldsymbol{x}_N)]^T$.

If available (such as, e.g., in [18] for polyharmonic splines and multiquadrics on infinite regular grids) cardinal functions trivialize the interpolation problem since the interpolant can then be written in the form

$$s_f(\boldsymbol{x}) = \sum_{j=1}^{N} y_j \boldsymbol{u}_j^*(\boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega \subseteq \mathbb{R}^d, \tag{8}$$

i.e., the interpolation matrix for this basis is an identity matrix by virtue of (6).

Cardinal functions are also very useful for theoretical purposes such as establishing error bounds for kernel-based scattered data interpolation as was done in the seminal paper [157] by Zongmin Wu and Robert Schaback. Other work on infinite lattices has been aided by the use of the Poisson summation formula which enables one to derive explicit representations of cardinal functions. For the Gaussian kernel this has been done, e.g., in [64], but there is also earlier work in, e.g., [5], and in [14, 99] one can find so-called approximate cardinal functions for the Gaussian. In [115] cardinal functions are provided for the Gaussian kernel at a set of points in a bounded interval, i.e., for

$d = 1$. If one has an explicit formula for the cardinal functions, then one immediately obtains information about the *Lebesgue constant*

$$\Lambda_{K,\mathcal{X}} = \max_{\boldsymbol{x}\in\Omega} \sum_{j=1}^{N} |\boldsymbol{u}_j^*(\boldsymbol{x})|,$$

which in turn provides information about *accuracy* and *stability* of a kernel interpolant via

$$\|f - s_f\| \leq (1 + \Lambda_{K,\mathcal{X}}) \, \|f - s^*\|,$$

where s^* is the L_∞-best approximation of f from the finite-dimensional space $H(K, \mathcal{X})$ and

$$\|s_f\|_\infty \leq \Lambda_{K,\mathcal{X}} \|\boldsymbol{y}\|_\infty,$$

where s_f is the kernel interpolant to the data $\boldsymbol{y}$ sampled from f on $\mathcal{X}$ so that the norm on the left is continuous and that on the right discrete. Unfortunately, the cardinal functions and associated Lebesgue constants are known only for some special cases. Some results on the behavior of the Lebesgue constant can be found in the paper [34] by Stefano De Marchi and Robert Schaback.

Figure 16 shows three Gaussian cardinal basis functions again identified with a corner, center and edge midpoint of the unit square, respectively. It is clear that these basis functions are no longer shifted copies of one single basic function. In fact, cardinal basis functions are inherently tied to the set $\mathcal{X}$ of data sites and the domain Ω. Nevertheless, if many interpolation problems with the same set $\mathcal{X}$ need to be solved, then one might consider precomputing the cardinal basis. A value of $\varepsilon = 1$ was used for these plots.

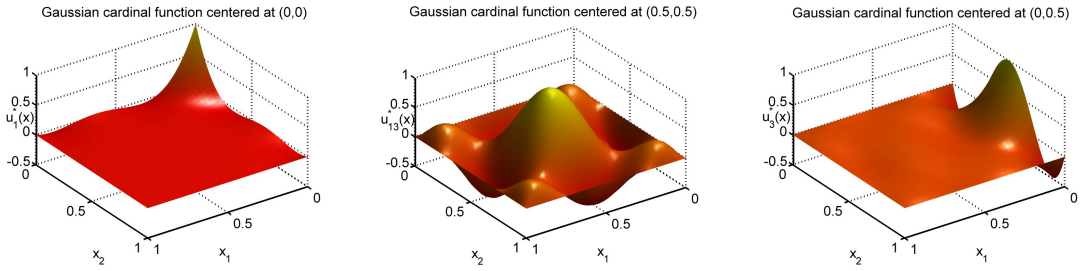


Figure 16: Gaussian cardinal basis functions centered at different points in the unit square.

4.3 Newton-type Basis Functions

In the setting of polynomial interpolation it is well known that the *Newton form* of the basis is "in-between" the Lagrange form and the monomial basis (which corresponds to our standard kernel basis) in the sense that the interpolation matrix associated with the Newton form is triangular.

In the field of kernel-based interpolation such a basis was missing for a long time, and only recently Stefan Müller in his Ph.D. thesis [107] under the direction of Robert Schaback (see also [108]) discovered such a basis and studied it in detail.

These Newton-type basis functions satisfy a *Newton property*, i.e.,

$$\boldsymbol{v}_j^*(\boldsymbol{x}_i) = \delta_{ij}, \qquad 0 \leq i \leq j \leq N.$$

One of the nice insights recently obtained in [107, 108] (and also [114] as well as Stefano DeMarchi's paper [32] in this issue) is the fact that a Newton-type basis can be computed via a pivoted LU-decomposition of the kernel matrix K. Moreover, the Newton basis provides an *orthogonal basis for*

the native Hilbert space $\mathcal{H}(K,\Omega)$. This latter fact may lead to many new algorithms for the efficient and stable computation of kernel-based interpolants, some of which have already been introduced in the papers just mentioned.

Figure 17 shows three Gaussian Newton-type basis functions once again identified with a corner, center and edge midpoint of the unit square, respectively. Clearly, these basis functions also cannot be obtained as shifted copies of a single basic function. A value of $\varepsilon = 3$ was used for these plots.

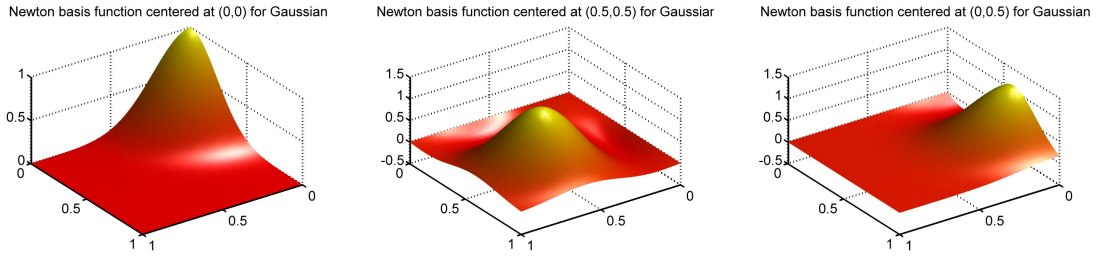


Figure 17: Gaussian Newton-type basis functions centered at different points in the unit square.

5 Accuracy of Kernel Interpolation

5.1 Standard Error Bounds

We now assume that the data y_i are sampled from a function $f \in \mathcal{H}(K,\Omega)$, so that $y_i = f(\boldsymbol{x}_i)$, $i = 1,\ldots,N$ and discuss the error between the kernel interpolant s_f and f measured in two different ways, a pointwise error, which we will discuss in this section, and a weighted L_2 error, which we will discuss in Section 5.4.

The following standard pointwise error bound for reproducing kernel interpolation probably was introduced into the RBF literature by Zongmin Wu and Robert Schaback in [157]. One can easily obtain it using the representation of the kernel interpolant s_f in terms of the cardinal basis (8), the reproducing property of the kernel and the Cauchy-Schwarz inequality. In fact, using the representation $\boldsymbol{u}^*(\boldsymbol{x}) = \mathsf{K}^{-1}\boldsymbol{k}(\boldsymbol{x})$ of the cardinal functions (7) we have that for any $\boldsymbol{x} \in \Omega$

$$
\begin{aligned}
|f(\boldsymbol{x}) - s_f(\boldsymbol{x})| &= \left| f(\boldsymbol{x}) - \sum_{j=1}^{N} f(\boldsymbol{x}_j) u_j^*(\boldsymbol{x}) \right| = \left| \langle f, K(\cdot,\boldsymbol{x})\rangle_{\mathcal{H}(K,\Omega)} - \sum_{j=1}^{N} \langle f, K(\cdot,\boldsymbol{x}_j)\rangle_{\mathcal{H}(K,\Omega)} u_j^*(\boldsymbol{x}) \right| \\
&= \left| \langle f, K(\cdot,\boldsymbol{x}) - \sum_{j=1}^{N} K(\cdot,\boldsymbol{x}_j) u_j^*(\boldsymbol{x})\rangle_{\mathcal{H}(K,\Omega)} \right| = \left| \langle f, K(\cdot,\boldsymbol{x}) - \boldsymbol{k}^T(\cdot)\mathsf{K}^{-1}\boldsymbol{k}(\boldsymbol{x})\rangle_{\mathcal{H}(K,\Omega)} \right| \\
&\le \|f\|_{\mathcal{H}(K,\Omega)} \left\| K(\cdot,\boldsymbol{x}) - \boldsymbol{k}^T(\cdot)\mathsf{K}^{-1}\boldsymbol{k}(\boldsymbol{x}) \right\|_{\mathcal{H}(K,\Omega)} = \|f\|_{\mathcal{H}(K,\Omega)} P_{K,\mathcal{X}}(\boldsymbol{x}),
\end{aligned}
$$

with $\boldsymbol{k}(\cdot) = [K(\cdot,\boldsymbol{x}_1),\ldots,K(\cdot,\boldsymbol{x}_N)]^T$, and where the term

$$
P_{K,\mathcal{X}}(\boldsymbol{x}) = \sqrt{K(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{k}^T(\boldsymbol{x})\mathsf{K}^{-1}\boldsymbol{k}(\boldsymbol{x})}
$$

was called the *power function* by Robert Schaback in [125].

Another optimality property of the kernel-based interpolant is given by the fact that it minimizes the pointwise error just introduced (see, e.g., [155]). In the stochastic literature, i.e., kriging,

Gaussian processes, etc. (see e.g., [52, 117, 146]), the power function appears as the *kriging variance* and it turns out that one obtains the *best linear predictor* of the random field with covariance kernel K by fitting the samples $y_1, \ldots, y_N$ of a sample path f of the random field by the same interpolation procedure described in the deterministic setting above. This analogy is also discussed in the recent paper by Michael Scheuerer, Robert Schaback and Martin Schlather [139].

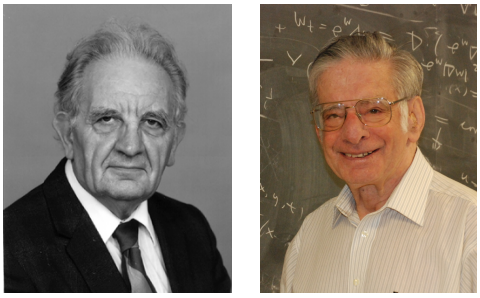


Figure 18: Michael Golomb (left) and Hans Weinberger (right).

The above error bound can be improved (as was already suggested in [59] by Michael Golomb and Hans Weinberger, who are shown in Figure 18) to

$$|f(\boldsymbol{x}) - s_f(\boldsymbol{x})| \leq \|f - s_f\|_{\mathcal{H}(K,\Omega)} P_{K,\mathcal{X}}(\boldsymbol{x}) \tag{9}$$

since $f - s_f$ is orthogonal (in the Hilbert space inner product) to both f and s_f. This latter (tighter) error bound does not seem to play a significant role in the RBF literature.

Since $\|f\|_{\mathcal{H}(K,\Omega)}$ usually is not computable (remember, we do not even know f, but want to reconstruct it from the data) the standard error bound is not very useful for practical situations. On the other hand, if we assume that our approximation s_f is not too bad, i.e., $\|f - s_f\|_{\mathcal{H}(K,\Omega)} \leq \delta \|s_f\|_{\mathcal{H}(K,\Omega)}$ for some not too large constant δ, then we obtain a *computable error bound*

$$|f(\boldsymbol{x}) - s_f(\boldsymbol{x})| \leq \delta \|s_f\|_{\mathcal{H}(K,\Omega)} P_{K,\mathcal{X}}(\boldsymbol{x}) \tag{10}$$

since it is easy to see that $\|s_f\|_{\mathcal{H}(K,\Omega)} = \sqrt{\boldsymbol{y}^T \mathsf{K}^{-1} \boldsymbol{y}}$, where $\boldsymbol{y} = [y_1, \ldots, y_N]^T$ is the vector of data values:

$$\begin{aligned} \|s_f\|^2_{\mathcal{H}(K,\Omega)} &= \langle \boldsymbol{y}^T \boldsymbol{u}^*(\cdot), \boldsymbol{y}^T \boldsymbol{u}^*(\cdot) \rangle_{\mathcal{H}(K,\Omega)} = \langle \boldsymbol{y}^T \mathsf{K}^{-1} \boldsymbol{k}(\cdot), \boldsymbol{y}^T \mathsf{K}^{-1} \boldsymbol{k}(\cdot) \rangle_{\mathcal{H}(K,\Omega)} \\ &= \boldsymbol{y}^T \mathsf{K}^{-1} \langle \boldsymbol{k}(\cdot), \boldsymbol{k}(\cdot) \rangle_{\mathcal{H}(K,\Omega)} \mathsf{K}^{-1} \boldsymbol{y} = \boldsymbol{y}^T \mathsf{K}^{-1} \mathsf{K} \mathsf{K}^{-1} \boldsymbol{y}. \end{aligned}$$

We will come back to the error bound (10) in Section 5.2 when we discuss how to choose a good value for the shape parameter ε which appears in the definition of many kernels.

In the RBF literature (for a summary of results see, e.g., [38, 155]) the generic error bound is adapted to specific kernels K and usually expressed in terms of some *fill distance* or *meshsize* h. More recent techniques (see, e.g., the recent survey by Christian Rieger, Robert Schaback and Barbara Zwicknagl [119] or Christian Rieger's presentation [118] at the Göttingen workshop) do not work with the power function, but use so-called *sampling inequalities* to arrive at similar error bounds.

Example 7. For well-distributed (quasi-uniform) data we can replace the fill distance h by $N^{-1/d}$ without affecting the order of an error bound. Then the error bounds for the Gaussian kernel published in the literature (such as [93, 120, 154]) are of the form

$$\|f - s_f\|_\infty \leq C_d N^{-p/d} \|f\|_{\mathcal{H}(K,\Omega)}.$$

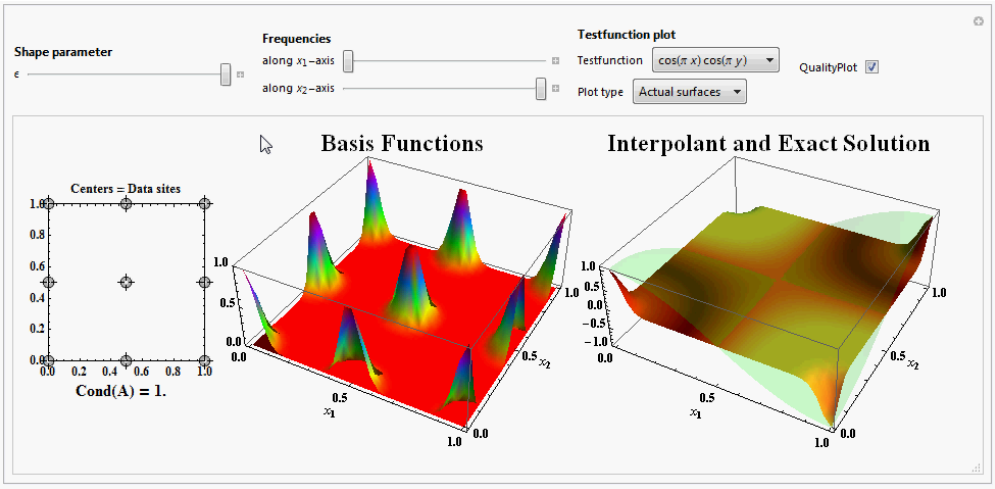


Figure 19: Screen shot of the Mathematica module used to illustrate effects of the Gaussian shape parameter.

This can be interpreted in the sense that Gaussian kernels provide *arbitrarily high approximation order* to infinitely smooth functions f, i.e., with $p = \infty$. However, one also sees that the rate of convergence decreases with increasing dimension, and what is less apparent, the constant C_d grows with d. This means that such bounds are *dimension-dependent* and are likely to suffer from the *curse of dimensionality*. This may not be of concern for low-dimensional applications such as those that occur in terms of the three physical dimensions surrounding us. However, if we are dealing with interpolation problems that arise in a computer experiment, then the number of variables is likely to be much higher, and dimension-dependence does matter.

5.2 Choosing a Good Shape Parameter

Up until now we have not said much about the shape parameter ε that appears in the definition of the Gaussian kernel (2). Other kernels have similar shape parameters, and what we are about to discuss for the Gaussian applies to those kernels in an analogous way.

We begin with the description of another movie, IsotropicGaussianInterpolation.avi, that shows how the shape parameter affects the basis functions, the condition number of the kernel matrix K and the accuracy of the kernel interpolant s_f. This movie is also a screen capture of explorations performed with a Mathematica notebook that has been posted online [40]. A screen shot of the Mathematica module is shown in Figure 19.

The screen shows three plots. The left-most plot displays the data sites and coinciding centers of the basis functions. Underneath this plot one also can monitor the L_2 condition number of the interpolation matrix K. In the middle plot we see the associated Gaussian basis functions. Both of these plots are analogous to those in the Mairhuber-Curtis movies. The right-most plot shows the interpolant s_f (in a goldish color) as well as the function f (shown in a partially transparent blueish color) that we are trying to recover from its samples obtained at the data sites shown in the left-most plot. In interactive mode, the sliders at the top can be used to change the (common) shape parameter ε of the basis functions, pick different test functions (which can be parametrized differently along the x and y-coordinate axes), and switch the right-most plot from the interpolating surface view to an error view.

The movie begins with the user decreasing the value of the shape parameter. This has several consequences:

- The basis functions become wider, or "flatter".
- The condition number of the matrix K increases.
- The interpolant s_f resembles the test function f more accurately.

None of these features were discussed in earlier parts of this paper, but are well-known in the literature.

About half-way through the movie, the right-most plot is switched to a view that displays the interpolation error instead of the actual surfaces. This reveals where the error is largest, and subsequently a few more points are added in the left-most plot. Again, this has a number of consequences:

- The shape of the basis functions does not change.
- The condition number of the matrix K increases even more.
- The error becomes even smaller.

These latter observations are in line with the error bounds discussed earlier, and the effects on the condition number are also well known (we will come back to that in Section 7).

Clearly, in addition to sampling the function f at sufficiently many and well-chosen points, an appropriate choice of the shape parameter is important, and we will now say a few words about that.

For the longest time, people in the approximation theory and numerical analysis community went mostly with ad-hoc choices of the shape parameter or ignored its effect by treating it as a constant. Much more systematic approaches have been suggested in the statistics literature for a long time (see, e.g., [151] and many other references). In the radial basis community one can find papers such as [44, 67, 121, 137] that employ some of the methods we are about to explain. An extended discussion of a "good" shape parameter choice was also included in [38], but we would now like to add some more recent insights based on [41, 66, 105].

5.2.1 Leave-One-Out Cross Validation (LOOCV)

This method was first proposed in the context of ridge regression or smoothing splines by Allen [1] and Wahba and collaborators [27, 60], respectively, to find the optimal *smoothing parameter* μ in the linear system

$$(\mathsf{K} + \mu \mathsf{I})\boldsymbol{c} = \boldsymbol{y}.$$

This system corresponds to a fundamentally different – but closely related – data fitting problem since the data for this problem are assumed to be contaminated with noise so that one does not want to interpolate, but instead seeks a *smoothed fit*.

Rippa [121] used *leave-one out cross validation* (LOOCV) to select the shape parameter of a radial basis interpolation system, but everything he did can be transferred to more general positive definite kernels. The idea behind LOOCV is to split the data into two different sets: the *training set* $\{y_1, \ldots, y_{\ell-1}, y_{\ell+1}, \ldots, y_N\}$, and the *validation set* consisting of only the single y_ℓ (which was

left out when creating the training set). Now, for a fixed $\ell \in \{1, \ldots, N\}$ and fixed ε, one defines the partial interpolant

$$s_f^{[\ell,\varepsilon]}(\boldsymbol{x}) = \sum_{\substack{j=1 \\ j\neq\ell}}^{N} c_j^{[\ell,\varepsilon]} K_\varepsilon(\boldsymbol{x}, \boldsymbol{x}_j)$$

whose coefficients $c_j^{[\ell,\varepsilon]}$ are determined by interpolating the training data, i.e.,

$$s_f^{[\ell,\varepsilon]}(\boldsymbol{x}_i) = y_i, \qquad i = 1, \ldots, \ell-1, \ell+1, \ldots, N.$$

In order to measure the quality of this attempt one defines

$$e_\ell(\varepsilon) = y_\ell - s_f^{[\ell,\varepsilon]}(\boldsymbol{x}_\ell),$$

which denotes the error at the one validation point $\boldsymbol{x}_\ell$ not used to determine the interpolant. The "optimal" value of ε is found as

$$\varepsilon_{opt} = \operatorname*{argmin}_{\varepsilon} \|\boldsymbol{e}(\varepsilon)\|, \qquad \boldsymbol{e} = [e_1, \ldots, e_N]^T,$$

where some norm is used in this minimization problem. Usually the 2-norm is chosen, but Rippa also discussed (and preferred) the 1-norm.

The important fact recognized by Rippa (and earlier also by Wahba) is that one can compute the error vector *without* solving N problems each of size $(N-1) \times (N-1)$. Instead, everything can be expressed in terms of the full interpolation matrix K_ε, i.e.,

$$e_\ell(\varepsilon) = \frac{c_\ell}{\mathsf{K}_{\ell\ell}^{-1}},$$

where c_ℓ is the ℓ^{th} coefficient of the full interpolant s_f and $\mathsf{K}_{\ell\ell}^{-1}$ is the ℓ^{th} diagonal element of the inverse of corresponding $N \times N$ interpolation matrix K_ε.

The *generalized cross-validation* method of [27] replaces the LOOCV cost function

$$\text{LOOCV}(\varepsilon) = \|\boldsymbol{e}(\varepsilon)\|_2 = \sqrt{\sum_{\ell=1}^{N} \left(\frac{c_\ell}{\mathsf{K}_{\ell\ell}^{-1}} \right)^2}$$

with

$$\text{GCV}(\varepsilon) = \sqrt{\sum_{\ell=1}^{N} \left(\frac{c_\ell}{\frac{1}{N}\sum_{j=1}^{N} \mathsf{K}_{jj}^{-1}} \right)^2} = \frac{N\|\boldsymbol{c}\|_2}{\text{trace}(\mathsf{K}_\varepsilon^{-1})} = \sqrt{\boldsymbol{y}^T \mathsf{K}_\varepsilon^{-2} \boldsymbol{y}}\, \mu_h\left(\lambda(\mathsf{K}_\varepsilon)\right),$$

i.e., the individual diagonal elements of $\mathsf{K}_\varepsilon^{-1}$ have been replaced by their averages. The notation $\mu_h(\lambda(\mathsf{K}_\varepsilon))$ denotes the *harmonic mean* of the eigenvalues of K_ε, and the relation $\|\boldsymbol{c}\|_2 = \sqrt{\boldsymbol{y}^T \mathsf{K}_\varepsilon^{-2} \boldsymbol{y}}$ follows immediately from the interpolation system $\mathsf{K}_\varepsilon \boldsymbol{c} = \boldsymbol{y}$.

5.2.2 Maximum Likelihood Estimation

The recent papers [137, 139] by Michael Scheuerer, Robert Schaback and Martin Schlather provide a discussion of *maximum likelihood estimation* (MLE) as it can be applied to kernel-based interpolation methods. This method rests on a solid statistical foundation.

We assume that our unknown function f is a realization of a *random field* Z, itself a collection of random variables, i.e., $Z = \{Z(\boldsymbol{x}),\ \boldsymbol{x} \in \Omega \subseteq \mathbb{R}^d\}$. This implies that the data $y_i = f(\boldsymbol{x}_i)$ are viewed as realizations of the random variables $Z(\boldsymbol{x}_i)$, $i = 1, \ldots, N$. Usually we assume that the random field is *Gaussian*, i.e., the N-dimensional random vectors $Z(\boldsymbol{x}) = [Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_N)]^T \in \mathbb{R}^N$ are *normally distributed* (or *Gaussian distributed*), i.e.,

$$Z(\boldsymbol{x}) \sim \mathcal{N}(\boldsymbol{\mu}, \mathsf{K})$$

with *mean* $\boldsymbol{\mu} = [\mathbb{E}[Z(\boldsymbol{x}_1)], \ldots, \mathbb{E}[Z(\boldsymbol{x}_N)]]^T$ and *covariance matrix*

$$\mathsf{K} = (\mathrm{Cov}[Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j)])_{i,j=1}^N .$$

In *simple kriging* we further assume that we are working with a *zero-mean* Gaussian field, i.e., $\boldsymbol{\mu} = \mathbf{0}$. Then $K(\boldsymbol{x}, \boldsymbol{z}) = \mathrm{Cov}[Z(\boldsymbol{x}), Z(\boldsymbol{z})] = \mathbb{E}[Z(\boldsymbol{x})Z(\boldsymbol{z})]$, where K is a positive definite covariance kernel.

If the probability density for an event for which we observe the data $\boldsymbol{y}$, given the parameter ε, is written as $\rho(\boldsymbol{y}|\varepsilon)$, then the $\mathcal{L}(\varepsilon|\boldsymbol{y})$ characterizes the *likelihood* of the parameter ε given the data $\boldsymbol{y}$.

The (Gaussian) joint *probability density function* of the vector $Z(\boldsymbol{x})$ of N random observations with covariance matrix K_ε can be written as

$$\rho(Z(\boldsymbol{x})|\varepsilon) = \rho(Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_N)|\varepsilon) = \frac{1}{\sqrt{(2\pi)^N \det(\mathsf{K}_\varepsilon)}} \mathrm{e}^{-\frac{1}{2}(Z(\boldsymbol{x})-\boldsymbol{\mu})^T \mathsf{K}_\varepsilon^{-1} (Z(\boldsymbol{x})-\boldsymbol{\mu})}.$$

If we evaluate this density function using the data vector $\boldsymbol{y} = [y_1, \ldots, y_N]^T$ we obtain the *log-likelihood function* (now for simplicity again assuming $\boldsymbol{\mu} = \mathbf{0}$)

$$\begin{aligned} \ln \mathcal{L}(\varepsilon|\boldsymbol{y}) &= \ln \left(\frac{\mathrm{e}^{-\frac{1}{2}\boldsymbol{y}^T \mathsf{K}_\varepsilon^{-1} \boldsymbol{y}}}{\sqrt{(2\pi)^N \det(\mathsf{K}_\varepsilon)}} \right) \\ &= -\frac{1}{2} \ln(\det(\mathsf{K}_\varepsilon)) - \frac{1}{2} \boldsymbol{y}^T \mathsf{K}_\varepsilon^{-1} \boldsymbol{y} + \text{constants}. \end{aligned}$$

The best shape parameter ε will be obtained by maximizing this log-likelihood function, since then observation of the data $\boldsymbol{y}$ is most likely under the assumed stochastic model defined by the family of covariance kernels K_ε parametrized by ε. Alternatively, we can minimize (twice) the negative log-likelihood function. In order to have a criterion invariant under an additional rescaling of the kernel in terms of a process variance parameter we can derive (see [66]) the *maximum likelihood criterion*

$$\mathrm{MLE}(\varepsilon) = \frac{1}{N} \ln(\det(\mathsf{K}_\varepsilon)) + \ln\left(\boldsymbol{y}^T \mathsf{K}_\varepsilon^{-1} \boldsymbol{y}\right).$$

Alternatively, we could also consider

$$\mathrm{GMLE}(\varepsilon) = \sqrt[N]{\det(\mathsf{K}_\varepsilon)}\left(\boldsymbol{y}^T \mathsf{K}_\varepsilon^{-1} \boldsymbol{y}\right) = \mu_g\left(\lambda(\mathsf{K}_\varepsilon)\right)\left(\boldsymbol{y}^T \mathsf{K}_\varepsilon^{-1} \boldsymbol{y}\right),$$

where $\mu_g\left(\lambda(\mathsf{K}_\varepsilon)\right)$ denotes the *geometric mean* of the eigenvalues of K_ε.

5.2.3 Other Shape Parameter Optimization Criteria

Based on the Golomb-Weinberger refined error estimate (10) we can determine a good value of the shape parameter by minimizing

$$\mathrm{EB}(\varepsilon) = \sqrt{\boldsymbol{y}^T \mathsf{K}_\varepsilon^{-1} \boldsymbol{y}} \| P_{K_\varepsilon, \mathcal{X}} \|_\infty,$$

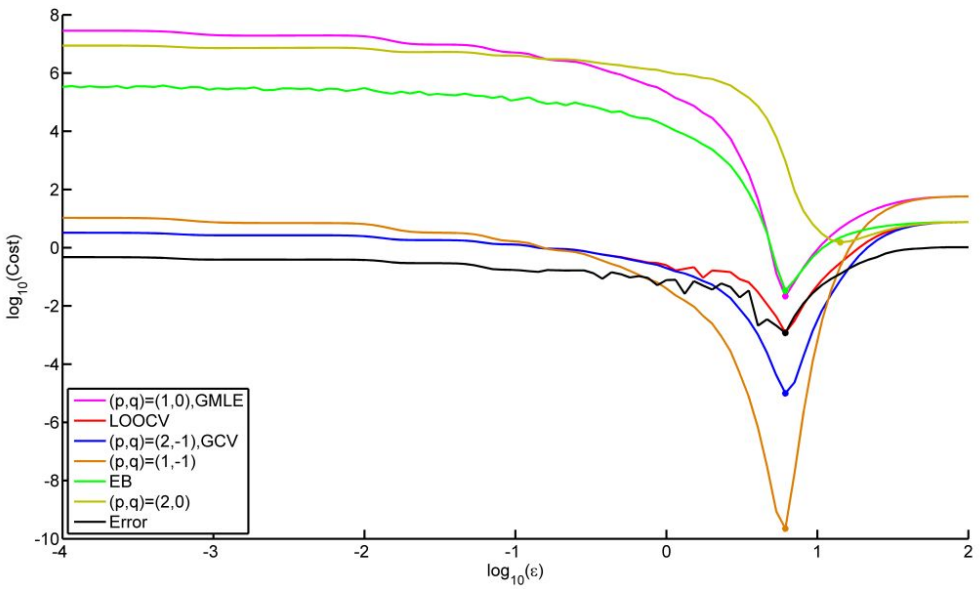


Figure 20: Comparison of different criteria to select an "optimal" shape parameter ε in Gaussian kernel interpolation.

where we compute the max-norm of the power function on a discrete evaluation grid with high resolution.

We can view GCV and GMLE as being at two ends of the spectrum in terms of the eigenvalues of K_ε: GCV is uses their harmonic mean, and GMLE the geometric mean. Using so-called *Hölder means* of the eigenvalues and p-type norms of the coefficient vector this can be generalized giving rise to a two-parameter family of shape parameter criteria:

$$\mathrm{Crit}_{p,q}(\varepsilon) = \left(\boldsymbol{y}^T \mathsf{K}_\varepsilon^{-p} \boldsymbol{y}\right)^{1/p} \left(\frac{1}{N} \sum_{\ell=1}^{N} \lambda_\ell^q(\mathsf{K}_\varepsilon) \right)^{1/q}$$

with GCV $= \mathrm{Crit}_{2,-1}$ and GMLE $= \mathrm{Crit}_{1,0}$. In particular, large eigenvalues are penalized for positive values of q and small eigenvalues for $q < 0$. Note also that $\left(\frac{1}{N}\sum_{\ell=1}^{N} \lambda_\ell^q(\mathsf{K}_\varepsilon)\right)^{1/q}$ corresponds to $\max\left(\lambda(\mathsf{K}_\varepsilon)\right)$ for $q = \infty$ and to $\min\left(\lambda(\mathsf{K}_\varepsilon)\right)$ for $q = -\infty$. This family of criteria was introduced in [41]. Figure 20 shows a comparison of different criteria for Gaussian kernel interpolation of the well-known Franke test function (see, e.g., [38]).

This figure suggests that – at least for this example – all but the $\mathrm{Crit}_{2,0}$ criterion perform equally well, i.e., they all locate the value of ε for which the actual error (black curve) is minimized. However, another important feature of RBF interpolation that we have not yet discussed is revealed by this figure as well: *there really is an optimal value of* ε. Our earlier observations based on the interpolation movie seemed to indicate that the error decreases – possibly monotonically – with ε. This example (and many others in the literature), show this is not so. As claimed (and observed) earlier, the conditioning of the interpolation matrix deteriorates with decreasing ε. Therefore, we use a regularization algorithm due to James Riley [122] that is discussed in [39] to stabilize the computations for small ε. Another, more effective way of dealing with the ill-conditioning of K_ε for small values of ε – actually circumventing this issue – will be discussed in Section 7.

Further investigations of shape parameter and kernel selection are reported in [105].

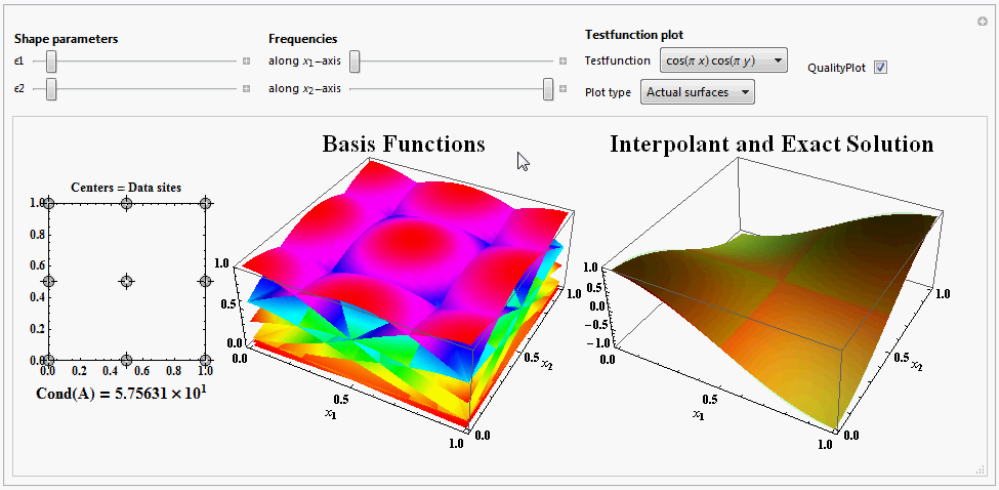


Figure 21: Screen shot of the Mathematica module used to illustrate anisotropic Gaussian interpolation.

5.3 Anisotropic Gaussian Interpolation

We have now seen how important it can be to select the single shape parameter of the Gaussian kernel (2) well.

To illustrate how much more flexibility is introduced by using *multiple shape parameters*, i.e., a different parameter for each coordinate direction, we discuss our last movie, AnisotropicGaussianInterpolation.avi. We emphasize that here we are interested with *anisotropic* kernels, i.e., different shape parameters for different coordinate directions. Another possible way of using multiple shape parameters would be to use spatially varying shape parameters. Such a strategy has been investigated by some researchers as well (see, e.g., [51, 105]). As for our other movies, this one is also a screen capture of explorations performed with a Mathematica notebook that has been posted online [40]. A screen shot of the Mathematica module is shown in Figure 21.

This movie more or less continues where IsotropicGaussianInterpolation.avi left off. The controls for this module are slightly refined as they now allow different shape parameters in the x_1 and x_2-coordinate directions.

We begin by creating a more anisotropic testfunction, i.e., the frequency of the cosine function is increased along the x_1-direction and decreased in x_2. This leads to a very poor surface fit using the initial nine data samples and relatively "flat" isotropic Gaussian kernels. The error plot shows that the error is largest along entire "tracks" parallel to the x_2-axis. We therefore add more points in the middle of the domain, parallel to the x_1-axis in order to increase the sampling in this direction with the goal of better capturing the higher frequency effects. As before, adding more points increases the condition number of K, but this time the error is not reduced markedly. We therefore change the shape parameters, creating anisotropic basis functions that are narrow in the x_1-direction (i.e., ε_1 is increased), and "flatter" in the x_2-direction (by decreasing the value of ε_2). This greatly improves the error and as a side-effect *also improves the condition number* of K. With these new anisotropic basis functions we no longer need multiple data points along "tracks" in the x_2-direction, so we remove some points and improve the condition number even further without affecting the error much. Finally, the last segment of the movie shows how poorly isotropic Gaussians would interpolate the data sampled along the central "track" selected earlier.

We remark that other point selection procedures might be even more effective for this example. One might, e.g., use Faure points, a set of well-distributed low discrepancy points frequently used in quasi-Monte Carlo methods (see, e.g., [58]). These points project roughly equally well in every coordinate direction.

5.4 Dimension-Independent Error Bounds

As pointed out in Section 5.1, the standard error bounds for interpolation with (isotropic) Gaussian kernels indicate a *dimension-dependent* convergence rate. In the recent paper [42] we investigated function approximation in a weighted L_2 sense using both isotropic and anisotropic Gaussian kernels in $\mathbb{R}^d$ and were able to derive *dimension-independent error bounds* for function approximation using N function values sampled from f of the form

$$\|f - s_f\|_{2,\rho} \leq \mathrm{err}_{2,\rho}^{wc} \|f\|_{\mathcal{H}(K,\mathbb{R}^d)} \quad \text{for all } f \in \mathcal{H}(K,\mathbb{R}^d),$$

where the *worst case error* decays like

$$\mathrm{err}_{2,\rho}^{wc} = \begin{cases} \mathcal{O}(N^{-1/4+\delta}) & \text{for isotropic Gaussians,} \\ \mathcal{O}(N^{-\max(\frac{\beta^2}{2+\beta},1/4)+\delta}) & \text{for anisotropic Gaussian provided } \varepsilon_\ell = \ell^{-\beta}. \end{cases}$$

Here δ is an arbitrarily small positive constant and β a possibly large positive constant indicating the decay of the shape parameters. This shows that we can have dimension-independent convergence rates both for isotropic and for anisotropic Gaussian kernels K provided the function f comes from the reproducing kernel Hilbert space of K. The (sufficient) requirement that the shape parameters decay like $\ell^{-\beta}$ translates into the fact that the function f we are trying to recover is *essentially low-dimensional*, i.e., the higher dimensions become less and less important (just as in the anisotropic interpolation movie above, where the contribution of f in the x_2-direction was insignificant). If we have sufficiently fast decay of the shape parameters, then the convergence rate essentially is arbitrarily high. The dimension-independent convergence rate we can guarantee for isotropic Gaussians, on the other hand, is rather slow. The fact that we are using finitely many samples, $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)$, on the infinite domain $\mathbb{R}^d$ is not problematic since the weight function ρ localizes the domain (see also Section 6.2).

The main tool used to prove these dimension-independent convergence rates is an eigenfunction expansion of the Gaussian kernel. That is discussed next.

6 Kernel Eigenfunction Expansions

A little over 100 years ago, infinite series expansions of the kernels of what are now called *Hilbert-Schmidt integral operators* were independently studied by Erhard Schmidt and James Mercer (see [140, 101] and also Figure 5). David Hilbert played a central role in this development since Schmidt was his Ph.D. student and Mercer states that his results were motivated by Hilbert's work [68].

6.1 Mercer's Theorem

We define the *Hilbert-Schmidt integral operator* $\mathcal{T}_K : L_2(\Omega,\rho) \to L_2(\Omega,\rho)$ on an open set $\Omega \subset \mathbb{R}^d$ with weight function ρ (which we will say more about below) by

$$(\mathcal{T}_K f)(\boldsymbol{x}) = \int_\Omega K(\boldsymbol{x},\boldsymbol{z}) f(\boldsymbol{z})\, \rho(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z}.$$

The operator $\mathcal{T}_K$ is compact if and only if $\iint_{\Omega\times\Omega} |K(\boldsymbol{x},\boldsymbol{z})|^2\rho(\boldsymbol{x})\rho(\boldsymbol{z})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{z} < \infty$ (see, e.g., [73]).

If the kernel K is symmetric, then according to *Mercer's theorem* we have a series expansion for the kernel K of the form

$$K(\boldsymbol{x},\boldsymbol{z}) = \sum_{n=0}^{\infty} \lambda_n \varphi_n(\boldsymbol{x})\varphi_n(\boldsymbol{z}),$$

where the λ_n and φ_n are the *eigenvalues* and $L_2(\Omega,\rho)$-orthonormal *eigenfunctions* of $\mathcal{T}_K$, i.e.,

$$(\mathcal{T}_K\varphi_n)(\boldsymbol{x}) = \lambda_n\varphi_n(\boldsymbol{x}) \quad \Longleftrightarrow \quad \langle K(\boldsymbol{x},\cdot),\varphi_n\rangle_{L_2(\Omega,\rho)} = \lambda_n\varphi_n(\boldsymbol{x}),$$

where

$$\langle \varphi_m,\varphi_n\rangle_{L_2(\Omega,\rho)} := \int_\Omega \varphi_m(\boldsymbol{x})\varphi_n(\boldsymbol{x})\rho(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \delta_{mn}.$$

6.2 Eigenfunctions for the Gaussian Kernel

In this paper we have focussed most of our examples on the Gaussian kernel. It is fortunate that in [117, Chapter 4] one can find an eigenfunction expansion for the Gaussian, i.e., for $x, z \in \mathbb{R}$ we have

$$\mathrm{e}^{-\varepsilon^2(x-z)^2} = \sum_{n=0}^{\infty} \lambda_n\varphi_n(x)\varphi_n(z), \tag{11}$$

where

$$\lambda_n = \frac{\alpha\varepsilon^{2n}}{\left(\frac{\alpha^2}{2}\left(1+\sqrt{1+\left(\frac{2\varepsilon}{\alpha}\right)^2}\right)+\varepsilon^2\right)^{n+\frac{1}{2}}}, \qquad n = 0,1,2,\ldots, \tag{12}$$

$$\varphi_n(x) = \frac{\sqrt[8]{1+\left(\frac{2\varepsilon}{\alpha}\right)^2}}{\sqrt{2^n n!}}\,\mathrm{e}^{-\left(\sqrt{1+\left(\frac{2\varepsilon}{\alpha}\right)^2}-1\right)\frac{\alpha^2x^2}{2}} H_n\left(\sqrt[4]{1+\left(\frac{2\varepsilon}{\alpha}\right)^2}\,\alpha x\right). \tag{13}$$

The eigenfunctions $\{\varphi_n\}_{n=0}^{\infty}$ are given in terms of Hermite polynomials H_n and are (weighted) L_2-orthonormal, i.e.,

$$\int_{-\infty}^{\infty} \varphi_m(x)\varphi_n(x)\,\frac{\alpha}{\sqrt{\pi}}\mathrm{e}^{-\alpha^2x^2}\,\mathrm{d}x = \delta_{mn}$$

with weight function $\rho(x) = \frac{\alpha}{\sqrt{\pi}}\mathrm{e}^{-\alpha^2x^2}$. The parameter α that appears in the weight function acts as a *global scale parameter* which essentially localizes the infinite domain $(-\infty,\infty)$ and provides a length scale that allows us to prove the convergence estimates for function approximation with finitely many pieces of data on an infinite domain mentioned in Section 5.4.

As mentioned in Example 2, the Gaussian kernel is of product form, and so we can apply the expansion (11) also in higher dimension with the multi-dimensional eigenvalues and eigenfunctions being products of the one-dimensional ones.

Figures 22 and 23 provide plots of the first few eigenfunctions of the Gaussian kernel in the case $d = 2$. The two different sets illustrate how a different choice of the length scale parameter α localizes the functions. All eigenfunctions in Figure 23 (corresponding to $\alpha = 6$) are much more localized than those in Figure 22 ($\alpha = 1$) even though the Gaussian kernel itself has a shape parameter of $\varepsilon = 3$ in both examples.

In the next section we will discuss how the eigenfunction expansion of the Gaussian kernel can also be used for practical purposes, namely to perform stable calculations with so-called "flat" Gaussians.

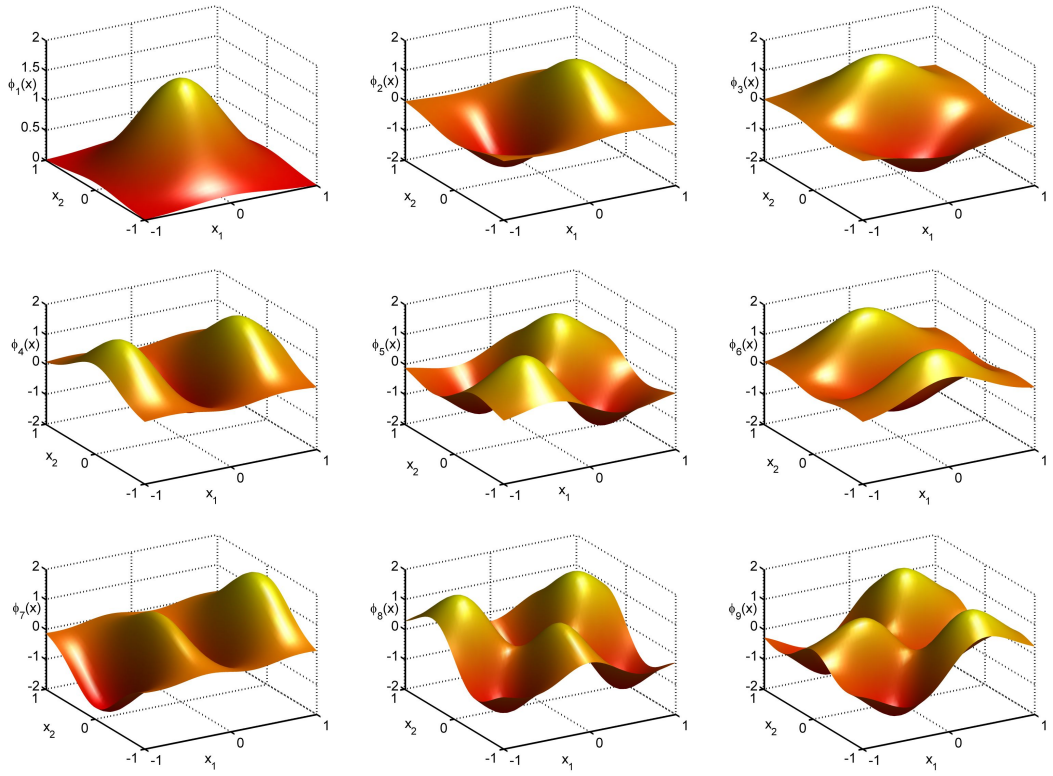


Figure 22: The first few Gaussian eigenfunctions for $\varepsilon = 3$ and $\alpha = 1$.

7 "Flat" RBFs and Stable Evaluation

7.1 Misconception II

Before we discuss how to compute with "flat" Gaussians, let us first consider *why* one might want to do this, and what happens if we do so with the standard approach as outlined in Section 3. In fact, let us begin with the latter. It has been a long-held belief that

Sound bite 4. *"One can't have high accuracy and stability at the same time."*

This belief is

1. based on the observations made by countless practitioners,
2. assumed to rest on a rigorous mathematical foundation, the so-called *uncertainty principle* due to Robert Schaback [127, 128].

Our experiments reported in Section 5.2 support item (1) since we observed there that as the basis functions become wider, or "flatter", the interpolant s_f approximates the test function f more accurately. However, this happened at the cost of an increase in the condition number of the matrix K. We did not show what happens as the shape parameter is pushed further toward zero, but the curves marked with symbols in Figure 24 below illustrate this for several typical interpolation problem.

Item (2), in fact, is a *second misconception*. One should summarize the uncertainty principle of Robert Schaback more carefully as something like

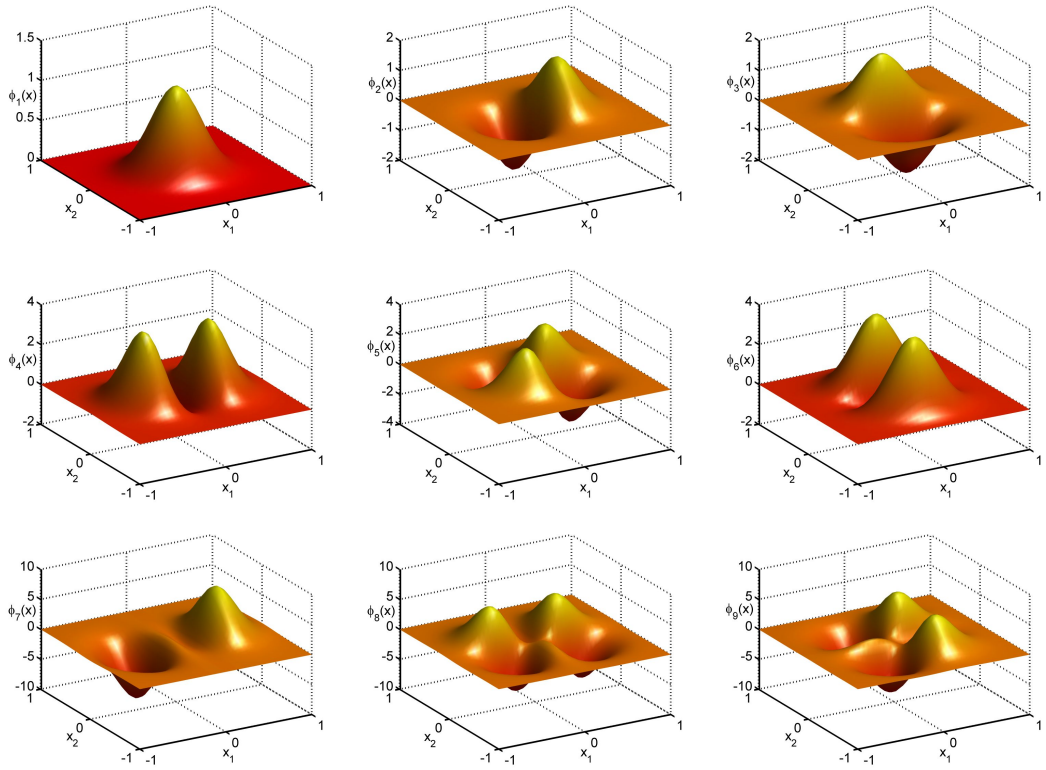


Figure 23: The first few Gaussian eigenfunctions for $\varepsilon = 3$ and $\alpha = 6$.

Sound bite 5 (Correction of Sound bite 4)**.** "Using the standard basis, *one can't have high accuracy and stability at the same time.*"

Therefore, there still might be grounds for believing to be able to stably perform scattered data fitting from the spaces associated with "flat" kernels. We will come back to this question below. So why would anyone want to do this?

7.2 Interpolation with "Flat" Radial Basis Functions

First, we point out that the results mentioned in this section are limited to *radial* kernels, or radial basis functions (RBFs). It was first observed by Bengt Fornberg and his co-workers (see, e.g., [36, 50, 83, 84]) that, if one manages to compute stably in the "flat" limit, then the RBF interpolant actually coincides with a corresponding *polynomial* interpolant. Several people, including Robert Schaback, were involved in rigorously proving this [13, 86, 133]. The previous papers all deal with infinitely smooth radial basis functions. Recently, the case of RBFs with finite smoothness has also been studied in [145], and in [17], a Diplom thesis under the supervision of Robert Schaback.

One of the most intriguing aspects associated with the polynomial limit of RBF interpolants seems to be the fact that RBF interpolants are most accurate (for a fixed number N of given samples) for a *positive* value of the shape parameter ε. Figure 20 shown during our earlier discussion on the optimization of the kernel's shape parameter clearly exhibits a minimum in the interpolation error (as well as for all the error predictors) distinctly away from zero. The reader, however, needs to remember that those plots were obtained using standard kernel interpolants. The fact that we used a regularization algorithm to stabilize the interpolation system does not avoid the ill-conditioning

associated with the "bad" standard basis. This regularization merely smoothes matters out a little. The observation that the interpolation error using infinitely smooth radial kernels – even when computed with a stable basis – is still minimized for a positive value of the shape parameter would imply that radial basis function interpolants are more accurate than polynomial interpolants (see, e.g., Figure 24). However, polynomials are the basis of traditional algorithms (usually referred to a *spectral methods*) for the numerical solution of equations whose solution is known to be smooth. We will come back to this point in Section 8.1 in the context of partial differential equations.

7.3 Stable Computation with "Flat" Radial Basis Functions

A number of approaches have been suggested for overcoming the limitations of the uncertainty principle. Clearly, the problem lies in the choice of basis used to represent the finite-dimensional kernel space $H(K, \mathcal{X}) = \text{span}\{K(\cdot, \boldsymbol{x}_1), \ldots, K(\cdot, \boldsymbol{x}_N)\}$ employed in the solution of the scattered data fitting problem.

Since piecewise polynomial splines of degree m correspond to a certain type of (conditionally) positive definite kernel, we can get some guidance from the spline literature in which it has been known for a long time that B-splines provide a stable basis for the space spanned by the kernels $|\cdot - x_j|_+^m$, where $(\cdot)_+$ is the cut-off function as before. Therefore, a better basis for $H(K, \mathcal{X})$ needs to be found since the standard kernel basis is often unstable. Some work in this direction was reported in [6] for polyharmonic splines, the most direct generalization of piecewise polynomial splines. Other approaches for stable computation have come from Bengt Fornberg and his co-workers such as the Contour-Padé algorithm of [50] and the RBF-QR algorithm of [47, 49]. Alternatively, Robert Schaback and co-workers have recently suggested the use of Newton basis functions whose computation can be coupled with a selection of "good" interpolation points (see, e.g., [33, 108, 114]).

The two papers [47, 49] inspired the work in [43], where a truncation of the Gaussian eigenfunction expansion (11) was used to develop two different algorithms. The first algorithm was inspired by [49] and truncates the (infinite) eigen-decomposition of the kernel interpolation matrix K in such a way that this truncation is accurate to within machine precision. Here one ends up using $M > N$ eigenfunctions to represent the $N \times N$ kernel matrix K. The value of M can be estimated based on the exact knowledge of the eigenvalues of the Gaussian kernel (12) and their fast decay. In Figure 24 the interpolation errors over a range of shape parameter values for this algorithm (QR, solid lines) are compared to those of the standard method (Direct, lines with symbols) and corresponding polynomial interpolants (horizontal lines). The test problems used 10, 20 and 30 samples, respectively, taken at appropriately scaled Chebyshev points from the 1D test function $f(x) = \sinh(x)/(1 + \cosh(x))$.

One can make the following observations:

- The standard (Direct) RBF method suffers from the ill-conditioning associated with sufficiently small values of the shape parameter ε as predicted by the uncertainty principle.
- The RBF interpolant computed stably with the RBF-QR algorithm converges to the polynomial interpolant of degree $N - 1$ at the same set of N Chebyshev points.
- The "optimal" choice of shape parameter is positive, and leads to an interpolation error that is significantly smaller than the one for the corresponding limiting polynomial interpolant.

Since the statistical techniques of Section 5.2 all led to criteria involving the kernel matrix K (or its discrete eigenvalues) there still remains work to be done in adapting these criteria to the stable

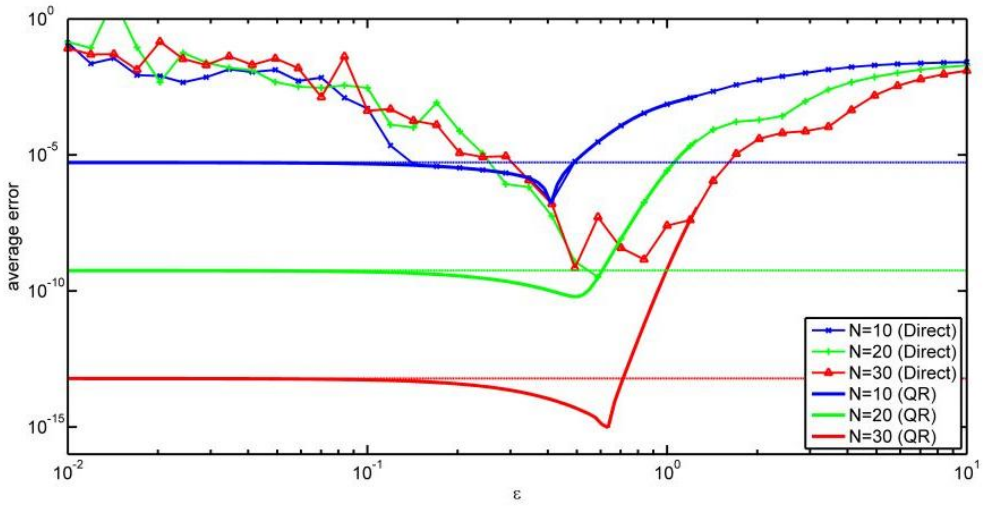


Figure 24: Illustration of spectral accuracy achievable with the RBF-QR interpolation algorithm on Chebyshev points.

evaluation techniques just mentioned, but it should be possible to couple the two approaches to obtain stable and highly accurate Gaussian kernel interpolants.

The second algorithm from [43] uses far fewer eigenfunctions of the Gaussian kernel K than data values. This argument is again justified by the fast decay of the eigenvalues. The value of M is chosen depending on the value of ε and α, the global scale parameter of the weight function ρ that shows up in the formulas for the eigenvalues and eigenfunctions of the Gaussian. This algorithm is similar in spirit to a truncated SVD and was called RBF-QRr (for regression) in [43].

In Figure 25 we compare the approximation errors over a range of shape parameter values for the RBF-QRr algorithm with those of the standard method (Direct, lines with symbols). The test problems used here were generated by sampling the fifth degree polynomial $f(u, v, w, x, y) = 1 + (u + v + w)^2(x - y)^2(u + x)$ in five variables, i.e., $d = 5$, at various sets of Halton points. The dimension of the space of degree 5 polynomials in five variables is 252, so it is not surprising that $N = 200$ points are not enough to recover the polynomial from its samples. However, using more samples, we can see that the RBF regression fit is able to recover the polynomial to within machine precision provided the shape parameter ε is sufficiently small.

8 Other Applications

8.1 Numerical Solution of Partial Differential Equations

One of the greatest application area of so-called *meshfree* methods is to the numerical solution of partial differential equations (PDEs). A seemingly endless alphabet soup of acronyms for different methods exists in the literature, such as EFG, hp-clouds, MLPG, MLS, PUFEM, RKPM, SPH, and XFEM (for more see, e.g., [102]). Some of these methods are closely related to positive definite kernels (such as MLPG, RKPM and SPH), others not so much (such as EFG, MLS and XFEM). For a relatively recent survey of many of these methods see [110].

We will briefly discuss some developments in the numerical solution of elliptic PDEs via a collocation or strong form kernel-based approach. There are essentially two different alternatives

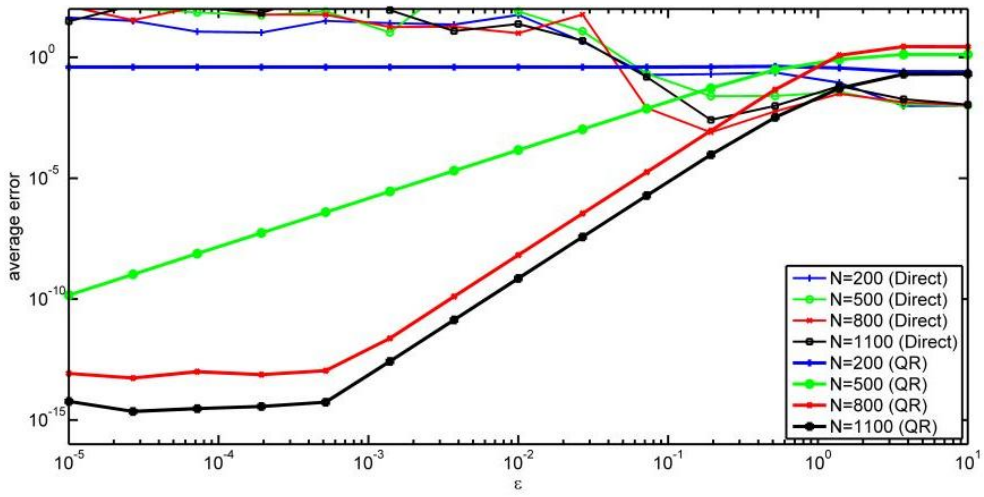


Figure 25: High-accuracy approximation of a 5D polynomial using the RBF-QRr approximation algorithm on Halton points.

proposed in the literature: (1) *non-symmetric collocation*, and (2) *symmetric collocation*. The non-symmetric approach was first proposed by Ed Kansa [77, 78] and has been extremely popular with many researchers, while the symmetric approach has been attributed to [37, 156]. Both methods are discussed in detail in, e.g., [38].

Taking a simple Poisson equation

$$\begin{aligned} \nabla^2 u(\boldsymbol{x}) &= f(\boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega, \\ u(\boldsymbol{x}) &= g(\boldsymbol{x}), \qquad \boldsymbol{x} \text{ on } \partial\Omega, \end{aligned}$$

the non-symmetric method assumes that the approximate solution is expanded in the form

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

where the $\boldsymbol{x}_j$, $j = 1, \ldots, N$, are so-called *collocation points* distributed both in the interior of Ω and along its boundary. For the symmetric method one assumes

$$\hat{u}(\boldsymbol{x}) = \sum_{j=1}^{N_I} c_j \nabla^2_{\boldsymbol{\xi}} K(\boldsymbol{x}, \boldsymbol{\xi})|_{\boldsymbol{\xi}=\boldsymbol{x}_j} + \sum_{j=N_I+1}^{N} c_j K(\boldsymbol{x}, \boldsymbol{x}_j),$$

where N_I denotes the number of collocation points in the interior of Ω. As a result, when one forms a linear system for the determination of the unknown coefficients c_j by "interpolating the PDE and its boundary condition" at the collocation points, then in the non-symmetric case one obtains a non-symmetric system matrix consisting of two blocks of the form

$$\begin{bmatrix} \left(\nabla^2 K(\boldsymbol{x}, \boldsymbol{x}_j)|_{\boldsymbol{x}=\boldsymbol{x}_i}\right)_{i,j=1}^{N_I,N} \\ \left(K(\boldsymbol{x}_i, \boldsymbol{x}_j)\right)_{i=N_I+1,j=1}^{N,N} \end{bmatrix}, \tag{14}$$

while for the symmetric method one obtains a symmetric matrix consisting of four blocks of the form

$$\begin{bmatrix} \left(\nabla^2_{\boldsymbol{x}}\nabla^2_{\boldsymbol{\xi}}K(\boldsymbol{x},\boldsymbol{\xi})|_{\boldsymbol{x}=\boldsymbol{x}_i,\boldsymbol{\xi}=\boldsymbol{x}_j}\right)_{i,j=1}^{N_I,N_I} & \left(\nabla^2_{\boldsymbol{x}}K(\boldsymbol{x},\boldsymbol{x}_j)|_{\boldsymbol{x}=\boldsymbol{x}_i}\right)_{i=1,j=N_I+1}^{N_I,N} \\ \left(\nabla^2_{\boldsymbol{\xi}}K(\boldsymbol{x}_i,\boldsymbol{\xi})|_{\boldsymbol{\xi}=\boldsymbol{x}_j}\right)_{i=N_I+1,j=1}^{N,N_I} & \left(K(\boldsymbol{x}_i,\boldsymbol{x}_j)\right)_{i,j=N_I+1}^{N} \end{bmatrix}. \tag{15}$$

Robert Schaback has been instrumental in providing a theoretical foundation for the well-posedness and error analysis of both the non-symmetric and symmetric collocation methods. Moreover, he has proposed several efficient algorithms for the practical implementation of these methods. Together with Carsten Franke, his Ph.D. student at the time, he established a convergence theory for the symmetric collocation approach [53, 54]. A few years later, together with Benny Hon, he provided a set of carefully selected and executed numerical examples which demonstrated that – contrary to the case of scattered data interpolation – there exist sets $\mathcal{X}$ of collocation points which lead to a singular collocation matrix (14) (for the symmetric case Zongmin Wu established nonsingularity of the collocation matrix (15) in [156]).

We have already seen that some positive definite kernels such as the Gaussian provide spectral rates of approximation. Polynomials also have this property, and this has led to the development of so-called *pseudospectral methods* (see, e.g., [46, 149]) for the numerical solution of PDEs. Since one can introduce differentiation matrices into the non-symmetric collocation approach in the same way as for polynomial methods, kernel-based collocation can also be formulated as a pseudospectral collocation method (see [38]). This is one of the main reasons why the non-symmetric collocation approach is very attractive for use in applications. However, the counterexamples of [69] presented somewhat of a discouragement for people to use this approach. On the other hand, the discussion in Section 7 on "flat limits" and the insight gained in the corresponding literature [13, 36, 50, 83, 84, 86, 133] show that there is not only a *formal* connection between RBF collocation and polynomial spectral methods, but that one should in fact interpret RBF collocation as a *generalization of polynomial spectral methods* with more flexibility to deal with problems in complex domains (both with complicated shapes and in high space dimensions) and with the potential for higher accuracy than their polynomial counterparts.

This great potential for practical applications has spurred a series of papers by Robert Schaback and co-workers [70, 85, 89, 132, 134] in which the authors have succeeded in providing a firm theoretical framework for the well-posedness and high accuracy of the non-symmetric collocation approach. As a by-product, a framework for the solution of more general operator equations – both in strong form as well as in weak form – have recently emerged [134]. Among other things, this provides a theoretical foundation for the "meshless local Petrov-Galerkin" (MLPG) method for which we listed two books in the introduction.

So far, the discussion has focussed on elliptic PDEs. Of course, many applications are dynamic in nature and will therefore either be of the parabolic or hyperbolic type. While the seminal paper by Kansa [78] already used a kernel-based approach to solve time-dependent PDEs of both types, there are only very few papers that discuss convergence and stability of the solution of time-dependent PDEs with a kernel-based approach. One such paper is the recent preprint [71] by Benny Hon and Robert Schaback in which they gave the equivalent of a *CFL condition* [26] (see also Figure 26) for the solution of the standard heat equation with a method of lines approach that uses positive definite kernels for the spatial discretization and an Euler method in time. The kernel-based CFL condition mirrors that for standard finite-difference methods and requires that the time step Δt satisfies $\Delta t \leq C(\Delta x)^2$, where C is some positive constant and Δx denotes the spacing of equidistant spatial collocation points.

The solution of *ill-posed problems* by kernel-based methods has received relatively little attention


Figure 26: Left to right: Kurt Friedrichs, Richard Courant and Hans Lewy.

(see, e.g., [25]). However, in [132] Robert Schaback provides a theoretical foundation for the numerical solution of a certain class of such problems. The solution of *nonlinear PDEs* with kernel-based methods is discussed in the presentation of Klaus Böhmer at the Göttingen workshop [12].

8.2 A Very Brief Discussion of Other Applications

As already alluded to in the introduction, positive definite kernels play a role in many other application areas. The books listed in the introduction provide a much richer source of information on these topics.

We referred to the scattered data fitting problem as the fundamental application. As already mentioned there, applications arise in many different areas of numerical analysis, statistics and engineering. For example, Michael Scheuerer's presentation at the Göttingen workshop [138] focusses on the *stochastic perspective* for data fitting. In the literature on *computer experiments* (see [63] for recent work using a multilevel kernel-based interpolation algorithm with the compactly supported Wendland functions of Section 2.4) and other engineering experiments one increasingly encounters models based on positive definite kernels, RBFs or kriging. One such topic is *response surface modeling* (see, e.g., [100]). Other types of applications that boil down to data fitting are *rapid prototyping* (see, e.g., [22]) and *computer graphics* (see, e.g., [87]). Here one often uses *implicit surface models* to approximate or interpolate point cloud data.

Machine learning or *statistical learning* is another area that very heavily relies on positive definite kernels. Algorithms such as *support vector machines* are used in many different applications such as *medical diagnostics* and *text analysis*. The presentation of Ding-Xuan Zhou at the Göttingen workshop [158] has its roots in statistical learning applications.

In the previous subsection we briefly talked about PDEs. Much work has been focussed on the use of positive definite kernels (or conditionally positive definite kernels, such as the *multiquadric*) for the numerical solution of PDEs.

Applications of positive definite kernels in various other branches of mathematics are in *multivariate integration*, *multivariate optimization* and in numerical analysis and scientific computing, where one studies fast, accurate and adaptive algorithms ideally implemented in *high-performance computing* environments. Unfortunately, the latter area has up to now received relatively little attention. Among the few contributions in this direction are [6, 61].

9 The Future

So where are the developments in this field likely to go in the near future? For one thing, hopefully there will be more interaction between different fields, such as approximation theory, statistics,

learning theory and engineering. Some conferences in recent years have been organized with this as one of their goals – and Robert Schaback has been present at most of them.

The recent paper [139] by Michael Scheuerer, Robert Schaback and Martin Schlather provides an excellent overview for someone who wants to see some of the connections between approximation (the RBF approach) and statistics (the kriging method). Many more such collaborations are needed. For example, numerical analysts can benefit from the knowledge of statisticians for the estimation of parameters and calibration of models. The choice of a "correct" kernel and its "correct" scale is still a major unknown in the numerical analysis world. On the other hand, statisticians may be able to benefit from the numerical analysis community when it comes to issues such as efficient algorithms and preconditioning.

High-dimensional problems are ubiquitous in statistical learning and in complexity theory, but the numerical analysis community has not focussed on such problems. Most convergence results and also many algorithms focus on the cases of $d = 1, 2, 3$ dimensions. However, as pointed out earlier, there are many important applications in the fields of computer experiments, response surface modeling and in finance which naturally "live" in high-dimensional spaces. Working successfully with such applications will most certainly involve *nonstandard kernels* such as anisotropic kernels, or kernels on manifolds.

In Section 8.1 we briefly discussed some of the accomplishments toward the numerical solution of PDEs. However, there remains much more to be done. The solution of time-dependent PDEs with kernel-based methods is still in its infancy. In particular, the real flexibility of meshfree methods should come to fruition when they are employed in an *adaptive algorithm*, where both the location of the discretization points and the scale of the kernels changes adaptively with time. Moreover, kernels with spatially varying scales are likely to be needed in the moving point framework. Kernels with spatially varying scales for static problems have been considered, e.g., by Robert Schaback together with Mira Bozzini and Licia Lenarduzzi [15] and in [51]. Another area that has to date received almost no attention is the solution of *stochastic* PDEs. This area is likely to benefit from interactions between numerical analysis and statistics.

Finally, kernel-based approximation methods have not entered the *scientific computing* field in a decisive manner. There is still much to be accomplished with respect to fast and stable algorithms that can be efficiently implemented in high-performance computing environments.

10 Robert Schaback

I would like to close this article with a few words about Robert Schaback. As should have become apparent during the course of this article, Robert Schaback is a *researcher* who has made many fundamental contributions at the interface of approximation theory, numerical analysis and computer science. Moreover, he has been a *bridge-builder* to various other disciplines. To emphasize this point, here is an excerpt from a poem entitled "The Bridge Builder" by Will Allen Dromgoole (1900):

> *"There followeth after me today,*
> *A youth, whose feet must pass this way.*
> *This chasm, that has been naught to me,*
> *To that fair-haired youth may a pitfall be.*
> *He, too, must cross in the twilight dim;*
> *Good friend, I am building this bridge for him."*

This poem also speaks to the many efforts Robert Schaback has spent as an *educator* who was active not only in mathematics, but also in teacher education and in the development of computer science

as an independent discipline at the University of Göttingen. His work with his many PhD students demonstrates how he has been building bridges enabling "fair-haired youths" to succeed in their future. For those of us who have not had the privilege of being students of Robert Schaback he has been a *mentor* and providing inspiration and motivation in many discussions and communications with encouragement such as

"Das sollte man mal durchrechnen."

Finally, Robert Schaback loves the outdoors and is a *hiking enthusiast*, so that a special issue in this journal is the fitting place to celebrate his 65^{th} birthday. Happy Birthday, Robert, and many more healthy, productive and joyful years!

Figure 27: Robert Schaback (back row, 4^{th} from the right) with a group of hiking mathematicians atop Piz Boe during the 1^{st} Dolomites Workshop on Constructive Approximation and Applications in 2006.

Acknowledgment. The author benefitted from discussions with many colleagues, students and visitors of the meshfree methods research seminar at IIT including Igor Cialenco, Fred Hickernell, Lulu Kang, Gene Lamie, Mike Machen, Mike McCourt, Paritosh Mokhasi, Mike Mongillo, John Riddle, Guohui Song, Josh Tate, Marjan Uddin, Henryk Woźniakowski and Qi Ye. All images included in this article were either created by the author or are, to the author's best knowledge, in the public domain and were primarily obtained from either the MacTutor History of Mathematics archive `http://www-history.mcs.st-andrews.ac.uk/`, the Oberwolfach Photo Collection `http://owpdb.mfo.de/` or Wikimedia Commons `http://commons.wikimedia.org/`.

References

[1] Allen, D. M., The relationship between variable selection and data agumentation and a method for prediction. *Technometrics* **16**/1 (1974), 125–127. http://dx.doi.org/10.2307/1267500.

[2] Aronszajn, N., Theory of reproducing kernels. *Transactions of the American Mathematical Society* **68** (1950), 337–404. http://dx.doi.org/10.2307/1990404.

[3] Atluri, S. N., *The Meshless Method (MLPG) for Domain & BIE Discretizations*. Tech Science Press (Encino, CA), 2004.

[4] Atluri, S. N., and S. Shen, *The Meshless Local Petrov-Galerkin (MLPG) Method*. Tech Science Press (Encino, CA), 2002.

[5] Baxter, B. J. C., and N. Sivakumar, On shifted cardinal interpolation by Gaussians and multiquadrics. *J. Approx. Theory* **87**/1 (1996), 36–59. http://dx.doi.org/10.1006/jath.1996.0091.

[6] Beatson, R. K., W. A. Light, and S. Billings, Fast solution of the radial basis function interpolation equations: domain decomposition methods. *SIAM J. Sci. Comput.* **22**/5 (2001), 1717–1740. http://dx.doi.org/10.1137/S1064827599361771.

[7] Belytschko, T., and J. S. Chen, *Meshfree and Particle Methods*. Wiley, 2007.

[8] Berg, C., J. P. R. Christensen, and P. Ressel, *Harmonic Analysis on Semigroups*. Springer-Verlag, Berlin-New York, 1984.

[9] Bergman, S., *The Kernel Function and Conformal Mapping*. Mathematical Surveys and Monographs, AMS, 1950.

[10] Berlinet, A., and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, 2003.

[11] Bochner, S., *Vorlesungen über Fouriersche Integrale*. Mathematik und ihre Anwendungen, Vol. 12, Akademische Verlagsgesellschaft, Leipzig, 1932. English translation: Lectures on Fourier Integrals, Princeton University Press, 1959.

[12] Böhmer, K., A nonlinear discretization theory with applications to meshfree methods: quasi-linear and fully nonlinear PDEs. Presented at *Workshop on Kernel Functions and Meshless Methods honoring the 65th birthday of Robert Schaback*, Göttingen, 2011.

[13] de Boor, C., On interpolation by radial polynomials. *Adv. Comput. Math.* **24**/1-4 (2006), 143–153. http://dx.doi.org/10.1007/s10444-004-7630-3.

[14] Boyd, J. P., and L. Wang, An analytic approximation to the cardinal functions of Gaussian radial basis functions on an infinite lattice. *Appl. Math. Comput.* **215**/6 (2009), 2215–2223. http://dx.doi.org/10.1016/j.amc.2009.08.037.

[15] Bozzini, M., L. Lenarduzzi, and R. Schaback, Adaptive interpolation by scaled multiquadrics. *Adv. Comput. Math.* **16** (2002), 375–387. http://dx.doi.org/10.1023/A:1014584220418.

[16] Braess, D., *Nonlinear Approximation Theory*. Springer, Berlin, 1986.

[17] Brink-Spalink, R., Flat Limits von radialen Basisfunktionen. Diplom thesis, Universität Göttingen, 2010.

[18] Buhmann, M. D., Multivariate cardinal interpolation with radial basis functions. *Constr. Approx.* **6** (1990), 225–255. http://dx.doi.org/10.1007/BF01890410.

[19] Buhmann, M. D., Radial functions on compact support. *Proc. Edin. Math. Soc. II* **41** (1998), 33-46. http://dx.doi.org/10.1017/S0013091500019416

[20] Buhmann, M. D., Radial basis functions. *Acta Numerica* **9** (2000), 1–38. http://journals.cambridge.org/abstract_S0962492900000015

[21] Buhmann, M. D., *Radial Basis Functions: Theory and Implementations.* Cambridge University Press, 2003.

[22] Carr, J. C, R. K Beatson, J. B Cherrie, T. J Mitchell, W. R Fright, B. C McCallum, and T. R Evans, Reconstruction and representation of 3D objects with radial basis functions. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '01*, ACM (New York, NY), 2001, 67–76. http://dx.doi.org/10.1145/383259.383266.

[23] Catoni, O., *Statistical Learning Theory and Stochastic Optimization.* Lecture Notes in Mathematics 1851, Springer Verlag, 2004.

[24] Cheney, E. W., and W. A. Light, *A Course in Approximation Theory.* Brooks/Cole (Pacific Grove, CA), 1999.

[25] Cheng, A. H.-D., and J. J. S. P. Cabral, Direct solution of certain ill-posed boundary value problems by collocation method. In *Boundary Elements XXVII*, A. Kassab, C. A. Brebbia, E. Divo, and D. Poljak (eds.), WIT Press (Southampton) 2005, 35–44. http://library.witpress.com/pages/PaperInfo.asp?PaperID=14449.

[26] Courant, R., K. Friedrichs, and H. Lewy, Über die partiellen Differenzengleichungen der mathematischen Physik. *Math. Ann.* **100**/1 (1928), 32–74. http://dx.doi.org/10.1007/BF01448839.

[27] Craven, P., and G. Wahba, Smoothing noisy data with spline functions. *Numer. Math.* **31** (1979), 377–403. http://dx.doi.org/10.1007/BF01404567.

[28] Cressie, N. A. C., *Statistics for Spatial Data*, revised edition. Wiley-Interscience, 1993.

[29] Cristianini, N., and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* Cambridge University Press, Cambridge, 2000.

[30] Cucker, F., and D. X. Zhou, *Learning theory: An approximation theory viewpoint.* Cambridge University Press, Cambridge, 2007.

[31] Curtis, P. C., Jr., n-parameter families and best approximation. *Pacific J. Math.* **9** (1959), 1013–1027. http://projecteuclid.org/euclid.pjm/1103038880.

[32] Bos, L., S. De Marchi, A. Sommariva and M. Vianello, On multivariate Newton interpolation at discrete Leja points. Dolomites Research Notes on Approximation (DRNA), this issue.

[33] De Marchi, S., and R. Schaback, Nonstandard kernels and their applications. *Dolomite Research Notes on Approximation* **2** (2009), 16–43. http://drna.di.univr.it/papers/2009/SchabackDeMarchi/Paper2_2009.pdf.

[34] De Marchi, S., and R. Schaback, Stability of kernel-based interpolation. *Adv. in Comput. Math.* **32**/2 (2010), 155-161. http://dx.doi.org/10.1007/s10444-008-9093-4.

[35] De Marchi, S., R. Schaback, and H. Wendland, Near-optimal data-independent point locations for radial basis function interpolation. *Adv. in Comput. Math.* **23**/3 (2005), 317–330. http://dx.doi.org/10.1007/s10444-004-1829-1.

[36] Driscoll, T. A., and B. Fornberg, Interpolation in the limit of increasingly flat radial basis functions. *Comput. Math. Appl.* **43**/3-5 (2002), 413–422. http://dx.doi.org/10.1016/S0898-1221(01)00295-4.

[37] Fasshauer, G. E., Solving partial differential equations by collocation with radial basis functions. In *Surface Fitting and Multiresolution Methods*, A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press (Nashville, TN) 1997, 131–138.

[38] Fasshauer, G. E., *Meshfree Approximation Methods with* MATLAB. World Scientific Publishers, 2007.

[39] Fasshauer, G. E., Tutorial on Meshfree Approximation Methods with MATLAB. *Dolomite Research Notes on Approximation* **1**, 2008. http://drna.di.univr.it/papers/2008/Fasshauer/Lecture5.pdf.

[40] Fasshauer, G. E., Mathematica CDFs for Positive Definite Kernels. http://math.iit.edu/~fass/PDKernelsMathematica.html.

[41] Fasshauer, G. E., F. J. Hickernell, E. Lamie, P. Mokhasi, and J. Tate, "Optimal" scaling and stable computation of meshfree kernel methods. presented at *3rd Workshop on High-Dimensional Approximation*, University of New South Wales, Sydney, 2009. http://math.iit.edu/~fass/FasshauerSydney.pdf.

[42] Fasshauer, G. E., F. J. Hickernell, and H. Woźniakowski, On dimension-independent rates of convergence for function approximation with Gaussian kernels. *SIAM J. Numer. Anal.*, to appear. http://math.iit.edu/~fass/gfh2011oct26.pdf.

[43] Fasshauer, G. E., and M. J. McCourt, Stable evaluation of Gaussian RBF interpolants. Illinois Institute of Technology, submitted, 2011. http://math.iit.edu/~fass/StableGaussianRBFs.pdf.

[44] Fasshauer, G. E., and J. G. Zhang, On choosing "optimal" shape parameters for RBF approximation. *Numer. Algorithms* **45**/1-4 (2007), 345–368. htpp://dx.doi.org/10.1007/s11075-007-9072-8.

[45] Fasshauer, G. E., and J. G. Zhang, Iterated approximate moving least squares approximation. In *Advances in Meshfree Techniques*, V. M. A. Leitão, C. Alves, and C. A. Duarte (eds.), Springer 2007, 221–240.

[46] Fornberg, B., *A Practical Guide to Pseudospectral Methods*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 1998.

[47] Fornberg, B., E. Larsson, and N. Flyer, Stable computations with Gaussian radial basis functions. *SIAM J. Sci. Comput.* **33** (2011), 869–892. http://dx.doi.org/10.1137/09076756X.

[48] Fornberg, B., E. Larsson, and G. Wright, A new class of oscillatory radial basis functions. *Comput. Math. Appl.* **51** (2006), 1209–1222. http://dx.doi.org/10.1016/j.camwa.2006.04.004.

[49] Fornberg, B., and C. Piret, A stable algorithm for flat radial basis functions on a sphere. *SIAM J. Sci. Comp.* **30** (2007), 60–80. http://dx.doi.org/10.1137/060671991.

[50] Fornberg, B., and G. Wright, Stable computation of multiquadric interpolants for all values of the shape parameter. *Comput. Math. Appl.* **47** (2004), 497–523. http://dx.doi.org/10.1016/j.camwa.2003.08.010.

[51] Fornberg, B., and J. Zuev, The Runge phenomenon and spatially variable shape parameters in RBF interpolation. *Comput. Math. Appl.* **54**/3 (2007), 379–398. http://dx.doi.org/10.1016/j.camwa.2007.01.028.

[52] Forrester, A. I. J., A. Sobester, and A. J. Keane, *Engineering Design via Surrogate Modelling – a Practical Guide*. John Wiley & Sons, 2008.

[53] Franke, C., and R. Schaback, Solving partial differential equations by collocation using radial basis functions. *Appl. Math. Comp.* **93** (1998), 73–82. http://dx.doi.org/10.1016/S0096-3003(97)10104-7.

[54] Franke, C., and R. Schaback, Convergence orders of meshless collocation methods using radial basis functions. *Adv. in Comput. Math.* **8** (1998), 381–399. http://dx.doi.org/10.1023/A:1018916902176.

[55] Freeden, W., T. Gervens, and M. Schreiner, *Constructive Approximation on the Sphere*. Oxford University Press (Oxford), 1998.

[56] Fuselier, E., and G. B. Wright, Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates. Preprint, arXiv:1007.2825v1. http://128.84.158.119/abs/1007.2825v1.

[57] Gauss, C. F., *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Hamburg: Friedrich Perthes and I.H. Besser, 1809. http://resolver.sub.uni-goettingen.de/purl?PPN236008730.

[58] Glasserman, P., *Monte Carlo Methods in Financial Engineering*. Applications of Mathematics, Vol. 53, Springer (New York), 2004.

[59] Golomb, M., and H. F. Weinberger, Optimal approximation and error bounds. In *On Numerical Approximation*, R. E. Langer (ed.), University of Wisconsin Press, 1959, 117–190.

[60] Golub, G. H., M. Heath, and G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21**/2 (1979), 215–223. http://dx.doi.org/10.2307/1268518.

[61] Gumerov, N. A., and R. Duraiswami, Fast radial basis function interpolation via preconditioned Krylov iteration. *SIAM J. Scient. Computing* **29**/5 (2007), 1876–1899. http://dx.doi.org/10.1137/060662083.

[62] Gutzmer, T., Interpolation by positive definite functions on locally compact groups with application to SO(3). *Results Math.* **29** (1996), 69–77.

[63] Haaland, B., and P. Z. G. Qian, Accurate emulators for large-scale computer experiments. *Ann. Stat.*, submitted (2010). http://www.stat.wisc.edu/~zhiguang/MultiStep_Revision.pdf.

[64] Hangelbroek, T., W. Madych, F. J Narcowich, and J. D Ward, Cardinal interpolation with Gaussian kernels. Preprint, arXiv:1008.3168v1. http://arxiv.org/abs/1008.3168.

[65] Hastie, T., R. Tibshirani, and J. Friedman, *Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd edition). Springer Verlag, 2009.

[66] Hickernell, F. J., Shape parameter problem. Technical report, Illinois Institute of Technology, 2009.

[67] Hickernell, F. J., and Y. C. Hon, Radial basis function approximations as smoothing splines. *Appl. Math. Comput.* **102**/1 (1999), 1–24. http://dx.doi.org/10.1016/S0096-3003(98)10012-7.

[68] Hilbert, D., Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen I. *Göttinger Nachrichten, Math.-Phys. Kl.* (1904), 49–91. http://resolver.sub.uni-goettingen.de/purl?GDZPPN002499967.

[69] Hon, Y. C., and R. Schaback, On unsymmetric collocation by radial basis functions. *Appl. Math. Comput.* **119** (2001), 177–186. http://dx.doi.org/10.1016/S0096-3003(99)00255-6.

[70] Hon, Y. C., and R. Schaback, Solvability of partial differential equations by meshless kernel methods. *Adv. in Comput. Math.* **28**/3 (2008), 283–299. http://dx.doi.org/10.1007/s10444-006-9023-2.

[71] Hon, Y. C., and R. Schaback, Meshless kernel techniques for the heat equation. Preprint, Universität Göttingen. http://num.math.uni-goettingen.de/schaback/research/papers/MKTftHE.pdf.

[72] Horn, R. A., and C. R. Johnson, *Matrix Analysis*. Cambridge University Press (Cambridge), 1985.

[73] Hunter, J. K., and B. Nachtergaele, *Applied Analysis*. World Scientific Publishing Company, 2001.

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

[75] Iske, A., A hitchhiker's guide through the zoo of radial kernels. Presented at *Workshop on Kernel Functions and Meshless Methods honoring the 65th birthday of Robert Schaback*, Göttingen, 2011.

[76] Joachims, T., *Learning to Classify Text Using Support Vector Machines: Methods, Theory, and Algorithms*. Kluwer Academic Publishers, 2002.

[77] Kansa, E. J., Application of Hardy's multiquadric interpolation to hydrodynamics. In *Proc. 1986 Simul. Conf.* **4** (1985), 111–117. http://www.osti.gov/bridge/servlets/purl/6373842-OnShrl/.

[78] Kansa, E. J., Multiquadrics — A scattered data approximation scheme with applications to computational fluid-dynamics — II: Solutions to parabolic, hyperbolic and elliptic partial differential equations. *Comput. Math. Appl.* **19** (1990), 147–161. http://dx.doi.org/10.1016/0898-1221(90)90271-K.

[79] Khinchin, A., Korrelationstheorie der stationären stochastischen Prozesse. *Math. Ann.* **109** (1934), 604–615. http://dx.doi.org/10.1007/BF01449156.

[80] Kitanidis, P. K., *Introduction to Geostatistics: Applications in Hydrogeology.* Cambridge University Press, 1997.

[81] Laplace, P. S. de, Suite de Mémoire sur les approximations des formules qui sont fonctions de très grands nombres. Histoire de lacademie royale des sciences de Paris, Année 1783 (1786), 423–467. Reprinted in [82, Vol. 10, pp. 295–338] http://gallica.bnf.fr/ark:/12148/bpt6k775981/f304.tableDesMatieres.

[82] Laplace, P. S. de, Œuvres complètes de Laplace, 14 Volumes, Gauthier-Villars, Paris, 1878–1912. http://gallica.bnf.fr/ark:/12148/bpt6k775981/.

[83] Larsson, E., and B. Fornberg, A numerical study of some radial basis function based solution methods for elliptic PDEs. *Comput. Math. Appl.* **46**/5-6 (2003), 891–902. http://dx.doi.org/10.1016/S0898-1221(03)90151-9.

[84] Larsson, E., and B. Fornberg, Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions. *Comput. Math. Appl.* **49** (2005), 103–130. http://dx.doi.org/10.1016/j.camwa.2005.01.010.

[85] Lee, C.-F., L. Ling, and R. Schaback, On convergent numerical algorithms for unsymmetric collocation. *Adv. in Comput. Math.* **30**/4 (2009), 339–354. http://dx.doi.org/10.1007/s10444-008-9071-x.

[86] Lee, Y. J., G. J. Yoon, and J. Yoon, Convergence of increasingly flat radial basis interpolants to polynomial interpolants. *SIAM J. Math. Anal.* **39** (2007), 537–553. http://dx.doi.org/10.1137/050642113.

[87] Lewis, J. P, F. Pighin, and K. Anjyo, Scattered data interpolation and approximation for computer graphics. In *ACM SIGGRAPH ASIA 2010 Courses*, ACM (New York, NY), 2010, 2:1–2:73. http://dx.doi.org/10.1145/1900520.1900522.

[88] Li, S., and W. K. Liu, *Meshfree Particle Methods.* Springer, 2007.

[89] Ling, L., and R. Schaback, Stable and convergent unsymmetric meshless collocation methods. SIAM J. Numer. Anal. **46**/3 (2008), 1097–1115. http://dx.doi.org/10.1137/06067300X.

[90] Liu, G. R., *Mesh Free Methods: Moving beyond the Finite Element Method.* CRC Press (Boca Raton, FL), 2002.

[91] Liu, G. R., and M. B. Liu, *Smoothed Particle Hydrodynamics: A Meshfree Particle Method.* World Scientific Publishers, 2003.

[92] Lowitzsch, S., Matrix-valued radial basis functions: stability estimates and applications. *Adv. in Comput. Math.* **23**/3 (2005), 299–315. http://dx.doi.org/10.1007/s10444-004-1786-8.

[93] Madych, W. R., and S. A. Nelson, Bounds on multivariate polynomials and exponential error estimates for multiquadric interpolation. *J. Approx. Theory* **70** (1992), 94–114. http://dx.doi.org/10.1016/0021-9045(92)90058-V.

[94] Mairhuber, J. C., On Haar's theorem concerning Chebychev approximation problems having unique solutions. *Proc. Amer. Math. Soc.* **7** (1956), 609–615. http://dx.doi.org/10.2307/2033359.

[95] Matérn, B., *Spatial Variation* (Second ed.). Lecture Notes in Statistics 36, Springer-Verlag (Berlin), 1986.

[96] Mathematics Genealogy Project. http://www.genealogy.math.ndsu.nodak.edu/.

[97] Matheron, G., *Les Variables Régionalisées et Leur Estimation*. Masson (Paris), 1965.

[98] Mathias, M., Über positive Fourier-Integrale. *Math. Zeit.* **16** (1923), 103–125. http://gdz.sub.uni-goettingen.de/dms/load/img/?PPN=GDZPPN002446022.

[99] Mazya, V., and G. Schmidt, *Approximate Approximations*, Mathematical Surveys and Monographs, Vol. 141. American Mathematical Society (Providence), 2007.

[100] McDonald, D. B., W. J. Grantham, W. L. Tabor, and M. J. Murphy, Global and local optimization using radial basis function response surface models. *Appl. Math. Model.* **31**/10 (2007), 2095–2110. http://dx.doi.org/10.1016/j.apm.2006.08.008.

[101] Mercer, J., Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London, Series A* **209** (1909), 415–446. http://dx.doi.org/10.1098/rsta.1909.0016.

[102] Meshfree Methods: List of methods and acronyms, Wikipedia. http://en.wikipedia.org/wiki/Meshfree_methods#List_of_methods_and_acronyms.

[103] Meschkowski, H., *Hilbertsche Räume mit Kernfunktion*. Springer (Berlin), 1962.

[104] Micchelli, C. A., Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constr. Approx.* **2** (1986), 11–22. http://dx.doi.org/10.1007/BF01893414.

[105] Mongillo, M., Choosing basis functions and shape parameters for radial basis function methods. SIAM Undergraduate Research Online, accepted 2011.

[106] Moore, E. H., *General Analysis, Part I*. Memoirs, Vol. 1, Amer. Philos. Soc. (Philadelphia), 1935.

[107] Müller, S., Komplexität und Stabilität von kernbasierten Rekonstruktionsmethoden. Ph.D. Dissertation, Universität Göttingen, 2009. http://webdoc.sub.gwdg.de/diss/2009/mueller/.

[108] Müller, S., and R. Schaback, A Newton basis for kernel spaces. *J. Approx. Th.* **161** (2009), 645–655. http://dx.doi.org/10.1016/j.jat.2008.10.014.

[109] Narcowich, F. J., and J. D. Ward, Generalized Hermite interpolation via matrix-valued conditionally positive definite functions. *Math. Comp.* **63** (1994), 661–687. http://dx.doi.org/10.2307/2153288.

[110] Nguyen, V. P., T. Rabczuk, S. Bordas, and M. Duflot, Meshless methods: A review and computer implementation aspects. *Math. Comput. Simulation* **79**/3 (2008), 763–813. http://dx.doi.org/10.1016/j.matcom.2008.01.003.

[111] Niu, B., Monte-Carlo Simulation of Infinite-Dimensional Integrals. Ph.D. Thesis, Illinois Institute of Technology, 2011.

[112] Opfer, R., Multiscale Kernels. Ph.D. Dissertation, Universität Göttingen, 2004.

[113] Opfer, R., Multiscale kernels. *Adv. in Comput. Math.* **25**/4 (2006), 357–380. http://dx.doi.org/10.1007/s10444-004-7622-3.

[114] Pazouki, M., and R. Schaback, Bases for kernel-based spaces. *J. Comput. Appl. Math.*, in press. http://dx.doi.org/10.1016/j.cam.2011.05.021.

[115] Platte, R. B., and T. A. Driscoll, Polynomials and potential theory for Gaussian radial basis function interpolation. *SIAM J. Numer. Anal.* **43**/2 (2005), 750–766. http://dx.doi.org/10.1137/040610143.

[116] The R Project for Statistical Computing. http://www.r-project.org/.

[117] Rasmussen, C. E., and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2006. http://www.gaussianprocess.org/gpml/.

[118] Rieger, C., Sampling inequalities and applications to regularization methods. Presented at *Workshop on Kernel Functions and Meshless Methods honoring the 65th birthday of Robert Schaback*, Göttingen, 2011.

[119] Rieger, C., R. Schaback, and B. Zwicknagl, Sampling and stability. In *Mathematical Methods for Curves and Surfaces 2008*, M. Dæhlen *et al.* (eds.), Lecture Notes in Computer Science 5862 (2010), 347–369. http://dx.doi.org/10.1007/978-3-642-11620-9_23.

[120] Rieger, C., and B. Zwicknagl, Sampling inequalities for infinitely smooth functions, with applications to interpolation and machine learning. *Adv. Comput. Math.* **32**/1 (2008), 103–129. http://dx.doi.org/10.1007/s10444-008-9089-0.

[121] Rippa, S., An algorithm for selecting a good value for the parameter c in radial basis function interpolation. *Adv. Comput. Math.* **11**/2-3 (1999), 193–210. http://dx.doi.org/10.1023/A:1018975909870.

[122] Riley, J. D., Solving systems of linear equations with a positive definite, symmetric, but possibly ill-conditioned matrix. *Mathematical Tables and Other Aids to Computation* **9**/51 (1955), 96–101. http://dx.doi.org/10.2307/2002065.

[123] Saitoh, S., *Theory of Reproducing Kernels and its Applications*. Longman Scientific, 1988.

[124] Saitoh, S., *Integral Transforms, Reproducing Kernels and Their Applications*. Chapman Hall, 1997.

[125] Schaback, R., Comparison of radial basis function interpolants. In *Multivariate Approximation: From CAGD to Wavelets*, K. Jetter and F. Utreras (eds.), World Scientific Publishing (Singapore) 1993, 293–305.

[126] Schaback, R., Creating surfaces from scattered data using radial basis functions. In *Mathematical Methods for Curves and Surfaces*, M. Dæhlen, T. Lyche, and L. Schumaker (eds.), Vanderbilt University Press (Nashville) 1995, 477–496.

[127] Schaback, R., Error estimates and condition numbers for radial basis function interpolation. *Adv. in Comput. Math.* **3** (1995), 251–264. http://dx.doi.org/10.1007/BF02432002.

[128] Schaback, R., Multivariate interpolation and approximation by translates of a basis function. In *Approximation Theory VIII, Vol. 1: Approximation and Interpolation*, C. Chui and L. Schumaker (eds.), World Scientific Publishing (Singapore) 1995, 491–514.

[129] Schaback, R., On the efficiency of interpolation by radial basis functions. In *Surface Fitting and Multiresolution Methods*, A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press (Nashville, TN) 1997, 309–318.

[130] Schaback, R., Native Hilbert spaces for radial basis functions I. In *New Developments in Approximation Theory*, M. W. Müller, M. D. Buhmann, D. H. Mache, and M. Felten (eds.), Birkhäuser (Basel) 1999, 255–282.

[131] Schaback, R., A unified theory of radial basis functions. Native Hilbert spaces for radial basis functions II. *J. Comput. Appl. Math.* **121** (2000), 165–177. http://dx.doi.org/10.1016/S0377-0427(00)00345-9.

[132] Schaback, R., Convergence of unsymmetric kernel-based meshless collocation methods. *SIAM J. Numer. Anal.* **45**/1 (2007), 333–351. http://dx.doi.org/10.1137/050633366.

[133] Schaback, R., Limit problems for interpolation by analytic radial basis functions. *J. Comp. Appl. Math.* **212**/2 (2008), 127–149. http:/dx.doi.org/10.1016/j.cam.2006.11.023.

[134] Schaback, R., Unsymmetric meshless methods for operator equations. *Numer. Math.* **114** (2010), 629–651. http://dx.doi.org/10.1007/s00211-009-0265-z.

[135] Schaback, R., and H. Wendland, Kernel techniques: From machine learning to meshless methods. *Acta Numerica* **15** (2006), 543–639. http://dx.doi.org/10.1017/S0962492906270016.

[136] Schaback, R., and Z. Wu, Operators on radial functions. *J. Comput. Appl. Math.* **73** (1996), 257–270. http://dx.doi.org/10.1016/0377-0427(96)00047-7.

[137] Scheuerer, M., An alternative procedure for selecting a good value for the parameter c in RBF-interpolation. *Adv. Comput. Math.* **34**/1 (2011), 105–126. http://dx.doi.org/10.1007/s10444-010-9146-3.

[138] Scheuerer, M., Kernel interpolation beyond the native space – a statistician's perspective. Presented at *Workshop on Kernel Functions and Meshless Methods honoring the 65th birthday of Robert Schaback*, Göttingen, 2011.

[139] Scheuerer, M., R. Schaback, and M. Schlather, Interpolation of spatial data – a stochastic or a deterministic problem? Preprint, Universität Göttingen. http://num.math.uni-goettingen.de/schaback/research/papers/IoSD.pdf.

[140] Schmidt, E., Über die Auflösung linearer Gleichungen mit unendlich vielen Unbekannten. *Rend. Circ. Mat. Palermo* **25** (1908), 53–77. http://dx.doi.org/10.1007/BF03029116.

[141] Schölkopf, B., and A. Smola, *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.

[142] Schoenberg, I. J., Metric spaces and completely monotone functions. *Ann. of Math.* **39** (1938), 811–841. http://dx.doi.org/10.2307/1968466.

[143] Schräder, D., and H. Wendland, A high-order, analytically divergence-free discretization method for Darcy's problem. *Math. Comp.* **80** (2011), 263–277. http://dx.doi.org/10.1090/S0025-5718-2010-02388-9.

[144] Shawe-Taylor, J., and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge Univ Press, 2004.

[145] Song, G., J. Riddle, G. E. Fasshauer, and F. J. Hickernell, Multivariate interpolation with increasingly flat radial basis functions of finite smoothness. *Adv. Comput. Math.* (2011). http://dx.doi.org/10.1007/s10444-011-9192-5.

[146] Stein, M. L., *Interpolation of Spatial Data. Some Theory for Kriging*. Springer Series in Statistics, Springer-Verlag (New York), 1999.

[147] Steinwart, I., and A. Christmann, *Support Vector Machines*. Springer Verlag, 2008.

[148] Stewart, J., Positive definite functions and generalizations, an historical survey. *Rocky Mountain J. Math.* **6** (1976), 409–434. http://dx.doi.org/10.1216/RMJ-1976-6-3-409.

[149] Trefethen, L. N., *Spectral Methods in* MATLAB. Society for Industrial and Applied Mathematics (Philadelphia, PA), 2000.

[150] Vapnik, V. N., *Statistical Learning Theory*. John Wiley & Sons, 1998.

[151] Wahba, G., *Splines Models for Observational Data*. Series in Applied Mathematics 59, SIAM, 1990.

[152] Wells, J. H., and L. R. Williams, *Embeddings and Extensions in Analysis*. Springer (Berlin), 1975.

[153] Wendland, H., Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. in Comput. Math.* **4** (1995), 389–396. http://dx.doi.org/10.1007/BF02123482.

[154] Wendland, H., Gaussian interpolation revisited. In *Trends in Approximation Theory*, K. Kopotun, T. Lyche, and M. Neamtu (eds.), Vanderbilt University Press, 2001, 417–426.

[155] Wendland, H., *Scattered Data Approximation*. Cambridge University Press, 2005.

[156] Wu, Z., Hermite-Birkhoff interpolation of scattered data by radial basis functions. *Approx. Theory Appl.* **8** (1992), 1–10. http://dx.doi.org/10.1007/BF02836101.

[157] Wu, Z., and R. Schaback, Local error estimates for radial basis function interpolation of scattered data. *IMA J. Numer. Anal.* **13** (1993), 13–27. http://dx.doi.org/10.1093/imanum/13.1.13.

[158] Zhou, D.-X., Sparsity of some learning algorithms. Presented at *Workshop on Kernel Functions and Meshless Methods honoring the 65th birthday of Robert Schaback*, Göttingen, 2011.

[159] Zwicknagl, B., Power series kernels, *Constr. Approx.* **29** (2009), 61–84. http://dx.doi.org/10.1007/s00365-008-9012-4.

[160] Zwicknagl, B., Series kernels and their approximation properties. Presented at *Workshop on Kernel Functions and Meshless Methods honoring the 65th birthday of Robert Schaback*, Göttingen, 2011.