Kernel–Based Meshless Methods

R. Schaback

July 18, 2007

Contents

1 Introduction

This is a text intended for use with my lecture "Approximationsverfahren II" in summer 2007. Though the basic background material is in the book [71] of Holger Wendland, some additional stuff is necessary at certain places, and I recycled larger parts of a 1997 lecture handout.

The text is under construction at various marked places, and it will evolve during the summer term. The chapter numbering is aligned with the numbering in the actual lecture and with the accompanying slides.

Göttingen, July 18, 2007

R. Schaback

2 Kernels

This chapter contains a collection of results on positive **semi**definite kernels, while the major literature focuses on positive definite kernels. See the old book of Meschkowski [42] and the recent dissertation of Roland Opfer [52].

2.1 Basics

Definition 2.1 *Let Ω be an arbitrary nonempty set. A function*

$$K \; : \; \Omega \times \Omega \to I\!R \text{ or } \mathbb{C}$$

is called a (real– or complex–valued) **kernel** *on Ω. We call K a* **hermitian kernel** *if*

$$K(x,y) = \overline{K(y,x)} \text{ for all } x,y \in \Omega.$$

If the kernel is real–valued, this property defines a **symmetric kernel**.

In most cases, we shall use only real–valued symmetric kernels, but for certain arguments we shall have to allow complex–valued kernels. For practical applications, one can consider hermitian kernels only, because for any kernel K we can go over to the hermitian kernel

$$\tilde{K}(x,y) := \frac{1}{2}(K(x,y) + \overline{K(y,x)}) \text{ for all } x,y \in \Omega.$$

To Do: convert everything as far as possible to allow complex–valued kernels.

Remember that Ω does not carry any structure at all. It can contain texts and images, for instance, and it will often be infinite. Some readers may consider this as being far too general. However, in the context of learning algorithms, the set Ω defines the possible *learning inputs*. Thus Ω should be general enough to allow Shakespeare texts or X-ray images, i.e. Ω should better have no predefined structure at all. Thus the kernels occurring in machine learning [65] are extremely general, but still they take a special form which can be tailored to meet the demands of applications. We shall later explain the recipes for their definition and usage.

In certain situations, a kernel is given a-priori, e.g. the *Gaussian*

$$K(x,y) := \exp(-\|x-y\|_2^2) \text{ for all } x,y \in \Omega := I\!R^d. \tag{2.2}$$

Each specific choice of a predefined kernel has a number of important and possibly unexpected consequences which we shall describe later.

If no predefined kernel is available for a certain set Ω, an application-dependent *feature map* $\Phi : \Omega \to \mathcal{F}$ with values in a Hilbert "*feature*" space $\mathcal{F}$ is defined. It should provide for each $x \in \Omega$ a large collection $\Phi(x)$ of *features* of x which are characteristic for x and which live in the Hilbert space $\mathcal{F}$ of high or even infinite dimension. Note the $\mathcal{F}$ has plenty of useful structure, while Ω has not. Feature maps $\Omega \to \mathcal{F}$ allow to apply linear techniques in their range $\mathcal{F}$, while their domain Ω is an unstructured set. They should be chosen carefully in an application-dependent way, capturing the essentials of elements of Ω.

With a feature map Φ at hand, there is a *kernel*

$$K(x,y) := (\Phi(x), \Phi(y))_{\mathcal{F}} \text{ for all } x,y \in \Omega \tag{2.3}$$

which is automatically hermitian.

In another important class of cases, the set Ω consists of random variables. Then the *covariance* between two random variables x and y from Ω is a standard choice of a kernel. These and other kernels arising in nondeterministic settings are dealt with in books on statistics. The connection to learning is obvious: two learning inputs x and y from Ω should be very similar, if they are closely "correlated", if they have very similar features, or if (2.3) takes large positive values. These examples again suggest to focus on symmetric kernels.

A kernel K on Ω defines a function $K(\cdot, y)$ for all fixed $y \in \Omega$. This allows to generate and manipulate spaces

$$\mathcal{K}_0 := \text{ span } \{K(\cdot, y) \ : \ y \in \Omega\}. \tag{2.4}$$

of functions on Ω. In Learning Theory, the function $K(\cdot, y) = (\Phi(\cdot), \Phi(y))_{\mathcal{F}}$ relates each other input object to a fixed object y via its essential features. But in general $\mathcal{K}_0$ just provides a handy linear space of *trial* functions on Ω which is extremely useful for most applications of kernels, e.g. when Ω consists of texts or images. For example, in meshless methods for solving partial differential equations, certain finite-dimensional subspaces of $\mathcal{K}_0$ are used as *trial* spaces to furnish good approximations to the solutions.

In certain other cases, the set Ω carries a measure μ, and then, under reasonable assumptions like $f,\ K(y,\cdot) \in L^2(\Omega, \mu)$, the generalized *convolution*

$$K *_\Omega f := \int_\Omega f(x) K(\cdot, x) d\mu(x) \tag{2.5}$$

defines an integral transform $f \mapsto K *_\Omega f$ which can be very useful. Note that Fourier or Hankel transforms arise this way, and recall the rôle of the Dirichlet kernel in Fourier analysis of univariate periodic functions. The above approach to kernels via convolution works on locally compact topological groups using Haar measure, but we do not want to pursue this detour into abstract harmonic analysis too far. For space reasons, we also have to exclude complex-valued kernels and all transform-type applications of kernels here, but it should be pointed out that

wavelets are special kernels of the above form, defining the *continuous wavelet transform* this way.

Note that discretization of the integral in the convolution transform leads to functions in the space $\mathcal{K}_0$ from (2.4). Using kernels as trial functions can be viewed as a discretized convolution. This is a very useful fact in the theoretical analysis of kernel-based techniques.

At this point, we skip over the various other occurrences of kernels in the mathematical literature and in applications (see the survey article [63]). Just keep in mind that kernels have three major application fields: they generate convolutions, trial spaces, and covariances. The first two are related by discretization.

But we recall that in machine learning and various other cases there are kernels of the **Hilbert–Schmidt** or **Mercer** form

$$K(x,y) = \sum_{i\in I} \lambda_i \varphi_i(x)\varphi_i(y) \text{ for all } x,y \in \Omega \tag{2.6}$$

with certain functions $\varphi : \Omega \to \mathbb{R}$, $i \in I$, certain positive **weights** λ_i, $i \in I$ and an **index set** I such that the summability conditions

$$K(x,x) := \sum_{i\in I} \lambda_i \varphi_i^2(x) < \infty \tag{2.7}$$

hold for all $x \in \Omega$. Note that this occurs in machine learning, if the functions φ_i each describe a **feature** of x, and if the **feature space** is the weighted ℓ_2 space

$$\ell_{2,I,\lambda} := \{\{\xi_i\}_{i\in I} : \sum_{i\in I} \lambda_i \xi_i^2 < \infty\} \tag{2.8}$$

of sequences with indices in I. But it also occurs when kernels generating positive integral operators are expanded into eigenfunctions φ_i on Ω (see Mercer's theorem 2.14 below), and such kernels can be viewed as arising from generalized convolution.

Note further that the summability condition (2.7) guarantees the well–definedness of the kernel by the Cauchy–Schwarz inequality

$$|K(x,y)| = \left| \sum_{i\in I} \left(\sqrt{\lambda_i}\varphi_i(x)\right) \cdot \left(\sqrt{\lambda_i}\varphi_i(y)\right) \right| \le \sqrt{K(x,x)K(y,y)} \text{ for all } x,y \in \Omega.$$

But there are many other kernels that have the above form. For instance, the univariate **Gaussian** kernel is

$$\begin{aligned} K(x,y) &:= \exp(-(x-y)^2) \\ &= \exp(-x^2)\exp(2xy)\exp(-y^2) \\ &= \exp(-x^2)\left(\sum_{n=0}^{\infty} \frac{2^n}{n!} x^n y^n\right)\exp(-y^2) \\ &= \sum_{n=0}^{\infty} \frac{2^n}{n!} \underbrace{x^n \exp(-x^2)}_{=:\varphi_n(x)} \underbrace{y^n \exp(-y^2)}_{=:\varphi_n(y)} \\ &= \sum_{n=0}^{\infty} \frac{2^n}{n!} \varphi_n(x)\varphi_n(y) \text{ for all } x,y \in \mathbb{R} \end{aligned}$$

without summability problems. But we shall postpone the construction of large classes of kernels to a later chapter.

2.2 Positive Definiteness

If we have an arbitrary set $X = \{x_1, \ldots, x_N\}$ of N distinct elements of Ω and a symmetric kernel K on Ω, we can form linear combinations

$$s(x) := \sum_{j=1}^{N} a_j K(x_j, x), \ x \in \Omega \tag{2.9}$$

of "translates" of the kernel. This is a very convenient technique to generate functions on an otherwise unstructured set Ω.

With such a set $X = \{x_1, \ldots, x_N\}$ we can form the symmetric $N \times N$ **kernel matrix**

$$A := A_{K,X,X} := (K(x_j, x_k))_{1 \le j,k \le N} \tag{2.10}$$

and pose the interpolation problem

$$\begin{aligned} y_k &= s(x_k), && 1 \le k \le N \\ y_k &= \sum_{j=1}^{N} a_j K(x_j, x_k), && 1 \le k \le N. \end{aligned} \tag{2.11}$$

In matrix notation, this is an $N \times N$ linear system

$$A_{K,X,X} a = y.$$

In general, solvability of such a system is a serious problem, but one of the central features of kernels and radial basis functions is to make this problem obsolete via

Definition 2.12 *A kernel K on Ω is called* **positive (semi–) definite,** *if for all sets $X = \{x_1, \ldots, x_N\}$ of N distinct elements of Ω and all N the $N \times N$* **kernel matrix** *(2.10) is positive (semi–) definite. This means in the complex–valued case that the quadratic form*

$$a \in \mathbb{C}^n \mapsto \sum_{j,k=1}^{N} a_j \overline{a_k} K(x_j, x_k)$$

is nonnegative, while in the positive definite case it is zero only if the vector a is zero.

Theorem 2.13 *Hilbert–Schmidt or Mercer kernels of the form (2.6) are positive semidefinite. Also, kernels arising from feature maps via (2.3) are positive semidefinite.*

Proof: The second statement is obvious, because kernels from feature maps generate kernel matrices that are Gramian matrices, and these are always positive semidefinite. To prove the first part, the quadratic form corresponding to the kernel matrix can be written as

$$\begin{aligned} a^T A_{K,X,X} a &= \sum_{j,k=1}^{N} a_j a_k K(x_j, x_k) \\ &= \sum_{j,k=1}^{N} a_j a_k \sum_{i \in I} \lambda_i \varphi_i(x_j) \varphi_i(x_k) \\ &= \sum_{i \in I} \lambda_i \sum_{j=1}^{N} a_j \varphi_i(x_j) \sum_{k=1}^{N} a_k \varphi_i(x_k) \\ &= \sum_{i \in I} \lambda_i \left(\sum_{j=1}^{N} a_j \varphi_i(x_j) \right)^2 \ge 0 \end{aligned}$$

for all vectors $a \in \mathbb{R}^N$. $\square$

At this point, we stick to positive semidefiniteness, but later we shall turn to positive definite kernels.

The basic connection of positive semidefinite kernels to a representation (2.6) is Mercer's

Theorem 2.14 *Suppose K is a continuous symmetric positive semidefinite kernel on a closed bounded interval $\Omega := [a,b] \subset \mathbb{R}$. Then there is an orthonormal basis $\{\varphi_i\}_{i\in\mathbb{N}}$ of $L_2[a,b]$ consisting of eigenfunctions of the linear integral operator defined by K such that the corresponding sequence of eigenvalues λ_i is nonnegative. This means*

$$\int_a^b K(x,y)\varphi_i(y)dy = \lambda_i\varphi_i(x) \text{ for all } x \in [a,b],\ i \in \mathbb{N}.$$

The eigenfunctions corresponding to non-zero eigenvalues are continuous on $[a,b]$ and K has the representation (2.6), where the convergence is absolute and uniform.

This theorem is contained in all reasonable books on Integral Equations or Functional Analysis. The background fact is that the operator

$$\varphi \mapsto \int_a^b K(x,y)\varphi(y)dy$$

is a compact "positive " integral operator on $L_2[a,b]$, and Mercer's theorem is a consequence of standard spectral theory in Hilbert spaces. Furthermore, all of this generalizes to domains and kernels in more than one dimension.

2.3 General Rules

We state some useful results on positive (semi)–definite kernels on some domain Ω.

Theorem 2.15 *Let K be a positive semidefinite kernel on Ω. Then*

$$\begin{array}{rcll} K(x,x) & \geq & 0 & \text{for all } x \in \Omega, \\ K(y,x) & = & \overline{K(x,y)} & \text{for all } x,y \in \Omega, \\ 2|K(x,y)|^2 & \leq & K(x,x)+K(y,y) & \text{for all } x,y \in \Omega, \\ |K(x,y)|^2 & \leq & K(x,x)\cdot K(y,y) & \text{for all } x,y \in \Omega. \end{array}$$

Furthermore, any finite linear combination of positive semidefinite kernels with nonnegative coefficients yields a positive definite kernel (this means that positive definite kernels form a **convex cone***). If one of the kernels is positive definite, and if its factor is positive, the superposition of kernels is positive definite. Finally, the product of two positive semidefinite kernels is positive semidefinite.*

Proof: For the first property, use $X = \{x\}$ in Definition 2.12. The second property implies that positive semidefinite kernels are always hermitian, and the proof uses $X = \{x,y\}$ and coefficients $(1,c)$ with $c \in \mathbb{C}$. Then

$$K(x,x) + K(y,y) + \overline{c}K(x,y) + c\overline{K(y,x)} \geq 0$$

is real–valued for all $c \in \mathbb{C}$, and setting $c = 1$ and $c = i$ we get the assertion. The third case follows if we set $c = -K(x, y)$ above. For the fourth assertion, we note that for x, y fixed, the sesquilinear semi–inner product

$$(\alpha, \beta) := \alpha A \overline{\beta} \text{ for all } \alpha, \beta \in \mathbb{C}^2$$

with A being the kernel matrix for $X = \{x, y\}$ satisfies the Cauchy–Schwarz inequality. Thus

$$|((1,0),(0,1))|^2 = |K(x,y)|^2 \leq |(1,0)||(0,1)| = K(x,x)K(y,y).$$

The statements on nonnegative linear superposition are very easy to see.

Thus we are left with the final assertion, which is nontrivial. Assume two positive semidefinite kernels K and L to be given, and take a set X of N points of Ω and a coefficient vector $a \in \mathbb{C}^N$. We have to prove nonnegativity of the quadratic form

$$Q := \sum_{j,k=1}^{N} a_j \overline{a_k} K(x_j, x_k) L(x_j, x_k).$$

Since the kernel matrix A for K is positive semidefinite, we can transform it to a diagonal matrix with nonnegative diagonal entries $\lambda_1, \ldots, \lambda_N$ by a unitary matrix S. This means that

$$K(x_j, x_k) = \sum_{m=1}^{N} \lambda_m s_{j,m} \overline{s_{k,m}}$$

with complex $s_{j,k}$ and we can insert this into our quadratic form to get

$$\begin{aligned} Q &= \sum_{j,k=1}^{N} a_j \overline{a_k} L(x_j, x_k) \sum_{m=1}^{N} \lambda_m s_{j,m} \overline{s_{k,m}} \\ &= \sum_{m=1}^{N} \lambda_m \sum_{j,k=1}^{N} \underbrace{a_j s_{j,m}}_{=:b_{j,m}} \overline{a_k s_{k,m}} L(x_j, x_k) \\ &= \sum_{m=1}^{N} \lambda_m \underbrace{\sum_{j,k=1}^{N} b_{j,m} \overline{b_{k,m}} L(x_j, x_k)}_{\geq 0} \geq 0. \end{aligned}$$

□

We leave it to the reader to use some linear algebra to prove

Corollary 2.16 *The product of two positive definite kernels is positive definite.* □

For later use, we add another superposition principle, applying **generalized convolution**. If

$$L : \Omega \times Z \to \mathbb{C}$$

is an arbitrary function, and if we take any set of points $z_1, \ldots, z_m \in Z$, we can form a kernel

$$K(x, y) := \sum_{\ell=1}^{m} c_\ell L(x, z_\ell) \overline{L(y, z_\ell)}$$

when taking nonnegative coefficients $c_1, \ldots, c_m$. The kernel K will be hermitian, and positive semidefinite due to

$$\begin{aligned} & \sum_{j,k=1}^{N} a_j \overline{a_k} K(x_j, x_k) \\ = & \sum_{j,k=1}^{N} a_j \overline{a_k} \sum_{\ell=1}^{m} c_\ell L(x_j, z_\ell) \overline{L(x_k, z_\ell)} \\ = & \sum_{\ell=1}^{m} c_\ell \sum_{j,k=1}^{N} a_j L(x_j, z_\ell) \overline{a_k L(x_k, z_\ell)} \\ = & \sum_{\ell=1}^{m} c_\ell \left| \sum_{j=1}^{N} a_j L(x_j, z_\ell) \right|^2 \geq 0. \end{aligned}$$

This generalizes easily to cases where the sum can be replaced by an integral, e.g.

$$K(x, y) := \int_Z c(z) L(x, z) \overline{L(y, z)} dz, \ x, y \in \Omega$$

with a nonnegative function c, provided that the above is well–defined and finite. This holds whenever

$$K(x, x) = \int_Z c(z) |L(x, z)|^2 dz$$

is well–defined and finite for all $x \in \Omega$, due to the Cauchy–Schwarz inequality. Applying measure theory, on can also go over to

$$K(x, y) := \int_Z L(x, z) \overline{L(y, z)} d\mu(z), \ x, y \in \Omega$$

with a nonnegative measure μ on Z, using

$$K(x, x) = \int_Z |L(x, z)|^2 d\mu(z)$$

as a sufficient condition for well–definedness of the new kernel.

But note that the above argument is nothing else than the transition to a suitable feature space. If

$$\Phi(x) := L(x, \cdot)$$

maps Ω into a suitable function space $\mathcal{F}$ consisting of functions on Z as a feature space, we can write each instance of the above construction in the form (2.3). Thus positive semidefiniteness of such kernels is no miracle.

2.4 Inner Product

The following construction is of utmost importance for kernel–based techniques. We assume K to be a symmetric real–valued positive semidefinite kernel on Ω, and we form the space

$$S := S_\Omega := \text{ span } \{K(x, \cdot) \ : \ x \in \Omega\} \tag{2.17}$$

of all finite linear combinations of the functions $K(x, \cdot) \ : \ \Omega \to I\!R$. Note that general elements from S take the form

$$f_{a,X}(\cdot) = \sum_{j=1}^{N} a_j K(x_j, \cdot) \tag{2.18}$$

with $a \in I\!R^N$ while $X = \{x_1, \ldots, x_N\} \subset \Omega$, but different N and all point sets X are allowed.

To Do: Convert to complex case...

On S we can define a bilinear form

$$\begin{aligned}\left(\underbrace{\sum_{j=1}^{M} a_j K(x_j, \cdot)}_{=: f_{a,X}(\cdot)}, \underbrace{\sum_{k=1}^{N} b_k K(y_k, \cdot)}_{=: f_{b,Y}(\cdot)} \right)_K &:= \sum_{j=1}^{M} \sum_{k=1}^{N} a_j b_k K(x_j, y_k) \\ &= \sum_{j=1}^{M} a_j f_{b,Y}(x_j) \\ &= \sum_{k=1}^{N} b_k f_{a,X}(y_k).\end{aligned} \tag{2.19}$$

To prove that it is well–defined, we re–represent the functions $f_{a,X}$ and $f_{b,Y}$ in different form as

$$\begin{aligned} f_{a,X} &= \sum_{j=1}^{\tilde{M}} \tilde{a}_j K(\tilde{x}_j, \cdot) = f_{\tilde{a},\tilde{X}} \\ f_{b,Y} &= \sum_{k=1}^{\tilde{N}} \tilde{b}_k K(\tilde{y}_k, \cdot) = f_{\tilde{b},\tilde{Y}} \end{aligned}$$

and check the result:

$$\begin{aligned}(f_{a,X}, f_{b,Y})_K &= \sum_{j=1}^{M} \sum_{k=1}^{N} a_j b_k K(x_j, y_k) \\ &= \sum_{j=1}^{M} a_j f_{b,Y}(x_j) \\ &= \sum_{j=1}^{M} a_j \sum_{k=1}^{\tilde{N}} \tilde{b}_k K(\tilde{y}_k, x_j) \\ &= \sum_{k=1}^{\tilde{N}} \tilde{b}_k \sum_{j=1}^{M} a_j K(\tilde{y}_k, x_j) \\ &= \sum_{k=1}^{\tilde{N}} \tilde{b}_k f_{a,X}(\tilde{y}_k) \\ &= \sum_{k=1}^{\tilde{N}} \tilde{b}_k \sum_{j=1}^{\tilde{M}} \tilde{a}_j K(\tilde{x}_j, \tilde{y}_k) \\ &= \sum_{j=1}^{\tilde{M}} \sum_{k=1}^{\tilde{N}} \tilde{a}_j \tilde{b}_k K(\tilde{x}_j, \tilde{y}_k) \\ &= (f_{\tilde{a},\tilde{X}}, f_{\tilde{b},\tilde{Y}})_K\end{aligned}$$

to see that it is independent of the representation. Furthermore, we have a positive semidefinite bilinear form due to the positive semidefiniteness of all kernel matrices.

When specializing (2.19) partially to functions $K(y, \cdot)$ based on single points, we get the extremely useful **reproduction equation**

$$(f, K(y, \cdot))_K = f(y) \text{ for all } y \in \Omega,\ f \in S \tag{2.20}$$

and its special case

$$(K(x,\cdot),K(y,\cdot))_K \;=\; K(x,y) \text{ for all } x,y\in\Omega. \tag{2.21}$$

Strangely enough, the bilinear form is even positive definite:

Theorem 2.22 *If K is a positive semidefinite symmetric kernel on Ω, the bilinear form $(.,.)_K$ of (2.19) is positive definite on the space S of (2.17) as a space of functions on Ω. Thus S is a pre–Hilbert or Euclidean space of functions on Ω.*

Proof: Assume that

$$(f_{a,X},f_{a,X})_K=\sum_{j,k=1}^N a_ja_kK(x_j,x_k)=\sum_{j=1}^N a_jf_{a,X}(x_j)=0$$

for $a\in\mathbb{R}^N$ and $X=\{x_1,\ldots,x_N\}\subset\Omega$. Then by (2.20) and the Cauchy–Schwarz inequality we have

$$|f_{a,X}(x)|^2=|(f_{a,X},K(x,\cdot))_K|^2\le(f_{a,X},f_{a,X})_K(K(x,\cdot),K(x,\cdot))_K=0$$

for all $x\in\Omega$. □

Note that the above argument does not imply $a=0$ for $f_{a,X}(\cdot)=0$. But there are some other useful implications:

$$\begin{aligned}|f(x)| &\le \|f\|_K\sqrt{K(x,x)} \text{ for all } x\in\Omega,\ f\in S\\ |f(x)-f(y)| &\le \|f\|_K\sqrt{K(x,x)-2K(x,y)+K(y,y)} \text{ for all } x,y\in\Omega,\ f\in S\end{aligned}$$

where we now can use the norm notation for $\|f\|_K^2=(f,f)_K$ and all $f\in S$.

2.5 Duality

We now consider the **dual space** S^* to S. It contains all bounded linear functionals λ on S and it has a **dual norm**

$$\|\lambda\|_{S^*}:=\sup_{f\in S\setminus\{0\}}\frac{\lambda(f)}{\|f\|_K} \text{ for all } \lambda\in S^*.$$

But we assert that we can write the dual norm via an inner product which we can define as

$$(\lambda,\mu)_K:=(\lambda^xK(x,\cdot),\mu^yK(y,\cdot))_K \text{ for all } \lambda,\mu\in S$$

where the notation $\lambda^xK(x,\cdot)$ means that λ acts with respect to the variable x. Clearly, the right–hand side is a well–defined bilinear form, but we postpone to prove its positive definiteness for a moment. Instead, we want to prove the generalized reproduction equation

$$\lambda(f)=(f,\lambda^xK(x,\cdot))_K \text{ for all } f\in S,\ \lambda\in S^*. \tag{2.23}$$

It suffices to do this on all functions $f_y:=K(y,\cdot)$, and we get

$$(f_y,\lambda^xK(x,\cdot))_K=(K(y,\cdot),\lambda^xK(x,\cdot))_K=\lambda^xK(x,y)=\lambda^xK(y,x)=\lambda(f_y).$$

Now if $(\lambda,\lambda)_K = 0$, we have $\lambda^x K(x,\cdot) = 0$ on Ω and $\lambda(f) = 0$ for all $f \in S$ by the generalized reproduction equation, leading to $\lambda = 0$ and proving definiteness of the dual bilinear form. Furthermore, by the definition of the dual norm and the dual bilinear form we have

$$\|\lambda\|_{S^*} = \sup_{f \in S\setminus\{0\}} \frac{\lambda(f)}{\|f\|_K} \leq \|\lambda^x K(x,\cdot)\|_K = \|\lambda\|_K,$$

but we get equality for $f = \lambda^x K(x,\cdot)$ due to

$$\frac{\lambda(\lambda^x K(x,\cdot))}{\|\lambda^x K(x,\cdot)\|_K} = \frac{\|\lambda^x K(x,\cdot)\|_K^2}{\|\lambda^x K(x,\cdot)\|_K} = \|\lambda^x K(x,\cdot)\|_K.$$

This implies that the dual space S^* of S is again a pre–Hilbert space under the dual inner product above.

By the reproduction equation, the point evaluation functionals

$$\delta_x \ : \ S \to I\!R, \ f \mapsto f(x)$$

satisfy

$$\delta_x(f) = f(x) = (f, K(x,\cdot)_K \text{ for all } f \in S, \ x \in \Omega$$

and thus are bounded and continuous via

$$|\delta_x(f)| = |f(x)| = |(f, K(x,\cdot)_K| \leq \|f\|_K \|K(x,\cdot)\|_K = \|f\|_K \sqrt{K(x,x)} \text{ for all } f \in S, \ x \in \Omega.$$

The dual S^* of S thus contains all point evaluation functionals, and we get

$$(\delta_x, \delta_y)_K = (K(x,\cdot), K(y,\cdot))_K = K(x,y) \text{ for all } x, y \in \Omega.$$

In particular, we shall often use the identity

$$\|\delta_x - \delta_y\|_K^2 = \|\delta_x\|_K^2 - 2(\delta_x, \delta_y)_K + \|\delta_y\|_K^2 = K(x,x) - 2K(x,y) + K(y,y) \text{ for all } x, y \in \Omega.$$

This leads to a notion of a **distance** on Ω via

$$dist(x,y) := \|\delta_x - \delta_y\|_K = \sqrt{K(x,x) - 2K(x,y) + K(y,y)} \text{ for all } x, y \in \Omega.$$

In this special distance, all functions in S are "continuous" due to

$$|f(x) - f(y)| \leq \|f\|_K \|\delta_x - \delta_y\|_K = \|f\|_K dist(x,y) \text{ for all } x, y \in \Omega, \ f \in S.$$

2.6 Native Space

We now know that S is an inner–product or semi–Hilbert space of functions on Ω under the inner product $(.,.)_K$, provided that K is a positive semidefinite symmetric kernel on Ω. Then we can invoke a classical argument from Hilbert space theory to go over the closure of S under $(.,.)_K$. This is an abstract space defined by equivalence classes of Cauchy sequences in S, but it is a complete space (thus a Hilbert space), and each continuous map from S to a Banach space Y extends uniquely to the closure.

Theorem 2.24 *Each symmetric positive semidefinite kernel K on a set Ω is the reproducing kernel of a Hilbert space called the* **native space** $\mathcal{N}_K$ *of the kernel. This Hilbert space is unique, and it is a space of functions on Ω. The kernel K is a* **reproducing kernel** *of $\mathcal{N}_K$ in the sense*

$$(f, K(y,\cdot))_K = f(y) \text{ for all } y \in \Omega,\ f \in \mathcal{N}_K$$

generalizing (2.20).

Proof: The existence of the native space follows from standard Hilbert space arguments we do not repeat here, see section 9.8. Since (2.20) is an equation with both sides being continuously dependent on $f \in S$, it carries over to the closure and thus to the native space, proving the reproduction formula above. But then it explains how an abstract element f of the native space can be interpreted as a function: just use the left–hand side as a definition of the right–hand side.

If K is reproducing in a possibly different Hilbert space T with an analogous reproduction equation, we can use (2.21) and the reproduction equation in T to conclude

$$K(x,y) = (K(x,\cdot), K(y,\cdot))_K = (K(x,\cdot), K(y,\cdot))_T,$$

and this proves that the inner products of T and $\mathcal{N}_K$ coincide on S. Since T is a Hilbert space, it must then contain the closure $\mathcal{N}_K$ of S as a closed subspace. If T were larger than $\mathcal{N}_K$, there must be a nonzero element $f \in T$ which is orthogonal to $\mathcal{N}_K$ and in particular to S. But then

$$f(y) = (f, K(y,\cdot))_T = 0 \text{ for all } y \in \Omega$$

is a contradiction. □

Note that usually the Hilbert space closure of an inner–product space is considerably "larger" than the space itself. This is very much like the transition from rational numbers to real numbers.

We should have a quick look at point evaluation functionals

$$\delta_x\ :\ \mathcal{N}_K \to I\!R,\ f \mapsto f(x) \text{ for all } f \in \mathcal{N}_K$$

for $x \in \Omega$. Note that the dual space $\mathcal{N}_K^*$ of the native space is again a Hilbert space with an inner product and norm, which is isometrically isomorphic to $\mathcal{N}_K$ itself via the Riesz map

$$\begin{aligned} R &: \mathcal{N}_K^* \to \mathcal{N}_K, \\ \lambda(f) &= (f, R(\lambda))_K \text{ for all } f \in \mathcal{N}_K,\ \lambda \in \mathcal{N}_K^*, \\ (\lambda,\mu)_K &= (R(\lambda), R(\mu))_K \text{ for all } \lambda,\mu \in \mathcal{N}_K^*, \end{aligned}$$

where we denote the dual inner product in $\mathcal{N}_K^*$ again by $(.,.)_K$ for simplicity.

The reproduction equation tells us that

$$\delta_x(f) = (f, K(x,\cdot))_K \text{ for all } f \in \mathcal{N}_K,\ x \in \Omega,$$

and we immediately see that $K(x,\cdot)$ is the Riesz representer $R(\delta_x)$ of δ_x in $\mathcal{N}_K$, leading directly to

$$(\delta_x, \delta_y)_K = (R(\delta_x), R(\delta_y))_K = (K(x,\cdot), K(y,\cdot))_K = K(x,y) \text{ for all } x,y \in \Omega$$

and

$$\|\delta_x\|_K = \|K(x,\cdot)\|_K = \sqrt{K(x,x)} \text{ for all } x \in \Omega \tag{2.25}$$

because the Riesz map is an isometry. Similarly, we have the extended reproduction property

$$\lambda(f) = (f, \lambda^x K(x,\cdot))_K \text{ for all } f \in \mathcal{N}_K,\ \lambda \in \mathcal{N}_K^*$$

telling us that $\lambda^x K(x,\cdot)$ is the Riesz representer of λ.

2.7 Reproducing Kernel Hilbert Spaces

The theoretical background for all of this is

Definition 2.26 *A Hilbert space $\mathcal{H}$ of functions on a set Ω with inner product $(.,.)_{\mathcal{H}}$ is called a* **reproducing kernel Hilbert space** *(RKHS), if there is a kernel function $K : \Omega \to \mathbb{R}$ with $K(x,\cdot) \in \mathcal{H}$ for all $x \in \Omega$ and the reproduction property*

$$f(x) = (f, K(x,\cdot))_{\mathcal{H}} \text{ for all } x \in \Omega,\ f \in \mathcal{H}.$$

This implies

$$(K(y,\cdot), K(x,\cdot))_{\mathcal{H}} = K(y,x) = K(x,y) \text{ for all } x, y \in \Omega,$$

and it is easy to verify that K is positive semidefinite. In fact, if we take $X = \{x_1 \ldots, x_N\} \subset \Omega$ and a vector $a \in \mathbb{R}^N$, we get

$$\begin{aligned}
& \sum_{j,k=1}^{N} a_j a_k K(x_j, x_k) \\
= & \sum_{j,k=1}^{N} a_j a_k (K(x_j,\cdot), K(x_k,\cdot))_K \\
= & \left(\sum_{j=1}^{N} a_j K(x_j,\cdot), \sum_{k=1}^{N} a_k K(x_k,\cdot) \right)_K \\
= & \left\| \sum_{j=1}^{N} a_j K(x_j,\cdot) \right\|_K^2 \geq 0
\end{aligned}$$

so that the kernel matrix is positive semidefinite.

In the previous section we have proven

Theorem 2.27 *Every positive semidefinite symmetric kernel K on some set Ω is the reproducing kernel of some ("native") Hilbert space $\mathcal{N}_K$ of functions on Ω in which the point evaluation functionals δ_x are continuous and have the kernel functions $K(x,\cdot)$ as Riesz representers.* □

Now we go for the converse:

Theorem 2.28 *Let $\mathcal{H}$ be a Hilbert space of functions on Ω such that all point evaluation functionals*

$$\delta_x : f \mapsto f(x) \text{ for all } f \in \mathcal{H}$$

for $x \in \Omega$ are continuous. Then $\mathcal{H}$ is a reproducing kernel Hilbert space with a positive semidefinite kernel K on Ω, and the kernel is uniquely defined by providing the Riesz representers of the point evaluation functionals. Finally, the space $\mathcal{H}$ is the native space for the kernel.

Proof: Under the hypothesis of the theorem, there must be a Riesz representer of δ_x, and by definition of the Riesz map it takes the form $K(x,\cdot) \in \mathcal{H}$ satisfying the reproduction equation. Thus any such Hilbert space has a positive semidefinite symmetric reproducing kernel. The final assertion follows from Theorem 2.24, because both the native space and $\mathcal{H}$ are Hilbert spaces which contain all $K(x,\cdot)$. □

2.8 Kernels for Orthogonal Expansions

Let us look at the special case where a Hilbert space $\mathcal{H}$ has a complete orthonormal basis $\{\varphi_i\}_{i\in I}$ of functions in Ω. A special case are trigonometric polynomials in the space of square–integrable 2π–periodic functions, or any space of functions spanned by orthogonal polynomials.

Then each $f \in \mathcal{H}$ has a unique expansion

$$f = \sum_{i\in I} (f,\varphi_i)_{\mathcal{H}} \varphi_i$$

with the Parseval equation

$$\|f\|_{\mathcal{H}}^2 = \sum_{i\in I} (f,\varphi_i)_{\mathcal{H}}^2 < \infty.$$

In many cases, including trigonometric or orthogonal algebraic polynomials, the expansions of functions in $\mathcal{H}$ do not converge pointwise, but only in the Hilbert space norm. Thus point–evaluation functionals are not continuous on $\mathcal{H}$. The situation is better if the coefficients of the expansion satisfies a decay condition, and this condition defines a closed space of $\mathcal{H}$ if posed correctly. An example for trigonometric series

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos(nx) + b_n \sin(nx)\right) \tag{2.29}$$

is the condition

$$\sum_{n=1}^{\infty} n^2 \left(a_n^2 + b_n^2\right) < \infty$$

because it implies that the series for $f(x)$ and $f'(x)$ are absolutely convergent.

If we have weights λ_i such that the summability condition (2.7) holds, we have boundedness of point evaluation via

$$\begin{aligned} |f(x)| &\le \sum_{i\in I} |(f,\varphi_i)_{\mathcal{H}}||\varphi_i(x)| \\ &= \sum_{i\in I} \frac{|(f,\varphi_i)_{\mathcal{H}}|}{\sqrt{\lambda_i}} |\varphi_i(x)| \sqrt{\lambda_i} \\ &\le \sqrt{\sum_{i\in I} \frac{(f,\varphi_i)_{\mathcal{H}}^2}{\lambda_i}} \sqrt{\sum_{i\in I} \varphi_i^2(x)\lambda_i} \end{aligned}$$

in the subspace

$$\mathcal{H}_\lambda := \left\{ f \in \mathcal{H} \;:\; \|f\|_\lambda^2 := \sum_{i\in I} \frac{(f,\varphi_i)_{\mathcal{H}}^2}{\lambda_i} < \infty \right\}$$

of functions with a suitable summability condition for the coefficients. This space has a norm which arises from the inner product

$$(f,g)_\lambda := \sum_{i\in I} \frac{(f,\varphi_i)_{\mathcal{H}} (g,\varphi_i)_{\mathcal{H}}}{\lambda_i} \text{ for all } f,g \in \mathcal{H}_\lambda.$$

We now define the Mercer kernel (2.6) and check whether all $f_x := K(x,\cdot)$ lie in $\mathcal{H}_\lambda$. This follows from the fact that it has the expansion coefficients

$$(f_x,\varphi_i)_{\mathcal{H}} = \lambda_i\varphi_i(x)$$

with the summability

$$\sum_{i\in I}\frac{(f_x,\varphi_i)_{\mathcal{H}}^2}{\lambda_i} = \sum_{i\in I}\varphi_i^2(x)\lambda_i < \infty.$$

Each function $f\in\mathcal{H}_\lambda$ satisfies the reproduction equation

$$\begin{aligned}(f,K(x,\cdot))_\lambda &= \sum_{i\in I}\frac{(f,\varphi_i)_{\mathcal{H}}(K(x,\cdot),\varphi_i)_{\mathcal{H}}}{\lambda_i}\\ &= \sum_{i\in I}\frac{(f,\varphi_i)_{\mathcal{H}}\lambda_i\varphi_i(x)}{\lambda_i}\\ &= f(x) \text{ for all } x\in\Omega,\ f\in\mathcal{H}_\lambda.\end{aligned}$$

Thus the Mercer kernel is reproducing in $\mathcal{H}_\lambda$. This proves

Theorem 2.30 *If a Hilbert space of functions on Ω has a countable orthonormal basis $\{\varphi_i\}_{i\in I}$, each summability property of the form (2.7) leads to a reproducing Mercer kernel for a suitable subspace of functions with continuous point evaluation.* □

We add without proof that spaces like $\mathcal{H}_\lambda$ are always complete because they are isometrically isomorphic to certain Hilbert spaces of weighted sequences (see section 9.9 for details). Thus

Corollary 2.31 *The spaces $\mathcal{H}_\lambda$ defined above are the native spaces for the corresponding Mercer kernels.*

Let us look at trigonometric polynomials as an example. The basic space $\mathcal{H}$ is the space of 2π–periodic square integrable functions with the inner product

$$(f,g)_{\mathcal{H}} := \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)g(t)dt$$

and with the orthonormal functions

$$\frac{1}{\sqrt{2}},\cos(nx),\sin(nx),\ n\in I\!N.$$

We can write these via the index set

$$I := (0,0)\cup(I\!N,0)\cup(0,I\!N)$$

as

$$\varphi_i(x) := \begin{cases}\frac{1}{\sqrt{2}} & i=(0,0)\\ \cos(nx) & i=(n,0),\ n\geq 1\\ \sin(nx) & i=(0,n),\ n\geq 1.\end{cases}$$

Note that all functions are uniformly bounded, such that the summability condition (2.7) works whenever the weights are summable. We fix some $m\geq 1$ and define

$$\lambda_i := \begin{cases}1 & i=(0,0)\\ n^{-2m} & i=(n,0),\ n\geq 1\\ n^{-2m} & i=(0,n),\ n\geq 1\end{cases}$$

to get the Mercer kernel

$$\begin{aligned} K_{2m}(x,y) &:= \frac{1}{\sqrt{2}} + \sum_{n=1}^{\infty} n^{-2m} \left(\cos(nx)\cos(ny) + \sin(nx)\sin(ny)\right) \\ &= \frac{1}{\sqrt{2}} + \sum_{n=1}^{\infty} n^{-2m} \cos(n(x-y)) \end{aligned}$$

which must be positive semidefinite on $\Omega = [0, 2\pi)$. Plotting the kernel K_2 (see Figure 1) reveals that it is a continuous piecewise parabola, and from $K_{2m}'' = -K_{2m-2}$ for large m we see that K_{2m} must be a piecewise polynomial of degree $2m$ which is still $2m-2$ times continuously differentiable.

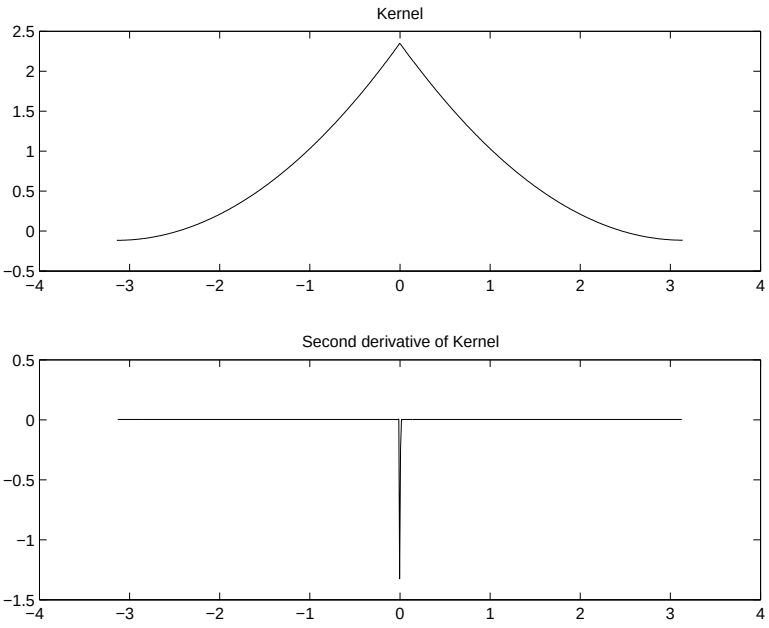

Figure 1: The kernel K_2 and its second derivative

To verify this, we suspect K_2 to be something like $g(t) := (\pi - t)^2$ on $[0, \pi]$ with periodic continuation to an even 2π–periodic function. We calculate the even Fourier coefficients as

$$\begin{aligned} & (g(t), \cos(nt))_{\mathcal{H}} \\ = \; & \frac{2}{\pi} \int_0^{\pi} (\pi - t)^2 \cos(nt) dt \\ = \; & \left[\frac{2}{n\pi}(\pi - t)^2 \sin(nt)\right]_0^{\pi} + \frac{4}{n\pi} \int_0^{\pi} (\pi - t)\sin(nt) dt \\ = \; & 0 + \frac{4}{n^2\pi}\left[-(\pi - t)\cos(nt)\right]_0^{\pi} - \frac{4}{n^2\pi} \int_0^{\pi} \cos(nt) dt \\ = \; & \frac{4}{n^2} \end{aligned}$$

and

$$\begin{aligned} & (g(t), \frac{1}{\sqrt{2}})_{\mathcal{H}} \\ = & \frac{2}{\pi} \int_0^{\pi} (\pi - t)^2 \frac{1}{\sqrt{2}} dt \\ = & \frac{\sqrt{2}\pi^2}{3} \end{aligned}$$

such that we get

$$K_2(t) = \frac{1}{4} g(t) + \frac{1}{\sqrt{2}} - \frac{\pi^2}{12}.$$

We note that periodic functions of this form arise in the context of Romberg integration. The native space for K_{2m} contains all functions with Fourier series coefficients satisfying the summability condition in $\mathcal{H}_\lambda$, which in case of (2.29) and K_{2m} takes the form

$$\sum_{n \in I\!N} n^{2m} \left(a_n^2 + b_n^2 \right) < \infty.$$

Thus the functions in the native space for K_{2m} get more and more smooth for increasing m. Readers familiar with Sobolev spaces will recognize that K_{2m} is the reproducing kernel of the **Sobolev space** of order $2m$ for univariate 2π–periodic functions.

From Anette Meyenburg's thesis [43] we cite some other cases, with possibly different additive constants than used here, and with $t = x - y$ and on $[0, 2\pi]$:

$$\begin{aligned} K_2(t) &= \frac{3t^2 - 6\pi t + 2\pi^2}{12} \\ K_4(t) &= -\frac{t^4}{90} + \frac{\pi t^3}{12} - \frac{\pi^2 t^2}{12} + \frac{\pi^4}{90} \\ K_6(t) &= \frac{t^6}{1440} - \frac{\pi t^5}{240} + \frac{\pi^2 t^4}{144} - \frac{\pi^4 t^2}{180} + \frac{\pi^6}{945}. \end{aligned}$$

Furthermore, there are the infinitely differentiable periodic kernels

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{1}{n!} \cos(nx) &= \cos(\sin(x)) \cdot \exp(\cos(x)) \\ \sum_{n=0}^{\infty} \frac{1}{2^n} \cos(nx) &= \frac{1 - \frac{1}{2}\cos(x)}{1 - \cos(x) + \frac{1}{4}}. \end{aligned}$$

Without any further work we know that their native spaces consist of 2π–periodic functions whose Fourier coefficients decay like $\frac{1}{n!}$ or $\frac{1}{2^n}$, respectively.

Note that users can specify the decay of the spectrum of the functions they work with by choosing an appropriate kernel. Furthermore, the above theory applies similarly to expansions into algebraic orthogonal polynomials (Chebyshev–, Legendre–, Jacobi–, Hermite–) and to expansions on the sphere into spherical harmonics. A particularly nice case is the formula

$$\sum_{n=0}^{\infty} H_n(x) H_n(y) \frac{t^n}{n!} = (1 - t^2)^{-1/2} \exp\left(-\frac{x^2 t^2 - 2txy + y^2 t^2}{2(1 - t^2)} \right), \; x, y \in I\!R, \; -1 < t < 1$$

using Hermite polynomials, and it is due to Mehler, as cited from Tricomi's nice book on Orthogonal Series, p. 254.

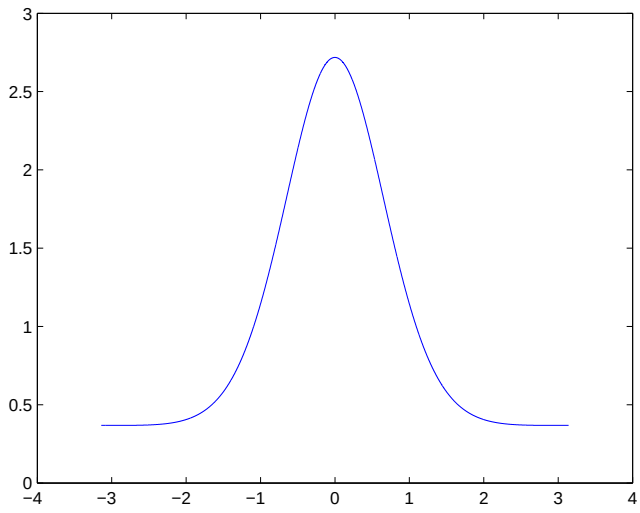

Figure 2: Kernel $\cos(\sin(x)) \cdot \exp(\cos(x))$

2.9 Native Spaces of Mercer Kernels

We now want to turn the above situation upside down, starting with a Mercer kernel and nothing else. We want to arrive at a Hilbert space of functions on Ω with continuous point evaluation such that the Mercer kernel is reproducing.

Thus we start with a Hilbert–Schmidt or Mercer kernel (2.6) with the summability condition (2.7) to make it pointwise well–defined. We want to write the kernel via a feature space, and this leads to the feature space $\ell_{2,\lambda,I}$ of (2.8) under the inner product

$$(\{\xi_i\}_{i\in I}, \{\eta_i\}_{i\in I})_{\lambda,I} := \sum_{i\in I} \lambda_i \xi_i \eta_i.$$

The feature map

$$\Phi(x) := \{\varphi_i(x)\}_{i\in I} \in \ell_{2,\lambda,I} \text{ for all } x \in \Omega$$

takes Ω into a set $\Phi(\Omega) \subseteq \ell_{2,\lambda,I}$, and the kernel is

$$K(x,y) = (\Phi(x), \Phi(y))_{\lambda,I} \text{ for all } x, y \in \Omega.$$

The functions $K(x,\cdot)$ take the form

$$\begin{aligned} K(x,\cdot) &= \sum_{i\in I} \underbrace{\lambda_i \varphi_i(x)}_{=:c_i(x)} \varphi_i(\cdot) \\ c_i(x) &= \lambda_i \varphi_i(x),\ i \in I \\ \sum_{i\in I} \frac{c_i^2}{\lambda_i} &= \sum_{i\in I} \lambda_i \varphi_i^2(x) < \infty \end{aligned}$$

and thus they are in the function space

$$\mathcal{H} := \left\{ \sum_{i\in I} c_i \varphi_i(\cdot) \ : \ \sum_{i\in I} \frac{c_i^2}{\lambda_i} < \infty \right\} = T(\ell_{2,1/\lambda,I})$$

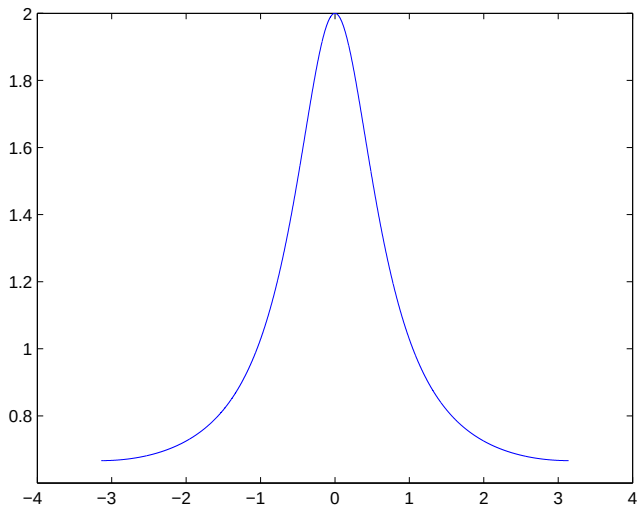

Figure 3: Kernel $\frac{1-\frac{1}{2}\cos(x)}{1-\cos(x)+\frac{1}{4}}$

with the surjective (but not necessarily injective) linear map

$$T \;:\; \ell_{2,1/\lambda,I} \to \mathcal{H},\; T(c) := \sum_{i\in I} c_i\varphi_i.$$

Within this space, our pre–Hilbert space $S = S_\Omega$ consists of all finite linear combinations

$$\begin{aligned} f_{a,X}(y) &= \sum_{j=1}^{N} a_j K(x_j, y) \\ &= \sum_{j=1}^{N} a_j \sum_{i\in I} \lambda_i \varphi_i(x_j)\varphi_i(y) \\ &= \sum_{i\in I} \varphi_i(y) \underbrace{\lambda_i \sum_{j=1}^{N} a_j \varphi_i(x_j)}_{=:c_i(a,X)} \end{aligned}$$

and the inner product is

$$\begin{aligned} (f_{a,X}, f_{b,Y})_K &= \sum_{j=1}^{M}\sum_{k=1}^{N} a_j b_k K(x_j, x_k) \\ &= \sum_{i\in I} \frac{1}{\lambda_i} \left(\lambda_i \sum_{j=1}^{M} a_j\varphi_i(x_j)\right)\left(\lambda_i \sum_{k=1}^{N} b_k\varphi_i(x_k)\right) \\ &= \sum_{i\in I} \frac{1}{\lambda_i} c_i(a,X)c_i(b,Y) \\ &= (c(a,X), c(b,Y))_{1/\lambda,I} \end{aligned}$$

if we define

$$c(a,X) := \{c_i(a,X)\}_{i\in I} = \left\{\lambda_i \sum_{j=1}^{N} a_j\varphi_i(x_j)\right\}_{i\in I} \in \ell_{2,1/\lambda,I}.$$

The native space $\mathcal{N}_K$ for K is the completion of S_Ω under this inner product, and it is a space of functions on Ω. By the above identity, it is clear that it is isometrically isomorphic to the Hilbert subspace of $\ell_{2,1/\lambda,I}$ obtained as the completion of the span of all $c(a,X)$.

We want to relate the native space to a function subspace of $\mathcal{H}$ now. To this end, we can form the closure of the span of all elements of $\Phi(\Omega)$ in $\ell_{2,\lambda,I}$ and denote it by $\mathcal{K}_{\lambda,\Omega}$. It clearly is a closed subspace of $\ell_{2,\lambda,I}$ and thus itself a Hilbert space. The elements $c \in \ell_{2,1/\lambda,I}$, when seen as functionals on $\ell_{2,\lambda,I}$, satisfy

$$(R(c), \Phi(x))_{\lambda,I} = c(\Phi(x)) = \sum_{i\in I} c_i\varphi_i(x) = T(c)(x)$$

for all $x \in \Omega$. A special case is

$$(R(c(a,X)), \Phi(x))_{\lambda,I} = f_{a,X}(x),$$

and by definition of $c(a,X)$ we also have $R(c(a,X)) \in \mathcal{K}_{\lambda,\Omega}$ for all $f_{a,X} \in S_\Omega$. This leads to

Theorem 2.32 *The native space for a Mercer kernel on Ω with weights $\lambda_i,\ i \in I$ and features $\varphi_i,\ i \in I$ is isometrically isomorphic to the space $\mathcal{H}_\lambda := T(R^{-1}(\mathcal{K}_{\lambda,\Omega})) \subset \mathcal{H}$ if $\mathcal{H}_\lambda$ is equipped with the inner product*

$$(T(c), T(d))_{\mathcal{H}_\lambda} := \sum_{i\in I} \frac{c_i d_i}{\lambda_i} = (c,d)_{1/\lambda,I}.$$

Proof: It is tempting to define the above bilinear form on all of $\mathcal{H}$, but the representation of functions in $\mathcal{H}$ in terms of coefficients of the φ_i is not unique, i.e. T is not necessarily injective. However, the representation is unique when restricted to the subspace $\mathcal{H}_\lambda \subset \mathcal{H}$. To see this, assume

$$T(c) = \sum_{i\in I} c_i\varphi_i = \sum_{j\in I} d_j\varphi_j = T(d)$$

as functions in $\mathcal{H}_\lambda \subset \mathcal{H}$, i.e. with $R(c),\ R(d) \in \mathcal{K}_{\lambda,\Omega}$. Then

$$(R(c) - R(d), \Phi(x))_{\lambda,I} \;=\; (T(c) - T(d))(x) = 0$$

proving $R(c) = R(d)$ as elements of $\mathcal{K}_{\lambda,\Omega}$ and finally $c = d$ due to bijectivity of the Riesz map. The rest follows from

$$(T(c(a,X)), T(c(b,Y)))_{\mathcal{H}_\lambda} = (c(a,X), c_b, Y))_{1/\lambda,I} = (f_{a,X}, f_{b,Y})_K.$$

□

2.10 Finite Case

We now specialize to the context of learning models on a **finite** set Ω consisting of $|\Omega|$ points and a **finite–dimensional** feature space. Instead of using point notation for Ω, we can identify Ω with the set $\Omega = \{1, \ldots, |\Omega|\}$ and use index notation instead, and we assume the feature space to be $\mathbb{R}^L$ for simplicity. Mercer kernels (2.6) then can be written as symmetric positive semidefinite matrices K with entries $K_{r,s},\ 1 \le r, s \le |\Omega|$ as

$$K = \Phi\Lambda\Phi^T$$

with an $L \times L$ diagonal matrix Λ containing positive weights $\lambda_1, \ldots, \lambda_L$ on its diagonal, while Φ is a $L \times |\Omega|$ matrix consisting of entries $\varphi_i(r),\ 1 \le i \le L,\ 1 \le r \le |\Omega|$.

The space $S = S_\Omega$ is then spanned by the $|\Omega|$ columns or rows of K, but we shall stick to column notation when we consider a function on Ω. Each function f in S thus is a linear combination of columns of K, and thus it has the form $f_{a,\Omega} := Ka$ with a vector $a \in I\!R^{|\Omega|}$. The inner product then is

$$(f_{a,\Omega}, f_{b,\Omega})_K = a^T K b = a^T \Phi \Lambda \Phi^T b \text{ for all } a, b \in I\!R^{|\Omega|}.$$

Our complicated proof of section 2.4 for the well–definedness of the inner product now takes a simpler form, since for $Ka = K\tilde{a}$ and $Kb = K\tilde{b}$ we get

$$\begin{aligned}(f_{a,\Omega}, f_{b,\Omega})_K &= a^T K b \\ &= \tilde{a}^T K b \\ &= \tilde{a}^T K \tilde{b} \\ &= (f_{\tilde{a},\Omega}, f_{\tilde{b},\Omega})_K,\end{aligned}$$

but readers will see that the basic argument is the same. Also, the positive definiteness of the inner product is simple to see, because from $\|f_{a,\Omega}\|_K^2 = a^T K a = 0$ we first get $\Phi^T a = 0$ from

$$0 = a^T K a = a^T \Phi \Lambda \Phi^T a = a^T \Phi \sqrt{\Lambda}\sqrt{\Lambda} \Phi^T a = \|\sqrt{\Lambda}\Phi^T a\|_2^2$$

with the nonsingular diagonal matrix $\sqrt{\Lambda}$ defined in an obvious way. But $\Phi^T a = 0$ implies $f_{a,\Omega} = Ka = \Phi\Lambda\Phi^T a = 0$.

In practical cases, the matrices Φ and K are much too large to be handled, but there are efficient methods for the reduction of dimensions via **principal component analysis** or **singular value decomposition**. We describe the basic principle now, but remark that practical applications will proceed differently.

A singular value decomposition splits K into a product

$$K = \Phi\Lambda\Phi^T = U\Sigma U^T$$

with an orthogonal $|\Omega| \times |\Omega|$ matrix U and a diagonal $|\Omega| \times |\Omega|$ matrix Σ of **singular values** of K, i.e. the nonnegative eigenvalues of $K^T K$. Note that this amounts to consider an equivalent setting with now $L = |\Omega|$, $U = \Phi$, and $\Lambda = \Sigma$, but now the diagonal of Σ may contain zero entries. The orthogonal matrix U just is a coordinate change in the native space, and thus does not matter in theory. The problem thus behaves exactly as in the case $K = \Sigma$, i.e. as if the kernel was diagonal. We shall come back to the finite situation later.

2.11 Kernels for Univariate Sobolev Spaces

Let us calculate the kernel for Sobolev space $H_2^k[a,b]$ for an interval $[a,b] \subset I\!R$. It has the inner product

$$(f,g)_k := \sum_{j=0}^{k} \int_a^b f^{(j)}(t) g^{(j)}(t) dt = \sum_{j=0}^{k} \left(f^{(j)}, g^{(j)}\right)_{L_2[a,b]}$$

and consists of all functions whose k–th derivative is in $L_2[a,b]$. For $k \geq 1$ these functions are continuous and have continuous point evaluation. To prove continuity, take $a \leq x \leq y \leq b$ to get

$$\begin{aligned} |f(x) - f(y)| &\leq \int_x^y |f'(t)|dt \\ &\leq \sqrt{\int_x^y 1dt}\sqrt{\int_x^y |f'(t)|^2 dt} \\ &\leq \sqrt{y-x}\sqrt{\int_a^b |f'(t)|^2 dt} \\ &\leq \sqrt{y-x}\|f\|_1. \end{aligned}$$

Continuous point evaluation follows everywhere, if we have it at a and use the above argument. To prove it at a, just verify

$$f(a) = \frac{1}{b-a}\int_a^b f(t)dt + \frac{1}{b-a}\int_a^b f'(t)(t-b)dt$$

via integration by parts and bound it like we did above.

The kernel K_k of the space for $k \geq 1$ must exist and should satisfy

$$f(x) = \sum_{j=0}^{k} \int_a^b f^{(j)}(t) \frac{\partial^j K_k(x,t)}{\partial t^j} dt$$

for all $x \in [a,b]$ and $f \in H_2^k[a,b]$.

We only look at the case $k = 1$ and have to care for

$$f(x) = \int_a^b f(t)K_1(x,t)dt + \int_a^b f'(t)K_1'(x,t)dt$$

where from now on we keep x fixed and use only derivatives with respect to t. We want to use integration by parts on the second integral in order to generate the two other terms. But then we have to assume that $K_1(x,t)$ has a derivative discontinuity at $t = x$, and we split the integral there. This yields

$$\begin{aligned} \textstyle\int_a^b f'(t)K_1'(x,t)dt &= [f(t)K'(x,t)]_a^x - \textstyle\int_a^x f(t)K_1''(x,t)dt \\ &\quad + [f(t)K'(x,t)]_x^b - \textstyle\int_x^b f(t)K_1''(x,t)dt \end{aligned}$$

and we manage to get the reproduction formula if we can satisfy

$$\begin{aligned} K_1''(x,t) &= K_1(x,t) && \text{for all } x,t \\ K_1'(x,a) &= 0 && \text{for all } x \\ K_1'(x,b) &= 0 && \text{for all } x \\ \frac{\partial K_1(x,t)}{\partial t}_{|t=x_-} &= \alpha(x) && \text{for all } x \\ \frac{\partial K_1(x,t)}{\partial t}_{|t=x_+} &= \beta(x) && \text{for all } x \\ \alpha(x) - \beta(x) &= 1 && \text{for all } x. \end{aligned}$$

The differential equation $K_1''(x,t) = K_1(x,t)$ has the general solution

$$K_1(x,t) = c_+(x)e^{+t} + c_-(x)e^{-t}$$

with certain coefficient functions we still can choose, but we can choose two sets, one each for $x \geq t$ and one for $x \leq t$.

We first look at $a \leq t \leq x$ with $x > a$ and take the derivative

$$K_1'(x,t) = c_+(x)e^{+t} - c_-(x)e^{-t}.$$

Then we have to satisfy

$$\begin{array}{rcccl} K_1'(x,a) & = & c_+(x)e^{a} - c_-(x)e^{-a} & = & 0 \\ K_1'(x,x) & = & c_+(x)e^{x} - c_-(x)e^{-x} & = & \alpha(x) \end{array}$$

leading to

$$\begin{array}{rcl} c_+(x) & = & \alpha(x)\dfrac{e^{-a}}{e^{x-a} - e^{-(x-a)}} \\ c_-(x) & = & \alpha(x)\dfrac{e^{a}}{e^{x-a} - e^{-(x-a)}} \end{array}$$

and

$$K(x,t) = \alpha(x)\frac{e^{t-a} + e^{-(t-a)}}{e^{x-a} - e^{-(x-a)}} = \alpha(x)\frac{\cosh(t-a)}{\sinh(x-a)} \text{ for all } a \leq t \leq x > a.$$

Similarly, for $a \leq x \leq t \leq b,\ x < b$ we have to satisfy the equations

$$\begin{array}{rcccl} K_1'(x,b) & = & c_+(x)e^{b} - c_-(x)e^{-b} & = & 0 \\ K_1'(x,x) & = & c_+(x)e^{x} - c_-(x)e^{-x} & = & \beta(x). \end{array}$$

We can write down the solution by replacing a by b and α by β in what we had before, resulting in

$$K(x,t) = \beta(x)\frac{e^{t-b} + e^{-(t-b)}}{e^{x-b} - e^{-(x-b)}} = \beta(x)\frac{\cosh(t-b)}{\sinh(x-b).}$$

We still need to satisfy

$$\begin{array}{rcl} \alpha(x) - \beta(x) & = & 1 \\ K(x,x_-) & = & K(x,x_+) \end{array}$$

which gives us two equations for two unknowns in the form

$$\begin{array}{rcccl} \alpha(x) & - & \beta(x) & = & 1 \\ \alpha(x)\dfrac{\cosh(x-a)}{\sinh(x-a)} & - & \beta(x)\dfrac{\cosh(x-b)}{\sinh(x-b)} & = & 0 \end{array}$$

which result in

$$\begin{array}{rcl} \alpha(x) & = & \dfrac{\sinh(x-a)\cosh(x-b)}{\sinh(x-a)\cosh(x-b) - \sinh(x-b)\cosh(x-a)} = \dfrac{\sinh(x-a)\cosh(x-b)}{\sinh(b-a)} \\ \beta(x) & = & \dfrac{\sinh(x-b)\cosh(x-a)}{\sinh(x-a)\cosh(x-b) - \sinh(x-b)\cosh(x-a)} = \dfrac{\sinh(x-b)\cosh(x-a)}{\sinh(b-a)} \end{array}$$

and finally

$$\begin{array}{rcll} K(x,t) & = & \dfrac{\cosh(x-b)\cosh(t-a)}{\sinh(b-a)} & a \leq t \leq x \leq b \\ K(x,t) & = & \dfrac{\cosh(x-a)\cosh(t-b)}{\sinh(b-a)} & a \leq x \leq t \leq b. \end{array}$$

We leave it to the reader to go all the way backwards to prove the reproduction equation.

But we can also look at the case of Sobolev space on $(-\infty, \infty)$. The kernel, as a function of t, then must have the form

$$\begin{aligned} K(x,t) &= c_+(x)e^t \quad -\infty < t \le x < \infty \\ K(x,t) &= c_-(x)e^{-t} \quad -\infty < x \le t < \infty \end{aligned}$$

because otherwise it would not be integrable. To get continuity at $t = x$ we need

$$c_+(x)e^x = c_-(x)e^{-x}$$

which is satisfied if we write

$$\begin{aligned} c_+(x) &= c(x)e^{-x} \\ c_-(x) &= c(x)e^{+x} \end{aligned}$$

with some function $c(x)$. We have

$$\begin{aligned} K'(x,t) &= c_+(x)e^t &&= c(x)e^{t-x} \quad -\infty < t \le x < \infty \\ K'(x,t) &= -c_-(x)e^{-t} &&= -c(x)e^{x-t} \quad -\infty < x \le t < \infty \end{aligned}$$

and we can satisfy $K'(x, x_-) - K'(x, x_+) = 1$ by simply setting $c(x) = \frac{1}{2}$. Thus

$$\begin{aligned} K(x,t) &= \tfrac{1}{2}e^{t-x} \quad -\infty < t \le x < \infty \\ K(x,t) &= \tfrac{1}{2}e^{x-t} \quad -\infty < x \le t < \infty \end{aligned}$$

or

$$K(x,t) = \frac{1}{2}e^{-|x-t|} \text{ for all } x, t \in I\!R,$$

see figure 4. The local kernel on $[-1, 1]$ is in Figure 5.

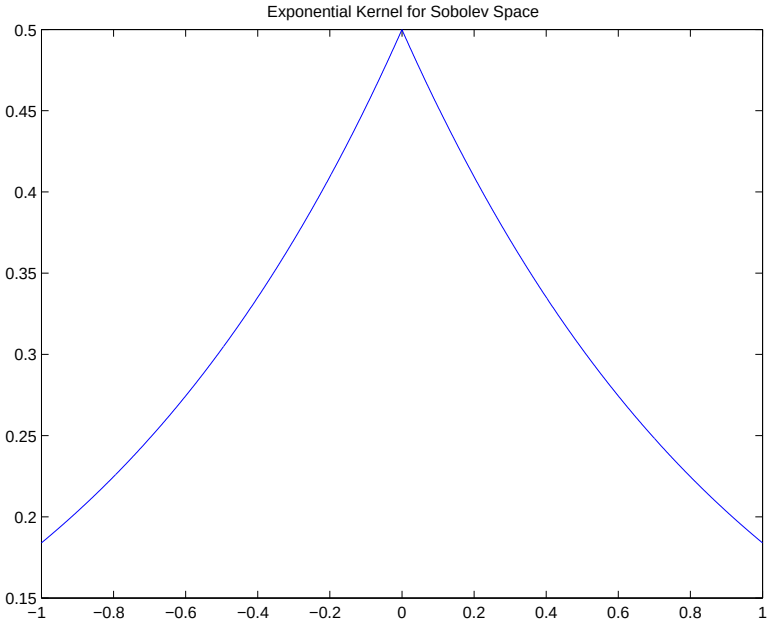

Figure 4: The kernel $\exp(-|x - y|)$

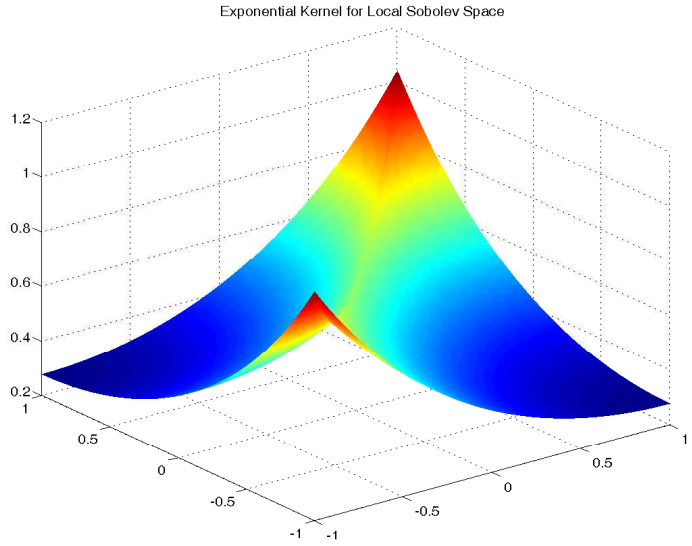

Figure 5: The local kernel for $W_2^1[-1,1]$

Just for fun let us check if this occurs as the limit of the previous kernel for $a \to -\infty$ and $b \to \infty$. We just check the case

$$K(x,t) = \frac{\cosh(x-b)\cosh(t-a)}{\sinh(b-a)} \text{ for all } a \le t \le x \le b$$

and leave the other case to the reader. We get

$$\begin{aligned} K(x,t) &= \frac{\cosh(x-b)\cosh(t-a)}{\sinh(b-a)} \\ &= \frac{1}{2}\frac{(e^{x-b}+e^{b-x})(e^{t-a}+e^{a-t})}{e^{b-a}-e^{a-b}} \\ &= \frac{1}{2}\frac{e^{b-a}}{e^{b-a}}\frac{(e^{x-2b}+e^{-x})(e^{t}+e^{2a-t})}{1-e^{2a-2b}} \\ &\to \frac{1}{2}e^{-x}e^{t} = \frac{1}{2}e^{t-x} \end{aligned}$$

as required.

The upshot of all of this is to show that certain well–known and useful spaces have useful reproducing kernels. But at other places we shall also show the opposite, i.e. that certain useful kernels have useful native spaces.

3 Optimal Recovery

This section concerns the reconstruction of functions from native spaces from pointwise data on subsets of Ω.

3.1 Optimality Properties

We now fix a finite set $X = \{x_1, \ldots, x_N\} \subset \Omega$ and consider the finite–dimensional space

$$S_X := \text{ span } \{K(x,\cdot) \ : \ x \in X\}.$$

Such spaces are useful in various contexts, because they provide **trial functions** for approximation and interpolation of functions, or the solution of differential equations. We first check the best approximations from S_X to functions in the Native Space.

Theorem 3.1 *Given an arbitrary function f from the native space $\mathcal{N}_K$ of a positive semidefinite kernel K on a set Ω, fix a finite set $X = \{x_1, \ldots, x_N\} \subset \Omega$. Then the solution $f_{a^*,X}$ of the finite–dimensional approximation problem*

$$\min_{s \in S_X} \|f - s\|_K = \min_{a \in \mathbb{R}^N} \|f - f_{a,X}\|_K$$

exists and is an interpolant to f on X. The coefficients solve the linear system

$$f_{a^*,X}(x_k) = \sum_{j=1}^{N} a_j^* K(x_j, x_k) = f(x_k),\ 1 \le k \le N. \tag{3.2}$$

Proof: From standard arguments of linear approximation in spaces with inner products it follows that the solution exists and has the orthogonality property

$$(f - f_{a^*,X}, s)_K = 0 \text{ for all } s \in S_X$$

which includes

$$0 = (f - f_{a^*,X}, K(x_k, \cdot))_K = f(x_k) - f_{a^*,X}(x_k),\ 1 \le k \le N$$

due to the reproduction property (2.20). But this implies the rest of the proof. □

Note that this guarantees that the linear system (3.2) with the kernel matrix $A_{K,X,X}$ is always solvable, though the matrix may be singular. The matrix is only positive semidefinite, but the right–hand side of (3.2), being a set of values on X of a function from the native space, always lies in the span of the columns. The coefficients a_j^* and the resulting function $f_{a^*,X}$ may not be unique unless the kernel is positive definite.

Theorem 3.3 *Given an arbitrary function f from the native space $\mathcal{N}_K$ of a positive semidefinite kernel K on a set Ω and a finite set $X = \{x_1, \ldots, x_N\} \subset \Omega$. Under all interpolants from $\mathcal{N}_K$ to f on X, the interpolating functions $f_{a^*,X}$ from the previous theorem minimize the native space norm, i.e.*

$$\min_{g \in \mathcal{N}_K,\ g = f \text{ on } X} \|g\|_K = \|f_{a^*,X}\|_K.$$

Proof: We have to show

$$\|g\|_K^2 \ge \|f_{a^*,X}\|_K^2$$

for all $g \in \mathcal{N}_K$ with $g = f$ on X. But we can write

$$\begin{aligned} \|g\|_K^2 &= \|f_{a^*,X} + (g - f_{a^*,X})\|_K^2 \\ &= \|f_{a^*,X}\|_K^2 + 2(f_{a^*,X}, g - f_{a^*,X})_K + \|g - f_{a^*,X}\|_K^2 \end{aligned}$$

and get the assertion from

$$\begin{aligned}(f_{a^*,X}, g - f_{a^*,X})_K &= \left(\sum_{j=1}^N a_j^* K(x_j,\cdot), g - f_{a^*,X}\right)_K \\ &= \sum_{j=1}^N a_j^* \left(K(x_j,\cdot), g - f_{a^*,X}\right)_K \\ &= \sum_{j=1}^N a_j^* \left(g(x_j) - f_{a^*,X}(x_j)\right) \\ &= \sum_{j=1}^N a_j^* \left(g(x_j) - f(x_j)\right) = 0\end{aligned}$$

proving the theorem. □

3.2 Lagrange Reformulation

We now fix K, $X = \{x_1, \ldots, x_N\} \subset \Omega$ and a point $x \in \Omega$. Then we define the function $f_x(\cdot) := K(x,\cdot) \in \mathcal{N}_K$ and carry the previous construction out for this function. Consequently, we get a solvable system of equations

$$\sum_{j=1}^N u_j^*(x) K(x_j, x_k) = K(x_k, x),\ 1 \le k \le N,\ x \in \Omega \tag{3.4}$$

for each $x \in \Omega$, and for reasons to become clear soon, we now denote the x–dependent solution coefficients by $u_j^*(x)$. In fact, this is the standard notation for a **cardinal** or **Lagrange** basis, and in case of nonsingularity of the kernel matrix we immediately see that the characteristic equations

$$u_j^*(x_k) = \delta_{j,k},\ 1 \le j,k \le N$$

for such a basis are satisfied.

If we only have positive semidefiniteness, we can still rewrite any solution $f_{a^*,X}$ of the original problem of interpolation or approximation of f on X in the form

$$\begin{aligned}f_{a^*,X}(x) &= \sum_{k=1}^N a_k^* K(x_k, x) \\ &= \sum_{k=1}^N a_k^* \sum_{j=1}^N u_j^*(x) K(x_j, x_k) \\ &= \sum_{j=1}^N u_j^*(x) \sum_{k=1}^N a_k^* K(x_j, x_k) \\ &= \sum_{j=1}^N u_j^*(x) f_{a^*,X}(x_j) \\ &= \sum_{j=1}^N u_j^*(x) f(x_j)\end{aligned} \tag{3.5}$$

which generalizes the usual Lagrange formulation of interpolation. Note that there still is nonuniqueness, but in our new form we have found a representation which separates the

influence of X and f. In particular, we can take an arbitrary $f = f_{b,X} \in S_X$ for an arbitrary $b \in I\!R^N$ and conclude

$$f_{b,X}(x) = \sum_{j=1}^{N} u_j^*(x) f_{b,X}(x_j).$$

Theorem 3.6 *The linear* **quasi–interpolation** *operator*

$$Q_X(f) := \sum_{j=1}^{N} u_j^*(\cdot) f(x_j)$$

on the native space $\mathcal{N}_K$ reproduces all functions from S_X using only their values on X. The pointwise error has the representation

$$f(x) - Q_X(f)(x) = (f, K_X(x, \cdot))_K \text{ for all } x \in \Omega$$

with the X–dependent **power kernel**

$$K_X(x, y) := K(x, y) - \sum_{j=1}^{N} u_j^*(x) K(x_j, y) \text{ for all } x, y \in \Omega$$

and the error bound

$$|f(x) - Q_X(f)(x)| \le \|f\|_K P_X(x) \text{ for all } x \in \Omega \tag{3.7}$$

with the **power function**

$$\begin{aligned} P_X^2(x) &:= \|K_X(x, \cdot)\|_K^2 \\ &= K(x, x) - 2 \sum_{j=1}^{N} u_j^*(x) K(x_j, x) + \sum_{j,k=1}^{N} u_j^*(x) u_k^*(x) K(x_j, x_k). \end{aligned} \tag{3.8}$$

Proof*: Clearly, the definition of the quasi–interpolation operator and the reproduction equation imply*

$$\begin{aligned} f(x) - Q_X(f)(x) &= (f, K(x, \cdot))_K - \sum_{j=1}^{N} u_j^*(x) (f, K(x_j, \cdot))_K \\ &= (f, K_X(x, \cdot))_K \text{ for all } x \in \Omega. \end{aligned}$$

□

Unfortunately, we cannot directly conclude that the functions u_j^* are in the native space or in S_X, because we have no invertibility of the kernel matrix.

To prove that we can choose the $u_j^*(x)$ in such a way that they lie in S_X, we start a detour and ask for the coefficients $a_1, \ldots, a_N$ which minimize

$$\sup_{f \in \mathcal{N}_K,\, \|f\|_K \le 1} |f(x) - \sum_{j=1}^{N} a_j f(x_j)|.$$

Clearly these coefficients, if they exist at all, will be dependent on x. We introduce the standard point–evaluation functional

$$\delta_x \,:\, \mathcal{N}_K \to I\!R,\ \delta_x(f) = f(x) \text{ for all } f \in \mathcal{N}_K$$

and consider its norm

$$\|\delta_x\|_K := \sup_{f\in\mathcal{N}_K,\ \|f\|_K\le 1} |f(x)| = \sup_{f\in\mathcal{N}_K,\ \|f\|_K\le 1} |(f, K(x,\cdot))_K| \le \|K(x,\cdot)\|_K = \sqrt{K(x,x)}$$

where we also used the index K to indicate the norm in the dual of the native space. But if we insert $f := K(x,\cdot)/\sqrt{K(x,x)}$, we get equality above. Our minimization problem then turns into

$$\min_{a\in\mathbb{R}^N} \|\delta_x - \sum_{j=1}^N a_j\delta_{x_j}\|_K^2 = \min_{a\in\mathbb{R}^N} \left(K(x,x) - 2\sum_{j=1}^N a_j K(x,x_j) + \sum_{j,k=1}^N a_j a_k K(x_j,x_k)\right). \tag{3.9}$$

By standard least–squares arguments, or by simple differentiation we see that the necessary equations in a minimum are

$$\sum_{j=1}^N a_j K(x_k,x_j) = K(x,x_k),\ 1\le k\le N$$

and they are satisfied if we take $a_j = u_j^*(x)$ from (3.4). Thus we know that the minimum problem is solved by what we had before, and get

Theorem 3.10 *Let K be a positive semidefinite kernel on Ω, and let $X = \{x_1,\dots,x_N\} \subset \Omega$ be a fixed set. For all points $x\in\Omega$ the coefficient functions $u_j^*(x)$ of (3.4) satisfy*

$$\inf_{a\in\mathbb{R}^N} \sup_{f\in\mathcal{N}_K,\ \|f\|_K\le 1} |f(x) - \sum_{j=1}^N a_j f(x_j)| = \sup_{f\in\mathcal{N}_K,\ \|f\|_K\le 1} |f(x) - \sum_{j=1}^N u_j^*(x) f(x_j)|,$$

i.e. they realize the optimal pointwise error for all interpolants. Furthermore, there is an error bound

$$|f(x) - \sum_{j=1}^N u_j^*(x) f(x_j)| \le \|f\|_K P_{X,K}(x) \text{ for all } x\in\Omega,\ f\in\mathcal{N}_K \tag{3.11}$$

with the **power function** *defined as*

$$\begin{aligned} P_{X,K}^2(x) &:= \|\delta_x - \sum_{j=1}^N u_j^*(x)\delta_{x_j}\|_K^2 \\ &= K(x,x) - 2\sum_{j=1}^N u_j^*(x)K(x_j,x) + \sum_{j,k=1}^N u_j^*(x)u_k^*(x)K(x_j,x_k) \\ &= K(x,x) - \sum_{j=1}^N u_j^*(x)K(x_j,x) \text{ for all } x\in\Omega \end{aligned} \tag{3.12}$$

and with values which are independent of the choice of the $u_j^*(x)$ in case of nonuniqueness.

Proof: If the necessary equations for the minimum of a positive semidefinite finite–dimensional quadratic form are satisfied somewhere, they are sufficient for a minimum. The location of the minimum may be nonunique, but the value is not. This proves the first assertion and the final part of the last. The middle part follows from the definition of the power function, and the simplification between the third and last line follows from (3.4). □

This result looks very theoretic, but it is of great practical importance, because the power function for fixed X can be calculated explicitly everywhere in Ω, and the error bound (3.11) allows to estimate the error of all possible interpolants based on the data locations in X.

Corollary 3.13 *In addition, the power function $P_{X,K}(x)$ always vanishes at the points of X.*

Proof: Clearly, for $x = x_k$ the coefficients $a_j := \delta_{jk}$ are admissible and lead to the value zero. Since the minimum value is unique (while the location of the minimum is not), we have the assertion in spite of the nonuniqueness of the solution. □

3.3 Calculation

We now want to take a closer look at the systems (3.2) or (3.4). To this end, we perform a singular–value–decomposition of the kernel matrix as

$$A = U\Sigma U^T$$

with an orthogonal matrix U and a diagonal matrix with nonnegative entries $\sigma_1, \ldots, \sigma_N$. We focus on (3.9) as minimization of a quadratic form. The latter is

$$\begin{aligned} 0 \le Q(a) &= K(x,x) - 2\sum_{j=1}^{N} a_j K(x.x_j) + \sum_{j,k=1}^{N} a_j a_k K(x_j, x_k) \\ &= K(x,x) - 2a^T K_X(x) + a^T A a \text{ with} \\ K_X(x) &:= (K(x_1,x), \ldots, K(x_N,x))^T \end{aligned}$$

and can be rewritten as

$$\begin{aligned} Q(a) &= K(x,x) - 2a^T UU^T K_X(x) + a^T UU^T AUU^T a \\ &= K(x,x) - 2a^T U \underbrace{U^T K_X(x)}_{=:z(x)} + a^T U\Sigma \underbrace{U^T a}_{=:b} \\ = R(b) &:= K(x,x) - 2b^T z(x) + b^T \Sigma b \\ &= K(x,x) + \sum_{j=1}^{N} \Big(b_j^2 \sigma_j - 2 b_j z_j(x) \Big). \end{aligned}$$

We know that this quadratic form is always nonnegative, and we can minimize it now by taking derivatives with respect to each b_j. The optimal values $b_j^*(x)$ have to satisfy

$$b_j^*(x)\sigma_j = z_j(x),\ 1 \le j \le N.$$

This leads to

$$b_j^*(x) := \frac{z_j(x)}{\sigma_j} \text{ for } \sigma_j > 0.$$

In case of $\sigma_j = 0$ we must (in theory) have $z_j(x) = 0$ because otherwise the quadratic form could take on negative values. For these j we can take any $b_j^*(x)$, and we formally write

$$b_j^*(x) := \begin{cases} \dfrac{z_j(x)}{\sigma_j} & \sigma_j > 0 \\ \lambda_j z_j(x) & \sigma_j = 0 \end{cases}$$

with arbitrary λ_j for the j with $\sigma_j = 0$. Thus we can write

$$b^*(x) = Dz(x)$$

with a diagonal matrix $D = D(\sigma, \lambda)$ having the entries

$$\begin{array}{ll} \frac{1}{\sigma_j} & \text{for} \quad \sigma_j > 0 \\ \lambda_j & \text{for} \quad \sigma_j = 0 \end{array}$$

on the diagonal. This yields the representation

$$a^*(x) = Ub^*(x) = UDz(x) = UDU^T K_X(x)$$

of the total solution, but we already know that this solution also arises as $u_j^*(x) = a_j^*(x)$ in the system (3.4) and the Lagrange type formula (3.5). But in the above form we see that the solution can in spite of the singular system be written in such a way that it lies in S_X and thus in the native space.

In practical situations, the right–hand side of a system (3.2) will not necessarily consist of values of a function from the native space. In such a case the system might be unsolvable, and this then proves that the data indeed do not come from a function in the native space. But one can always go for a quasi–interpolant of the form (3.5) with the $u_j^*(x)$ constructed as above. This will not necessarily interpolate the data, but probably be a good reconstruction strategy anyway.

3.4 Regularization

Let A be an $m \times n$ matrix and consider the linear system

$$Ax = b \in I\!R^m \tag{3.14}$$

which is to be solved for a vector $x \in I\!R^n$. The system may arise from any method using kernels, including (3.4) and (3.2), but we allow for more equations than unknowns here. Then the system will have $m \geq n$ and it usually is overdetermined. Furthermore, for later cases, we allow the matrix A to be unsymmetric.

The previous section told us that even in the case $m = n$ with a positive semidefinite matrix, chances are good there is an approximate solution $\hat{x}$ which at least yields $\|A\hat{x} - b\|_2 \leq \eta$ with a small tolerance η, and which has a coefficient vector $\hat{x}$ representable on a standard computer. Note that η may also contain noise of a certain unknown level. The central problem is that there are many vectors $\hat{x}$ leading to small values of $\|A\hat{x} - b\|_2$, and the selection of just one of them is an unstable process. But the reproduction quality is much more important than the actual accuracy of the solution vector $\hat{x}$, and thus questions like the nonsingularity or the condition of the matrix are not the right aspects here.

Clearly, any reasonably well-programmed least-squares solver [26] should do the job, i.e. produce a numerical solution $\tilde{x}$ which solves

$$\min_{x \in I\!R^n} \|Ax - b\|_2 \tag{3.15}$$

or at least guarantees $\|A\tilde{x} - b\|_2 \leq \eta$. It should at least be able not to overlook or discard $\hat{x}$. This **regularization** by **optimization** works in many practical cases, but we shall take a closer look at the joint error and stability analysis, because even an optimizing algorithm will

recognize that it has problems to determine $\hat{x}$ reliably if columns of the matrix A are close to being linearly dependent.

By **singular-value decomposition** [26], the matrix A can be decomposed into

$$A = U\Sigma V^T \tag{3.16}$$

where U is an $m \times m$ orthogonal matrix, Σ is an $m \times n$ matrix with zeros except for **singular values** $\sigma_1, \ldots, \sigma_n$ on the diagonal, and where V^T is an $n \times n$ orthogonal matrix. Due to some sophisticated numerical tricks, this decomposition can under normal circumstances be done with $\mathcal{O}(mn^2 + nm^2)$ complexity, though it needs an eigenvalue calculation. One can assume

$$\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_n^2 \geq 0,$$

and the σ_j^2 are the nonnegative eigenvalues of the positive semidefinite $n \times n$ matrix $A^T A$.

The condition number of the non-square matrix A is then usually defined to be σ_1/σ_n. This is in line with the usual **spectral condition number** $\|A\|_2\|A^{-1}\|_2$ for the symmetric case $m = n$. The numerical computation of U and V usually is rather stable, even if the total condition is extremely large, but the calculation of small singular values is hazardous. Thus the following arguments can rely on U and V, but not on small singular values.

Using (3.16), the solution of either the minimization problem (3.15) or, in the case $m = n$, the solution of (3.14) can be obtained and analyzed as follows. We first introduce new vectors

$$c := U^T b \in I\!R^m \text{ and } y := V^T x \in I\!R^n$$

by transforming the data and the unknowns orthogonally. Since orthogonal matrices preserve Euclidean lengths, we rewrite the squared norm as

$$\begin{aligned} \|Ax - b\|_2^2 &= \|U\Sigma V^T x - b\|_2^2 \\ &= \|\Sigma V^T x - U^T b\|_2^2 \\ &= \|\Sigma y - c\|_2^2 \\ &= \sum_{j=1}^{n} (\sigma_j y_j - c_j)^2 + \sum_{j=n+1}^{m} c_j^2 \end{aligned}$$

where now $y_1, \ldots, y_n$ are variables. Clearly, the minimum exists and is given by the equations

$$\sigma_j y_j = c_j,\ 1 \leq j \leq n,$$

but the numerical calculation runs into problems when the σ_j are small and imprecise in absolute value, because then the resulting y_j will be large and imprecise. The final transition to the solution $x = Vy$ by an orthogonal transformation does not improve the situation.

If we assume existence of a good solution candidate $\hat{x} = V\hat{y}$ with $\|A\hat{x} - b\|_2 \leq \eta$, we have

$$\sum_{j=1}^{n} (\sigma_j \hat{y}_j - c_j)^2 + \sum_{j=n+1}^{m} c_j^2 \leq \eta^2. \tag{3.17}$$

A standard **regularization** strategy to construct a reasonably stable approximation y is to choose a positive tolerance ϵ and to define

$$y_j^\epsilon := \begin{cases} \frac{c_j}{\sigma_j} & |\sigma_j| \geq \epsilon \\ 0 & |\sigma_j| < \epsilon \end{cases}$$

i.e. to ignore small singular values, because they are usually polluted by roundoff and hardly discernible from zero. This is called the **truncated singular value decomposition** (TSVD). Fortunately, one often has small c_j^2 whenever σ_j^2 is small, and then chances are good that

$$\|Ax^\epsilon - b\|_2^2 = \sum_{\substack{1 \le j \le n \\ |\sigma_j| \ge \epsilon}} c_j^2 + \sum_{j=n+1}^{m} c_j^2 \le \eta^2$$

holds for $x^\epsilon = Vy^\epsilon$.

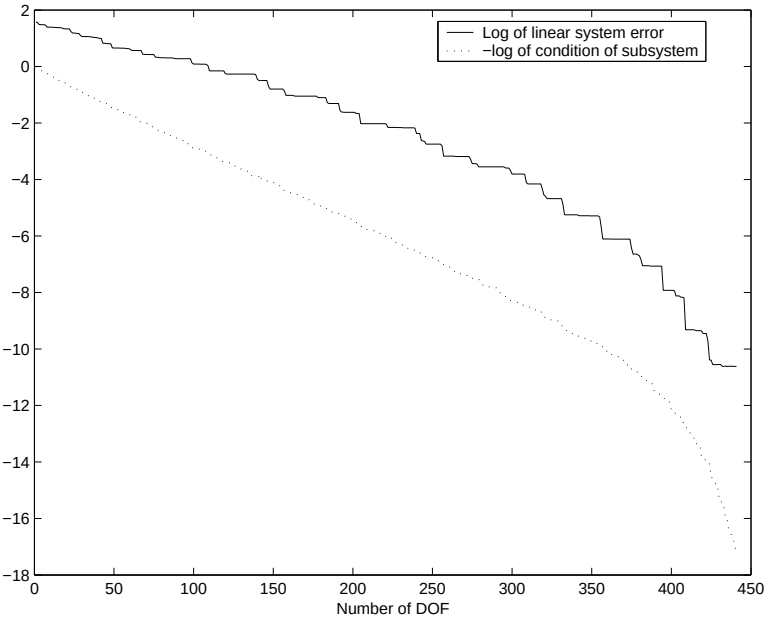

Figure 6: Error and condition of linear subsystems via SVD

Figure 6 is an example interpolating the MATLAB `peaks` function in $m = n = 441$ regular points on $[-3, 3]^2$ by Gaussians with scale 1, using the standard system (3.2). Following a fixed 441×441 singular value decomposition, we truncated after the k largest singular values, thus using only k degrees of freedom (DOF). The results for $1 \le k \le 441$ show that there are low-rank subsystems which already provide good approximate solutions.

But now we proceed with our analysis. In case of large c_j for small σ_j, truncation is insufficient, in particular if the dependence on the unknown noise level η comes into focus. At least, the numerical solution should not spoil the reproduction quality guaranteed by (3.17), which is much more important than an exact calculation of the solution coefficients. Thus one can minimize $\|y\|_2^2$ subject to the essential constraint

$$\sum_{j=1}^{n} (\sigma_j y_j - c_j)^2 + \sum_{j=n+1}^{m} c_j^2 \le \eta^2, \tag{3.18}$$

but we suppress details of the analysis of this optimization problem. Another, more popular possibility is to minimize the objective function

$$\sum_{j=1}^{n} (\sigma_j y_j - c_j)^2 + \delta^2 \sum_{j=1}^{n} y_j^2$$

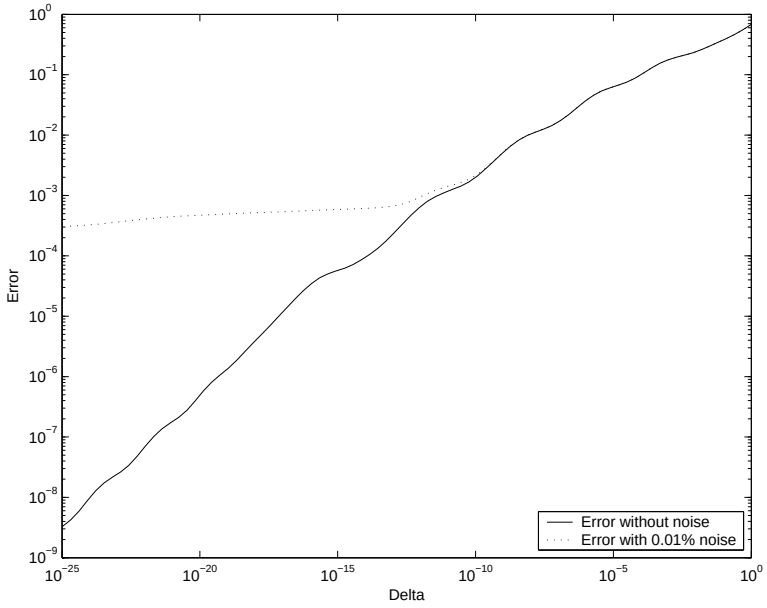

Figure 7: Error as function of regularization parameter δ^2

where the positive weight δ allows to put more emphasis on small coefficients if δ is increased. This is called **Tikhonov regularization**.

The solutions of both settings coincide and take the form

$$y_j^\delta := \frac{c_j \sigma_j}{\sigma_j^2 + \delta^2},\ 1 \le j \le n$$

depending on the positive parameter δ of the Tikhonov form, and for $x^\delta := V y^\delta$ we get

$$\|Ax^\delta - b\|_2^2 = \sum_{j=1}^{n} c_j^2 \left(\frac{\delta^2}{\delta^2 + \sigma_j^2} \right)^2 + \sum_{j=n+1}^{m} c_j^2,$$

which can me made smaller than η^2 for sufficiently small δ. The optimal value δ^* of δ for a known noise level η in the sense of (3.18) would be defined by the equation $\|Ax^{\delta^*} - b\|_2^2 = \eta^2$, but since the noise level is only rarely known, users will be satisfied to achieve a tradeoff between reproduction quality and stability of the solution by inspecting $\|Ax^\delta - b\|_2^2$ for varying δ experimentally.

We now repeat the example leading to Figure 6, replacing the truncation strategy by the above regularization. Figure 7 shows how the error $\|Ax^\delta - b\|_{\infty,X}$ depends on the regularization parameter δ. In case of noise, users can experimentally determine a good value for δ even for an unknown noise level. The condition of the full matrix was calculated by MATLAB as $1.46 \cdot 10^{19}$, but it may actually be higher. Figure 8 shows that the coefficients $|c_j|$ are indeed rather small for large j, and thus regularization by truncated SVD will work as well in this case.

From Figures 8 and 7 one can see that the error $\|Ax^\delta - b\|$ takes a sharp turn at the noise level. This has led to the ***L*-curve method** for determining the optimal value of δ, but the L-curve is defined differently as the curve

$$\delta \mapsto (\log \|y^\delta\|_2^2, \log \|Ax^\delta - b\|_2^2).$$

The optimal choice of δ is made where the curve takes its turn, if it does so, and there are various way to estimate the optimal δ, see [27, 28, 29] including a MATLAB software package.

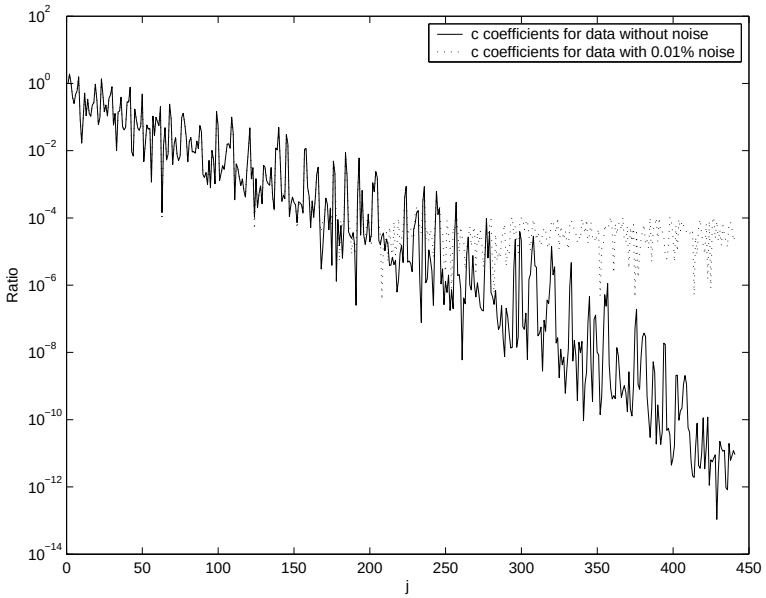

Figure 8: Coefficients $|c_j|$ as function of j

Figure 9 shows the typical L-shape of the L-curve in case of noise, while in the case of exact data there is no visible sharp turn within the plot range. The background problem is the same as for the previous figures.

Consequently, users of kernel techniques are strongly advised to take some care when choosing a linear system solver. The solution routine should incorporate a good regularization strategy or at least automatically project to stable subspaces and not give up quickly due to bad condition. Further examples for this will follow in later chapters.

But for large systems, the above regularization strategies are debatable. A singular-value decomposition of a large system is computationally expensive, and the solution vector will usually not be sparse, i.e. the evaluation of the final solution at many points is costly. In many cases, linear systems arising from kernels often have good approximate solutions with only few nonzero coefficients, and the corresponding numerical techniques are other, and possibly preferable regularizations which still are under investigation.

4 Conditionally Positive Definite Kernels

So far, we looked at positive semidefinite symmetric kernels. But this is not the end of the story. We need the more general notion of **conditional** positive (semi–) definite kernels, and there are several ways to introduce them. They do not fall directly out of a simple (non–distributional) Hilbert space setting, because otherwise they would be unconditionally positive semidefinite. Instead, the most important conditionally positive definite kernels like the **thin–plate spline** $K(x,y) = \log(\|x-y\|_2^2)$ arise directly from applications, or as certain fundamental solutions of partial differential equations. Thus we have to begin with kernels first and then work our way towards a Hilbert space.

We start with the most important univariate function class leading to conditionally positive definite kernels, i.e. the standard **polynomial splines**.

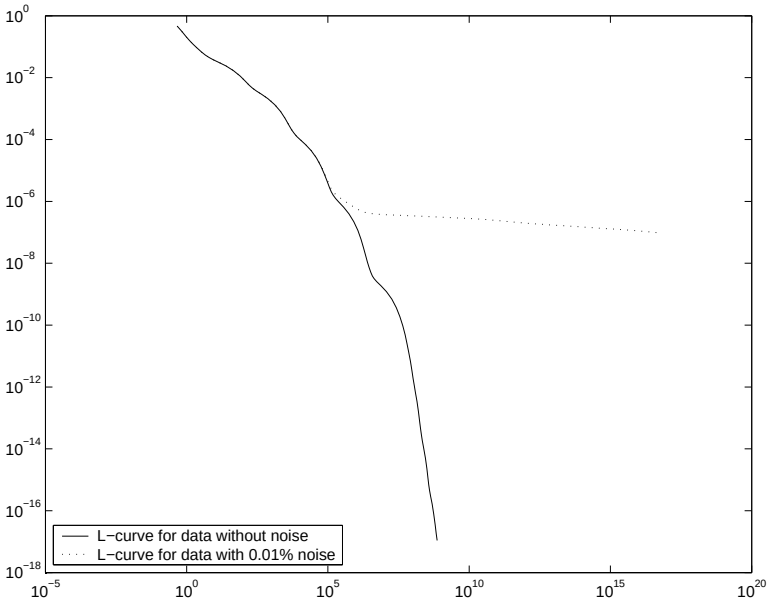

Figure 9: The L-curve for the same problem

4.1 Splines

The following is a somewhat nonstandard introduction to splines, modeled for extensions to general multivariate kernel-based function spaces.

First we fix a positive integer k and denote the space of real–valued **polynomials** with order (= degree -1) at most k by $I\!P_k$. In the d–variate case we shall use the notation $I\!P_k^d$.

4.1.1 Semi–inner product

As a function space, we use the vector space $\mathcal{C}^k[a,b]$ of all real-valued functions f with piecewise continuous k-th derivatives for which

$$|f|_k^2 := \int_a^b \left(\frac{d^k f(t)}{dt^k} \right)^2 dt \tag{4.1}$$

is finite. We leave it to the reader that this defines a reasonable vector space of functions on $[a,b]$.

Equation (4.1) defines a semi-norm, i.e. it has the properties of a norm except for the definiteness, and there is a semi-inner product

$$(f,g)_k := \int_a^b \frac{d^k f(t)}{dt^k} \frac{d^k g(t)}{dt^k} dt.$$

Lemma 4.2 *The seminorm $|f|_k$ is zero if and only if f is a polynomial of order at most k.*

Proof: Clearly, the seminorm $|f|_k$ is zero if f is a polynomial of order at most k. Conversely, if the seminorm $|f|_k$ is zero for some function $f \in \mathcal{C}^k[a,b]$, then $f^{(k)}$ is zero except for its points of discontinuity. Then f consists of polynomial pieces of order at most k which are glued together in such a way that the $(k-1)$st derivative still is continuous. But then f is a global polynomial of order at most k. □

4.1.2 Taylor's Formula

We want to align the above starting point with what we know about positive semidefinite kernels and reproducing kernel Hilbert spaces, but so far we have no inner product and no kernel. But we can go for a **reproduction property** which everybody should be well acquainted with.

Every function f on $[a,b]$ with k continuous derivatives satisfies

$$f(x) = \sum_{j=0}^{k-1} \frac{f^{(j)}(a)}{j!}(x-a)^j + \int_a^x f^{(k)}(t)\frac{(x-t)^{k-1}}{(k-1)!}dt,\ x \in [a,b]$$

and this generalizes to functions in $\mathcal{C}^k[a,b]$ (without proof here). This is a reproduction formula, and in the integral we can see what could later be a kernel, but we still have to work a little.

The upper bound x of the integral can be eliminated by defining the **truncated power** as

$$(z)_+^k := \begin{cases} z^k & z > 0 \\ 0 & z < 0 \\ \frac{1}{2} & z = 0,\ k = 0 \\ 0 & \text{else} \end{cases}$$

to get

$$f(x) = \sum_{j=0}^{k-1} \frac{f^{(j)}(a)}{j!}(x-a)^j + \int_a^b f^{(k)}(t)\frac{(x-t)_+^{k-1}}{(k-1)!}dt,\ x \in [a,b].$$

With the **kernel function**

$$K_k(x,t) := (-1)^k \frac{(x-t)_+^{2k-1}}{(2k-1)!}$$

the above equation takes the form

$$\begin{aligned} f(x) &= \underbrace{\sum_{j=0}^{k-1} \frac{f^{(j)}(a)}{j!}(x-a)^j}_{=:(P_k f)(x)} + (f, K_k(x,\cdot))_k \\ &= (P_k f)(x) + (f, K_k(x,\cdot))_k,\ x \in [a,b]. \end{aligned} \tag{4.3}$$

This is a **reproduction formula**, i.e. it allows f to be reproduced from $f^{(k)}$ in $[a,b]$ and the derivatives at a up to order $k-1$. We also have a kernel now, but it is unsymmetric, and thus it does not fit into our framework.

4.1.3 Taylor's Formula Symmetrized

But note that we have tackled a symmetric problem in an unsymmetric way, which is a mathematical crime. We should also use Taylor's formula at b. This is

$$\begin{aligned} f(x) &= \sum_{j=0}^{k-1} \frac{f^{(j)}(b)}{j!}(x-b)^j + \int_b^x f^{(k)}(t)\frac{(x-t)^{k-1}}{(k-1)!}dt,\ x \in [a,b] \\ &=: (Q_k f)(x) + \int_x^b f^{(k)}(t)(-1)^k \frac{(t-x)^{k-1}}{(k-1)!}dt \\ &= (Q_k f)(x) + \int_a^b f^{(k)}(t)(-1)^k \frac{(t-x)_+^{k-1}}{(k-1)!}dt. \end{aligned}$$

To get something symmetric, we take the mean of the two Taylor formulae. This is

$$\begin{aligned} f(x) &= \tfrac{1}{2}(P_k f)(x) + \tfrac{1}{2}(Q_k f)(x) \\ &\quad + \tfrac{1}{2}\int_a^b f^{(k)}(t) \left(\frac{(x-t)_+^{k-1}}{(k-1)!} + (-1)^k \frac{(t-x)_+^{k-1}}{(k-1)!} \right) dt \\ &=: (R_k f)(x) + (f, \Phi_k(x,\cdot))_k \end{aligned} \tag{4.4}$$

with

$$\begin{aligned} (R_k f)(x) &:= \tfrac{1}{2}(P_k f)(x) + \tfrac{1}{2}(Q_k f)(x) \\ &= \frac{1}{2}\sum_{j=0}^{k-1} \frac{f^{(j)}(a)}{j!}(x-a)^j + \frac{1}{2}\sum_{j=0}^{k-1} \frac{f^{(j)}(b)}{j!}(x-b)^j \\ \Phi_k(x,t) &:= \tfrac{1}{2}(-1)^k \frac{|x-t|^{2k-1}}{(2k-1)!}. \end{aligned}$$

To see that the form of the new symmetric kernel Φ_k is correct, we take its k-th derivative with respect to t for the two cases

$$\begin{aligned} \Phi_k(x,t) &= \tfrac{1}{2}(-1)^k \frac{(x-t)^{2k-1}}{(2k-1)!} \qquad x \geq t \\ \Phi_k(x,t) &= \tfrac{1}{2}(-1)^k \frac{(t-x)^{2k-1}}{(2k-1)!} \qquad t \geq x \end{aligned}$$

and get

$$\begin{aligned} \tfrac{d^k}{dt^k}\tfrac{1}{2}(-1)^k \frac{(x-t)^{2k-1}}{(2k-1)!} &= \tfrac{1}{2}\frac{(x-t)^{k-1}}{(k-1)!} \qquad x \geq t \\ \tfrac{d^k}{dt^k}\tfrac{1}{2}(-1)^k \frac{(t-x)^{2k-1}}{(2k-1)!} &= \tfrac{1}{2}(-1)^k \frac{(t-x)^{k-1}}{(k-1)!} \qquad t \geq x \end{aligned}$$

where we can add the $_+$ subscript in both cases in order to arrive at (4.4).

Note that the two reproduction formulae (4.3) and (4.4) can both be used to our convenience. The different kernels are linked to different polynomial projectors.

4.1.4 Smoothest Interpolation

We assume M points $x_1 < x_2 < \cdots < x_M$ in $[a,b] \subset I\!R$ and corresponding real values $y_1, \ldots, y_M$ to be given, and we want to find a function $s^* \in \mathcal{C}^k[a,b]$ which minimizes $|s|_k^2$ under all functions $s \in \mathcal{C}^k[a,b]$ satisfying the **interpolation conditions**

$$s(x_j) = y_j, \ 1 \leq j \leq M.$$

Note that this is a somewhat more specific case of Theorem 3.3.

In contrast to standard polynomial interpolation, we keep the smoothness k fixed and allow very large numbers M of data points, asking for the "smoothest" possible interpolant. Note that this is an infinite-dimensional quadratic optimization problem with linear constraints. But we shall not plunge deeply into optimization here and try to solve the problem single-handed.

If the data are values $p(x_j) = y_j$ of a polynomial $p \in I\!P_k$, the solution obviously is p with $|p|_k = 0$. To assure uniqueness of interpolation even in such a simple case, we need the additional assumption $M \geq k$.

We shall not directly prove the existence of a smoothest interpolant s^*. Instead, we first assume it exists, then derive its necessary form, and finally prove that it can be numerically calculated in its necessary (and simplified) form, proving existence constructively.

If s^* is our "smoothest" interpolant, we now repeat the "parabola argument" used for proving the characterization of best approximants in Euclidean spaces. Take any real number λ and any function $v \in \mathcal{C}^k[a,b]$ with $v(x_j) = 0,\ 1 \le j \le M$. Then for all such λ and v we have

$$\begin{aligned} |s^* + \lambda v|_k^2 &= |s^*|_k^2 + 2\lambda(s^*, v)_k + \lambda^2 |v|_k^2 \\ &\ge |s^*|_k^2 \end{aligned}$$

and this implies

$$(s^*, v)_k = 0 \text{ for all } v \in \mathcal{C}^k[a,b] \text{ with } v(x_j) = 0,\ 1 \le j \le M. \tag{4.5}$$

This argument can be put upside down and proves that any interpolating function s^* with (4.5) must be a smoothest interpolant.

If we define the linear data map $T : \mathcal{C}^k[a,b] \to I\!R^M$ with

$$Tv := (v(x_1), \ldots, v(x_M)),\ v \in \mathcal{C}^k[a,b],$$

and the linear functional $\mu^*(v) := (s^*, v)_k$, the property (4.5) is

$$\mu^*(v) = 0 \text{ for all } v \in \mathcal{C}^k[a,b] \text{ with } T(v) = 0.$$

But then there is a vector $\alpha \in I\!R^M$ with

$$\mu^*(v) = \alpha^T T(v) \text{ for all } v \in \mathcal{C}^k[a,b].$$

This is a standard argument of linear algebra, see Lemma 9.10 in the Hilbert Space section. It applies since T is surjective and thus the range $I\!R^M = T(\mathcal{C}^k[a,b])$ is isomorphic to the quotient space via $\mathcal{C}^k[a,b]/\ker\ T \xrightarrow{Q} T(\mathcal{C}^k[a,b])$. Since it vanishes on $\ker\ T$, the functional μ^* can be safely defined on the quotient space and thus be written via the range of T as $\mu^* = Q\alpha = \alpha^T T$. We now know that

$$(s^*, v)_k = \alpha^T T(v) = \sum_{j=1}^{M} \alpha_j v(x_j) \tag{4.6}$$

holds for all $v \in \mathcal{C}^k[a,b]$, and we insert (4.3) to get

$$\begin{aligned} (s^*, v)_k &= \sum_{j=1}^{M} \alpha_j \left((P_k v)(x_j) + (v, K_k(x_j, \cdot))_k \right) \\ &= \sum_{j=1}^{M} \alpha_j (P_k v)(x_j) + (v, \sum_{j=1}^{M} \alpha_j K_k(x_j, \cdot))_k \\ (s^* - \sum_{j=1}^{M} \alpha_j K_k(x_j, \cdot), v)_k &= \sum_{j=1}^{M} \alpha_j (P_k v)(x_j). \end{aligned}$$

If we replace v in (4.6) by $P_k v$, we see that

$$0 = (s^*, P_k v)_k = \alpha^T T(P_k v) = \sum_{j=1}^{M} \alpha_j P_k v(x_j)$$

for all $v \in \mathcal{C}^k[a,b]$. Since P_k clearly is surjective, this implies

$$\sum_{j=1}^{M} \alpha_j q(x_j) = 0 \text{ for all } q \in I\!P_k. \tag{4.7}$$

Furthermore,

$$0 = (s^* - \sum_{j=1}^{M} \alpha_j K_k(x_j, \cdot), v)_k$$

for all $v \in \mathcal{C}^k[a,b]$. For the special case

$$v := p := s^* - \sum_{j=1}^{M} \alpha_j K_k(x_j, \cdot)$$

Lemma 4.2 now implies that p is a polynomial in $I\!P_k$. This proves the first part of

Theorem 4.8 *The "smoothest" interpolant s^*, if it exists, has the form*

$$s^* = p + \sum_{j=1}^{M} \alpha_j K_k(x_j, \cdot) \tag{4.9}$$

with a polynomial $p \in I\!P_k$ and M coefficients $\alpha_1, \ldots, \alpha_M$ satisfying (4.7). Conversely, if a function s^* of the form (4.9) with (4.7) interpolates the data, it is the "smoothest" interpolant.

Proof of the converse: Just follow the above argument backwards to arrive at the "parabola argument". Details are left to the reader. □

Note that exactly the same argument works when using the symmetric kernel Φ_k instead of K_k.

4.1.5 Primitive Construction

We still have to prove that the "smoothest interpolant" exists. But since we now know what it should look like, we prove existence constructively. But please keep in mind that there are better algorithms to construct the solution, given in treatises devoted to splines, e.g. Carl de Boor's Practical Guide to Splines [16].

If we introduce a basis $p_1, \ldots, p_k$ for $I\!P_k$, we can write the candidate for a smoothest interpolant as

$$s^* := \sum_{j=1}^{M} \alpha_j K_k(x_j, \cdot) + \sum_{\ell=1}^{k} \beta_\ell p_\ell$$

with the additional conditions (4.7) in the form

$$\sum_{j=1}^{M} \alpha_j p_\ell(x_j) = 0,\ 1 \le \ell \le k.$$

Again, the following argument works similarly for the symmetric kernel Φ_k instead of K_k.

Together with the usual interpolation conditions

$$s^*(x_i) = \sum_{j=1}^{M} \alpha_j K_k(x_j, x_i) + \sum_{\ell=1}^{k} \beta_\ell p_\ell(x_i) = y_i,\ 1 \le i \le M$$

we get the $(M+k)\times(M+k)$ block system

$$\begin{pmatrix} A & P \\ P^T & 0_{\ell\times\ell} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} y \\ 0_\ell \end{pmatrix} \tag{4.10}$$

with the matrices and vectors

$$\begin{aligned} A &:= (K_k(x_j, x_i))_{1\le i,j\le M} \\ P &:= (p_\ell(x_i))_{1\le i\le M,\, 1\le \ell\le k} \\ y^T &:= (y_1, \ldots, y_M). \end{aligned}$$

Theorem 4.11 *If $M \ge k$ holds, the system (4.10) is uniquely solvable.*

Proof: We show that the homogeneous system has only the trivial solution. Assume that a homogeneous solution is given by vectors $\alpha \in I\!R^M$ and $\beta \in I\!R^k$. We then define s^* and p as in the above argument and see that s^* is the smoothest interpolant to zero data. Since the zero function also does the job, we necessarily have $|s^*|_k = 0$ and $s^* \in I\!P_k$. But since s^* interpolates zero in $M \ge k$ points, it must be zero everywhere.

Then, for every $v \in \mathcal{C}^k[a,b]$ we have

$$\begin{aligned} 0 &= (s^*, v)_k \\ &= 0 + \left(\sum_{j=1}^M \alpha_j K_k(x_j, \cdot), v \right)_k \\ &= \sum_{j=1}^M \alpha_j \left(K_k(x_j, \cdot), v \right)_k \\ &= \sum_{j=1}^M \alpha_j \left(v(x_j) - (P_k v)(x_j) \right) \\ &= \sum_{j=1}^M \alpha_j v(x_j) \end{aligned}$$

due to (4.7). By picking some useful v, e.g. as Lagrange interpolating polynomials, we get that all α_j must vanish. But the remaining equations then are $P\beta = 0$ and imply that the polynomial

$$p := \sum_{\ell=1}^k \beta_\ell p_\ell$$

vanishes at all $M \ge k$ data points. Thus its coefficients must all be zero. □

4.1.6 Properties

From Theorem 4.8 and equation (4.9) we see that the smoothest interpolant is of the form

$$s^*(x) = \sum_{\ell=1}^k \beta_\ell p_\ell(x) + \sum_{j=1}^M \alpha_j (x_j - x)_+^{2k-1}$$

or, equivalently, but with different coefficients,

$$s^*(x) = \sum_{\ell=1}^k \beta_\ell p_\ell(x) + \sum_{j=1}^M \alpha_j |x_j - x|_+^{2k-1}$$

with the additional conditions (4.7). Thus it is a piecewise polynomial of order at most $2k$ with "breakpoints" or "knots" at the data locations x_j. It still has $2k-2$ continuous derivatives, which is roughly twice the smoothness originally postulated in the space $\mathcal{C}^k[a,b]$ except for $k=1$ and $k=2$.

Furthermore, the first form tells us that it is a polynomial of order at most k in $[a,x_0]$. Since the equivalent second form is symmetric, we conclude in general that s^* is a polynomial of order at most k outside the data locations.

Altogether, the conditions

1. s^* interpolates in $x_0 < \ldots < x_M$ in $[a,b]$ and
2. is a C^{2k-2} function
3. consisting of polynomials of order at most $2k$ in each data interval $[x_j, x_{j+1}]$ and
4. a polynomial of order at most k outside $[x_0, x_M]$

uniquely define the solution to our problem, which is traditionally called the "natural interpolating spline of order $2k$".

4.1.7 Symmetrization

In view of multivariate kernel–based methods, we take a closer look at the symmetric kernel Φ_k. In particular,

$$(\Phi_k(x,\cdot),\Phi_k(y,\cdot))_k \;=\; \Phi_k(x,y) - (R_k\Phi_k(x,\cdot))(y)$$

and due to symmetry of the two other parts of the above identity,

$$(R_k\Phi_k(x,\cdot))(y) = (R_k\Phi_k(y,\cdot))(x).$$

Lemma 4.12 *If $M \geq k$ holds, and if formed with Φ_k, the matrix A defines a quadratic form which is positive definite on the subspace of vectors $\alpha \in I\!R^M$ with (4.7).*

Proof: The quadratic form defined by A and taken on the vectors $\alpha \in I\!R^M$ with (4.7) is

$$\begin{aligned}
\alpha^T A\alpha &= \sum_{i=0}^{M}\sum_{j=0}^{M}\alpha_i\alpha_j\Phi_k(x_i,x_j) \\
&= \sum_{i=0}^{M}\sum_{j=0}^{M}\alpha_i\alpha_j(\Phi_k(x_i,\cdot),\Phi_k(x_j,\cdot))_k + 0 \\
&= \left(\sum_{i=0}^{M}\alpha_i\Phi_k(x_i,\cdot),\sum_{j=0}^{M}\alpha_j\Phi_k(x_j,\cdot)\right)_k \\
&= \left|\sum_{i=0}^{M}\alpha_i\Phi_k(x_i,\cdot)\right|_k^2 \\
&\geq 0
\end{aligned}$$

and thus positive semidefinite. If it vanishes, then

$$p(x) := \sum_{i=0}^{M}\alpha_i\Phi_k(x_i,x)$$

must be a polynomial in $\mathbb{P}_k$. With the same argument as in the proof of Theorem 4.11, now taking p instead of s^*, we get that all α_i must vanish if (4.7) holds. □

Definition 4.13 *A kernel with the property described by Lemma 4.12 for all matrices arising on $M \geq k$ points is called* **conditionally positive definite of order** k.

This property will come up later in multivariate settings. But we can also push symmetry somewhat further by defining the symmetric kernel

$$\Psi_k(x,y) := \Phi_k(x,y) - (R_k \Phi_k(x,\cdot))(y)$$

which satisfies

$$(\Psi_k(x,\cdot), \Psi_k(y,\cdot))_k = \Psi_k(x,y) \text{ for all } x,y \in [a,b] \tag{4.14}$$

because we only modified Φ by a polynomial. The kernel matrix defined by Ψ instead of Φ also is conditionally positive definite of order k. This follows from the above proof, but with the additional step that if

$$p(x) := \sum_{i=0}^{M} \alpha_i \Psi_k(x_i, x)$$

is a polynomial, then

$$\sum_{i=0}^{M} \alpha_i \Phi_k(x_i, x)$$

is also a polynomial, since the two differ only by a polynomial. The rest is as above. But there is more:

Theorem 4.15 *The kernel Ψ_k is both conditionally positive definite of order k and unconditionally positive* **semidefinite**.

Proof: This is an elementary consequence of (4.14). □

Altogether, this sets the stage for conditional positive definiteness along the lines of Lemma 4.12 and Definition 4.13.

4.1.8 Kernel Regularization

In the above situation, we can take the $k+1$ polynomials $p_\ell(t) := t^{\ell-1}$ for $1 \leq \ell \leq k+1 =: Q = \dim \mathbb{P}_k$ for defining an extended kernel in "Mercer style" as

$$K_k(x,y) := \Phi_k(x,y) + \sum_{\ell=1}^{Q} p_\ell(x) p_\ell(y),$$

where we also could replace Φ_k by Ψ_k. We shall later show that this trick leads to an unconditionally positive definite kernel in the spline situation.

Kernel $\phi(r),\ r = \|x-y\|_2$	Order	Conditions	Name
$(-1)^{\lceil\beta/2\rceil}(c^2+r^2)^{\beta/2}$	$\lceil\beta/2\rceil$	$\beta > 0,\ \beta \notin 2\mathbb{N}$	Multiquadrics
$(-1)^{\lceil\beta/2\rceil}r^\beta$	$\lceil\beta/2\rceil$	$\beta > 0,\ \beta \notin 2\mathbb{N}$	polyharmonic splines
$(-1)^{k+1}r^{2k}\log r$	$k+1$	$k \in \mathbb{N}$	thin–plate splines

Table 1: Orders of conditional positive definiteness

4.1.9 Other Cases

We saw that spline theory leads to univariate conditionally positive definite kernels. But there are other cases, in particular **multivariate** ones which we shall handle in detail later. These are **radial** kernels $K(x,y) = \phi(\|x-y\|_2)$ with scalar functions $\phi : [0.\infty) \to \mathbb{R}$ and orders of conditional positive definiteness given by Table 1.

Like the special spline kernels, such kernels arise naturally and are not directly identifiable as reproducing kernels of certain Hilbert spaces, because otherwise they would be unconditionally positive semidefinite. Thus they have no direct link to Hilbert space theory, and we have to repeat the construction of sections 2.4 and 2.6 to see their connection to Hilbert spaces.

4.2 General Case

To define a sufficiently general notion of conditional positive (semi–) definiteness, we fix a finite–dimensional space $\mathbb{P}$ of functions on a set Ω, denote its dimension by Q and select a basis $p_1, \ldots, p_Q$.

Definition 4.16 *A subset $X = \{x_1, \ldots, x_N\}$ of Ω is called $\mathbb{P}$–***unisolvent***, if zero is the only function in $\mathbb{P}$ that vanishes on X.*

This means that functions from $\mathbb{P}$ are completely determined by their values on unisolvent sets X, and it implies that X must have at least Q elements for this to work. Therefore Ω must have at least Q points, because otherwise the dimension of $\mathbb{P}$ as a space of functions on Ω could not be Q. From now on, all subsets X of Ω we shall only consider must be $\mathbb{P}$–unisolvent and thus have at least Q points. Furthermore, to exclude certain degenerate situations, we assume Ω to admit at least one unisolvent set and at least one additional point. Otherwise, we could get away working with $\mathbb{P}$ alone.

Definition 4.17 *Let a finite–dimensional space $\mathbb{P}$ of real–valued functions on a set Ω be given. A symmetric kernel $K : \Omega \times \Omega \to \mathbb{R}$ is called $\mathbb{P}$–***conditionally positive (semi–) definite***, if for all $\mathbb{P}$–unisolvent subsets $X = \{x_1, \ldots, x_N\}$ of Ω the kernel matrices $A_{K,X,X}$ with entries $K(x_j, x_k),\ 1 \le j,k \le N \ge Q$ are positive (semi–) definite on the subspace of $\mathbb{R}^N$ of vectors $a \in \mathbb{R}^N$ with the* **moment conditions**

$$\sum_{j=1}^{N} a_j p(x_j) = 0 \text{ for all } p \in \mathbb{P}. \tag{4.18}$$

For notational convenience, we also define the $N \times Q$ matrices P_X by

$$P_X := (p_\ell(x_j))_{1 \le j \le N,\ 1 \le \ell \le Q}$$

and note that these have rank Q for $\mathbb{P}$–unisolvent sets X. Furthermore, the moment conditions 4.18 on a vector $a \in \mathbb{R}^N$ have the form $P_X^T a = 0$.

We now describe a general technique for the recovery of functions from data on unisolvent stes X, as we did for splines in (4.10).

Theorem 4.19 *If K is a $\mathbb{P}$–conditionally positive definite kernel on Ω, the $(N+Q)\times(N+Q)$ matrices*

$$\tilde{A}_{K,X,X,\mathbb{P}} := \begin{pmatrix} A_{K,X,X} & P_X \\ P_X^T & 0_{Q\times Q} \end{pmatrix}$$

with

$$A_{K,X,X} = (K(x_k, x_k))_{1\le j,k\le N}$$

are nonsingular for each $\mathbb{P}$–unisolvent finite set $X \subseteq \Omega$.

Proof: Consider a solution $(a^T, b^T)^T$ of the linear homogeneous system

$$\begin{pmatrix} A_{K,X,X} & P_X \\ P_X^T & 0_{Q\times Q} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0_{N\times 1} \\ 0_{Q\times 1} \end{pmatrix}.$$

Then $P_X^T a = 0$ and $A_{K,X,X}a + P_X b = 0$, leading to

$$0 = a^T(A_{K,X,X}a + P_X b) = a^T A_{K,X,X} a$$

and $a = 0$ due to $P_X^T a = 0$ and the definition of conditional positive definiteness. But then $P_X b = 0$, and unisolvence implies $b = 0$. □

We now proceed like in the positive semidefinite case and set up the system

$$\begin{pmatrix} A_{K,X,X} & P_X \\ P_X^T & 0_{Q\times Q} \end{pmatrix} \begin{pmatrix} u^*(x) \\ v^*(x) \end{pmatrix} = \begin{pmatrix} K_X(x) \\ p(x) \end{pmatrix} \tag{4.20}$$

with vectors

$$K_X(x) := (K(x, x_1), \ldots, K(x, x_N))^T,\ p(x) := (p_1(x), \ldots, p_Q(x))^T$$

for all $x \in \Omega$. The system is solvable, and thus we get functions with

$$u^*(x) = (u_1^*(x), \ldots, u_N^*(x))^T,\ v^*(x) = (v_1^*(x), \ldots, v_Q^*(x))^T,\ x \in \Omega$$

which are in the space

$$\tilde{S}_{\mathbb{P},X} :=\ \text{span}\ (\{K(x_j, \cdot)\ :\ x_j \in X\} \cup \{p_1, \ldots, p_Q\}). \tag{4.21}$$

Specializing to the first N columns of the system and using unique solvability, we get

$$u_j^*(x_k) = \delta_{jk},\ 1 \le j,k \le N,\ v_\ell^*(x_k) = 0,\ 1 \le \ell \le Q,\ 1 \le k \le N. \tag{4.22}$$

These Lagrange conditions lead us to define the interpolation operator

$$Q_X(f)(x) := \sum_{j=1}^N u_j^*(x) f(x_j)$$

which acts linearly on all functions f which have values on X and provides an interpolating function from the space $\tilde{S}_{\mathbb{P},X}$. The system also implies $P_X^T u(x) = p(x)$, and this is

$$p_\ell(x) = \sum_{j=1}^N u_j^*(x) p_\ell(x_j) = Q_X(p_\ell)(x), 1 \le \ell \le Q$$

and means that the functions from $\mathbb{P}$ are reproduced by Q_X. But the range of Q_X cannot have a dimension larger than N, while the space $\tilde{S}_{\mathbb{P},X}$ is spanned by $N+Q$ functions. To see that Q_X maps to the subspace

$$S_{\mathbb{P},X} := \mathbb{P} + \{a^T K_X(\cdot) \ : \ P_X^T a = 0, \ a \in \mathbb{R}^N\} \tag{4.23}$$

of $\tilde{S}_{\mathbb{P},X}$ with additional moment conditions, we write the inverse of $\tilde{A}_{K,X,X,\mathbb{P}}$ in the form

$$\begin{pmatrix} B & R \\ R^T & S \end{pmatrix}$$

and get the equations

$$\begin{array}{rcl} u(x) & = & BK_X(x) \quad + \quad Rp(x) \\ v(x) & = & R^T K_X(x) \quad + \quad Sp(x) \\ BP_X & = & 0 \end{array}$$

proving that the rows of B satisfy the moment conditions via

$$P_X^T (e_j^T B)^T = (e_j^T B P_X)^T = 0, \ 1 \le j \le N.$$

Therefore, the interpolation operator must necessarily have the alternative form

$$Q_X(f)(x) := \sum_{j=1}^{N} a_j(f) K(x_j, x) + \sum_{\ell=1}^{Q} b_\ell(f) p_\ell(x) \tag{4.24}$$

with $P_X^T a(f) = 0$ since the result lies in $S_{\mathbb{P},X}$. The coefficients solve the system

$$\begin{pmatrix} A_{K,X,X} & P_X \\ P_X^T & 0_{Q\times Q} \end{pmatrix} \begin{pmatrix} a(f) \\ b(f) \end{pmatrix} = \begin{pmatrix} f_X \\ 0 \end{pmatrix} \tag{4.25}$$

with $f_X := (f(x_1), \ldots, f(x_N))^T$, and this is how the interpolant is calculated in practice.

We note some facts we need later:

Lemma 4.26 *1. In (4.23), the sum is* **direct**, *i.e. there is no nonzero function in $\mathbb{P}$ which can be written in the form $a^T K_X(x)$ with a unisolvent set X.*

2. The dimension of $S_{\mathbb{P},X}$ is N.

3. Both operators Q_X and $Id - Q_X$ are **projectors**, *i.e. they are linear and* **idempotent**.

Proof of the lemma: The representation (4.24) of the interpolation operator via the solution of the system (4.25) is unique. A function in $\mathbb{P}$, when written as its own interpolant, thus coincides with the right–hand sum, and thus the first sum must be zero, proving the first assertion.

The second follows similarly: each function in $S_{\mathbb{P},X}$ has exactly the form (4.24) and satisfies the system (4.25) with its own values on X in the right–hand side. Providing N arbitrary values on the right–hand side will produce a function in $S_{K,X}$, and the corresponding map is bijective due to nonsingularity of the matrix. Thus the space spanned by the solution coefficients, and thus also the space $S_{\mathbb{P},X}$ are isomorphic to $\mathbb{R}^N$. The third assertion follows from the Lagrange property (4.22). □

4.3 Inner Products

There are many ways to construct a "native" Hilbert space when starting from a $\mathbb{P}$–conditionally positive definite kernel K. They all have some advantages and some drawbacks. Here, we take an approach which is symmetric with respect to duality, but we shall not yet go over to the Hilbert space completion.

We go back to the notation used in section 2.4 and (2.18), but we impose the additional moment conditions. We fix Ω, $\mathbb{P}$ and K in the notation and define

$$\begin{aligned}
M &:= \Big\{(a,X) \,:\, X \subset \Omega,\ \mathbb{P}\text{–unisolvent},\ |X| =: N,\ a \in \mathbb{R}^N,\ P_X^T a = 0\Big\},\\
\lambda_{a,X}(f) &:= \sum_{j=1}^{N} a_j f(x_j) \text{ for all } (a,X) \in M,\ f \,:\, \Omega \to \mathbb{R},\\
f_{a,X}(x) &:= \lambda_{a,X}^t K(t,x) = a^T K_X(x) = \sum_{j=1}^{N} a_j K(x_j,x) \text{ for all } (a,X) \in M,\ x \in \Omega\\
L &:= \{\lambda_{a,X} \,:\, (a,X) \in M\},\\
F &:= \{f_{a,X} \,:\, (a,X) \in M\}.
\end{aligned}$$

Note that point evaluation functionals are not in L.

To prove that L is a vector space, we have to show that we can add two functionals $\lambda_{a,X} \in L$ and $\lambda_{b,Y} \in L$ to get a functional $\lambda_{c,X\cup Y} \in L$. But this is easy if we sum the appropriate factors for points in $X \cap Y$ and keep the single weights for the rest, i.e. we write

$$\lambda_{a_X,X} + \lambda_{b_Y,Y} = \lambda_{a_{X\cap Y}+b_{X\cap Y},X\cap Y} + \lambda_{a_{X\setminus(X\cap Y)},X\setminus(X\cap Y)} + \lambda_{b_{Y\setminus(X\cap Y)},Y\setminus(X\cap Y)}$$

in a sloppy but understandable notation. Thus we get another functional in L satisfying the moment condition, because it holds for both summands.

Now we define the bilinear form

$$(\lambda_{a,X}, \lambda_{b,Y})_K := \lambda_{a,X}^x \lambda_{b,Y}^y K(x,y)$$

on L and see that it defines also a bilinear form on F via

$$\begin{aligned}
(\lambda_{a,X}, \lambda_{b,Y})_K &= \lambda_{a,X}^x \lambda_{b,Y}^y K(x,y)\\
&= \lambda_{a,X}(f_{b,Y})\\
&= \lambda_{b,Y}(f_{a,X})\\
&=: (f_{a,X}, f_{b,Y})_K
\end{aligned}$$

and it is easy to see that it is well–defined (just repeat the proof following (2.19)). By the definition of conditional positive definiteness, the bilinear forms are definite on L and F.

Remember that in the unconditionally positive **semi**definite case we got a positive **definite** inner product. As a little detour, let us now check what happens if K is only conditionally positive **semi**definite. To prove that the form on L is positive definite, we have to look at the $\lambda_{a,X} \in L$ with $(\lambda_{a,X}, \lambda_{a,X})_K = 0$. These can have the property that $f_{a.X} = \lambda_{a,X}^t K(t,\cdot) \in \mathbb{P}$ for nonzero a, and we can only arrive at $\lambda_{a,X} = 0$ in case of $\mathbb{P} = \{0\}$, i.e. in the unconditionally positive semidefinite case.

Thus L and F can be completed to abstract Hilbert spaces $\overline{L}$ and $\overline{F}$, but so far we have no idea how to interpret their elements as functionals or functions, respectively. But a later section will help.

To care for our recovery functions, we should go back to the spaces $S_{I\!P,X}$. They are subspaces of

$$N := I\!P + F \tag{4.27}$$

which is a direct sum due to Theorem 4.26. We call it the **Pre–Native Space** for K here and postpone its completion. Only the F part carries an inner product, but we can define a bilinear form on N via

$$(p + f_{a,X}, q + f_{b,Y})_N := (f_{a,X}, f_{b,Y})_K$$

for all $p + f_{a,X}, q + f_{b,Y} \in N = I\!P + F$ with $p, q \in I\!P$. This does not generate a norm on N, because it vanishes for functions in $I\!P$. Thus it only generates a seminorm on N via

$$|p + f_{a,X}|_N := \|f_{a,X}\|_K \text{ for all } p + f_{a,X} \in N.$$

Any functional $\lambda_{b,Y} \in L$ acts on functions $p + f_{a,X} \in N$ as

$$\lambda_{b,Y}(p + f_{a,X}) = \lambda_{b,Y} f_{a,X} = (\lambda_{a,X}\lambda_{b,Y})_L = (p + f_{a,X}, \lambda^t_{b,Y} K(t,\cdot))_N$$

which is a generalized reproduction equation

$$\lambda(f) = (f, \lambda^t K(t,\cdot))_N \text{ for all } \lambda \in L,\ f \in N. \tag{4.28}$$

At this point, we could already complete F under the above inner product to some new space $\overline{F}$, and then define the native space for K to be

$$\mathcal{N}_K := I\!P + \overline{F}.$$

But then it needs some further work to see that the abstract completion is a useful space of functions on Ω. Furthermore, this space is not a Hilbert space unless we define a useful inner product there. Thus we postpone the completion to a later section.

4.4 Optimal Recovery

Before we go over to a full definition of a native space, we want to look at optimality principles we had before. We fix a finite $I\!P$–unisolvent set $X = \{x_1, \ldots, x_N\} \subset \Omega$ and consider the finite–dimensional space $S_{I\!P,X}$ of (4.23). Then we have an analog of Theorem 3.1:

Theorem 4.29 *Given a $I\!P$–conditionally positive definite kernel K on some set Ω, and let X be a $I\!P$–unisolvent subset of Ω. For all functions $f \in N$, the solution $s^* = p^* + f_{a^*,X} \in S_{I\!P,X}$ of the finite–dimensional approximation problem*

$$\min_{s \in S_{P,X}} |f - s|_N = \min_{p \in I\!P, a \in I\!R^N,\ P_X^T a = 0} |f - p - f_{a,X}|_N$$

exists and is an interpolant to f on X. The coefficients solve the linear system (4.25).

Proof: We already know that there is an interpolant solving (4.25), and we can write it as $s^* = p^* + f_{a^*,X} \in S_{I\!P,X}$. If we take any other $s = p + f_{b,X} \in S_{I\!P,X}$, we can insert into the inner product to get

$$\begin{aligned}(f - s^*, s)_N &= (f - s^*, f_{b,X})_N \\ &= \lambda_{b,X}(f - s^*) = 0\end{aligned}$$

by interpolation and the reproduction formula. With the usual "parabola" argument we see that this proves that s^* is an optimal solution. □

Theorem 4.30 *Given an arbitrary function f from N, the interpolant s^* on a $I\!P$–unisolvent set X is the function with minimal seminorm under all other interpolants to f on X taken from N, i.e.*

$$\min_{g \in N,\, g=f \text{ on } X} |g|_N = |s^*|_N.$$

Proof: When writing $s^* = p^* + f_{a^*,X}$ we have to show

$$|g|_N^2 \geq |p^* + f_{a^*,X}|_N^2$$

for all $g \in N$ with $g = f$ on X. But we can write

$$\begin{aligned}|g|_N^2 &= |p^* + f_{a^*,X} + (g - p^* - f_{a^*,X})|_N^2 \\ &= |p^* + f_{a^*,X}|_N^2 + 2(p^* + f_{a^*,X}, g - p^* - f_{a^*,X})_N + |g - f_{a^*,X}|_N^2 \\ &= |p^* + f_{a^*,X}|_N^2 + 2(f_{a^*,X}, g - p^* - f_{a^*,X})_N + |g - f_{a^*,X}|_N^2\end{aligned}$$

and get the assertion from

$$\begin{aligned}(f_{a^*,X}, g - p^* - f_{a^*,X})_N &= \lambda_{a^*,X}(g - p^* - f_{a^*,X}) \\ &= \sum_{j=1}^N a_j^* \left(g(x_j) - p^*(x_j) - f_{a^*,X}(x_j)\right) \\ &= \sum_{j=1}^N a_j^* \left(g(x_j) - f(x_j)\right) = 0\end{aligned}$$

proving the theorem. □

The third optimality principle described by Theorem 3.10 is not as easy to recover here, because we cannot evaluate norms of point evaluation functionals. But we can use functionals which vanish on $I\!P$, which is sufficient for our purposes. Take any $I\!P$–unisolvent set X with $|X| =: N$ and any fixed point $x \in \Omega$. We know that we can recover all functions in $I\!P$ from their values on X, and thus there are coefficients $a_1(x), \ldots, a_N(x)$ such that

$$p(x) = \sum_{j=1}^N a_j(x) p(x_j) \text{ for all } p \in I\!P.$$

Thus there are functionals of the form $\lambda_{b,X \cup \{x\}} \in L$, e.g. the one defined as

$$\lambda_{b,X \cup \{x\}}(f) := f(x) - \sum_{j=1}^N a_j(x) f(x_j)$$

with the special coefficient vector

$$b = (1, -a_1(x), \ldots, a_N(x))^T \in I\!R^{N+1}$$

and the set

$$M(x,X) := \left\{ a \in I\!R^N \ : \ p(x) = \sum_{j=1}^N a_j p(x_j) \text{ for all } p \in I\!P \right\}$$

is not empty. Thus we can pose the minimization problem

$$\min_{a \in M(x,X)} \sup_{f \in N,\, |f|_N \neq 0} \frac{\left| f(x) - \sum_{j=1}^N a_j f(x_j) \right|}{|f|} \tag{4.31}$$

and prove

Theorem 4.32 *The minimum of the expression (4.31) is attained for the Lagrange interpolant on X based on K, i.e. the optimal $a_j^*(x)$ are the $u_j^*(x)$ of (4.20).*

Proof: If we take any admissible vector $a \in M(x,X)$, we have

$$f(x) - \sum_{j=1}^N a_j f(x_j) \ = \ (f, K(x,\cdot) - \sum_{j=1}^N a_j K(x_j,\cdot))_N$$

because the left–hand side is the action of a functional

$$\mu_{a,x}(f) := f(x) - \sum_{j=1}^N a_j f(x_j)$$

in L. Then we invoke the Cauchy–Schwarz inequality to get

$$\left| f(x) - \sum_{j=1}^N a_j f(x_j) \right| \leq |f|_N \left| K(x,\cdot) - \sum_{j=1}^N a_j K(x_j,\cdot) \right|_N$$

and the inner sup in the objective function satisfies

$$\sup_{f \in N,\, |f|_N \neq 0} \frac{\left| f(x) - \sum_{j=1}^N a_j f(x_j) \right|}{|f|} \leq \left| K(x,\cdot) - \sum_{j=1}^N a_j K(x_j,\cdot) \right|_N .$$

To prove equality above, we take $f := \mu_{a,x}^t K(t,\cdot)$ and see that

$$\begin{aligned} (f,f)_N &= (\mu_{a,x}, \mu_{a,x})_L \\ &= (\mu_{a,x}^t K(t,\cdot), \mu_{a,x}^t K(t,\cdot))_N \\ &= \left| K(x,\cdot) - \sum_{j=1}^N a_j K(x_j,\cdot) \right|_N^2 , \\ f(x) - \sum_{j=1}^N a_j f(x_j) &= \mu_{a,x}^t f(t) \\ &= (\mu_{a,x}, \mu_{a,x})_L . \end{aligned}$$

Thus the minimization problem boils down to minimizing

$$\left| K(x,\cdot) - \sum_{j=1}^N a_j K(x_j,\cdot) \right|_N^2 = \|\mu_{a,x}\|_L^2$$

under all $a \in M(x,X)$. We can write any $a \in M(x,X)$ as $a = u + v$ with our $u = (u_1^*(x), \ldots, u_N^*(x))^T$ and an arbitrary v with $(v,X) \in M$. This means

$$\mu_{a,x} = \mu_{u+v,x} = \mu_{u,x} + \lambda_{v,X}$$

and we get

$$\begin{aligned} \|\mu_{a,x}\|_L^2 &= \|\mu_{u+v,x}\|_L^2 \\ &= (\mu_{u+v,x}, \mu_{u+v,x})_L \\ &= (\mu_{u,x} + \lambda_{v,X}, \mu_{u,x} + \lambda_{v,X})_L \\ &= (\mu_{u,x}, \mu_{u,x})_L + 2(\lambda_{v,X}, \mu_{u,x})_L + (\lambda_{v,X}, \lambda_{v,X})_L \\ &\geq (\mu_{u,x}, \mu_{u,x})_L \end{aligned}$$

if we can prove $(\lambda_{v,X}, \mu_{u,x})_L = 0$. But this is

$$\begin{aligned} (\lambda_{v,X}, \mu_{u,x})_L &= \lambda_{v,X}^t \left(K(x,t) - \sum_{j=1}^N u_j^*(x) K(x_j,t) \right) \\ &= \sum_{k=1}^N v_k \left(K(x,x_k) - \sum_{j=1}^N u_j^*(x) K(x_j,x_k) \right) = 0. \end{aligned}$$

□

We also note the following error bound:

Corollary 4.33 *For interpolation on any $\mathbb{P}$–unisolvent set X there is the error bound*

$$\left| f(x) - \sum_{j=1}^N u_j^*(x) f(x_j) \right| \leq |f|_N P_X(x) \text{ for all } f \in N,\ x \in \Omega$$

with the **power function** *defined exactly as in (3.12).* □

4.5 Projector to $\mathbb{P}$

We still have a choice of basis in $\mathbb{P}$. But for use below, we want to make a special selection. Due to conditional positive definiteness, we can take a fixed unisolvent set $\Xi \subset \Omega$ consisting of Q points $\xi_1, \ldots, \xi_Q$. Such sets can be constructed from any other unisolvent set X by dropping rows of P_X keeping the rank at Q. Then we can replace the basis $p_1, \ldots, p_Q$ of $\mathbb{P}$ by a Lagrange basis. Thus we shall assume from now on that the basis $p_1, \ldots, p_Q$ of $\mathbb{P}$ is chosen in such a way that it is a Lagrange basis with respect to some unisolvent subset $\Xi = \{\xi_1, \ldots, \xi_Q\} \subset \Omega$, i.e.

$$p_j(\xi_k) = \delta_{jk},\ 1 \leq j,k \leq Q. \tag{4.34}$$

This defines a linear interpolating projector

$$Q_\Xi\ :\ f \mapsto \sum_{\ell=1}^Q p_\ell(\cdot) f(\xi_\ell)$$

on all functions f on Ω, mapping them to their unique interpolating function from $\mathbb{P}$ on Ξ.

It will later turn out to be very useful to apply the mapping $Id - Q_\Xi$ to both arguments of the conditionally positive definite kernel K to generate a new symmetric **normalized kernel**

$$\begin{aligned} K_\Xi(x,y) &:= ((Id-Q_\Xi)^s(Id-Q_\Xi)^t K(t,s))(x,y) \\ &= K(y,x) - (Q_\Xi^t K(t,x))(y) - Q_\Xi^s(K(y,s))(x) + Q_\Xi^s((Q_\Xi^t K(t,s))(y))(x) \\ &= K(y,x) - \sum_{j=1}^{Q} p_j(y)K(\xi_j,x) - \sum_{k=1}^{Q} p_k(x)K(y,\xi_k) + \sum_{j,k=1}^{Q} p_j(y)p_k(x)K(\xi_j,\xi_k) \end{aligned} \tag{4.35}$$

where the first line is somewhat sloppy, but due to symmetry, there is no ambiguity. By definition, this kernel is symmetric and satisfies

$$K_\Xi(\xi_j,\cdot) = K_\Xi(\cdot,\xi_j) = 0,\ 1 \le j \le Q.$$

Thus it does not make sense to use it on Ξ, but on $\Omega \setminus \Xi$ it works perfectly:

Theorem 4.36 *If K is a $\mathbb{P}$–conditionally positive definite kernel on Ω, and if Ξ is a $\mathbb{P}$–unisolvent subset of Ω, then the kernel K_Ξ of (4.35) is unconditionally positive definite on $\Omega \setminus \Xi$.*

Proof: We take any subset $Y = \{y_1, \ldots, y_N\}$ of $\Omega \setminus \Xi$ and any vector $a \in \mathbb{R}^N$ and look at

$$\begin{aligned} & a^T A_{Y,Y,K_\Xi} a \\ = & \sum_{r,s=1}^{N} a_r a_s \left(K(y_r,y_s) - \sum_{j=1}^{Q} p_j(y_r)K(\xi_j,y_s) \right. \\ & \left. - \sum_{k=1}^{Q} p_k(y_s)K(y_r,\xi_k) + \sum_{j,k=1}^{Q} p_j(y_r)p_k(y_s)K(\xi_j,\xi_k) \right) \\ = & \sum_{r,s=1}^{N} a_r a_s K(y_r,y_s) - \sum_{r=1}^{N} a_r \sum_{j=1}^{Q} p_j(y_r) \sum_{s=1}^{N} a_s K(\xi_j,y_s) \\ & - \sum_{s=1}^{N} a_s \sum_{k=1}^{Q} p_k(y_s) \sum_{r=1}^{N} a_r K(y_r,\xi_k) + \sum_{j,k=1}^{Q} \underbrace{\sum_{r=1}^{N} a_r p_j(y_r)}_{=:b_j} \underbrace{\sum_{s=1}^{N} a_s p_k(y_s)}_{=:b_k} K(\xi_j,\xi_k) \\ = & \sum_{r,s=1}^{N} a_r a_s K(y_r,y_s) - \sum_{j=1}^{Q} b_j \sum_{s=1}^{N} a_s K(\xi_j,y_s) \\ & - \sum_{k=1}^{Q} b_k \sum_{r=1}^{N} a_r K(y_r,\xi_k) + \sum_{j,k=1}^{Q} b_j b_k K(\xi_j,\xi_k). \end{aligned}$$

This is a quadratic form, defined for the unisolvent set $Y \cup \Xi$ with the coefficients $a_r,\ 1 \le r \le N$ and $-b_j,\ 1 \le j \le Q$. In order to apply conditional positive definiteness, we have to check the moment conditions there. They are

$$\begin{aligned} & \sum_{r=1}^{N} a_r p_\ell(y_r) - \sum_{j=1}^{Q} b_j p_\ell(\xi_j) \\ = & \sum_{r=1}^{N} a_r p_\ell(y_r) - \sum_{j=1}^{Q} \left(\sum_{r=1}^{N} a_r p_j(y_r) \right) p_\ell(\xi_j) \\ = & \sum_{r=1}^{N} a_r \left(p_\ell(y_r) - \sum_{j=1}^{Q} p_j(y_r) p_\ell(\xi_j) \right) = 0,\ 1 \le \ell \le Q \end{aligned}$$

due to the Lagrange property. Thus the quadratic form is positive definite. □

Let us now go back to the projector Q_Ξ. For each $x \in \Omega$ we can define the functional

$$\delta_{(x)} f := (Id - Q_\Xi)(f)(x) = f(x) - \sum_{\ell=1}^{Q} p_\ell(x) f(\xi_\ell)$$

which clearly lies in L. We can insert this functional twice into the inner product on L and get

$$(\delta_{(x)}, \delta_{(y)})_L = \delta^s_{(x)} \delta^t_{(y)} K(s,t) = K_\Xi(x,y)$$

for all $x, y \in \Omega$, proving that K_Ξ is a good candidate for a reproducing kernel in a space whose dual should be L. If we pick an arbitrary $\lambda \in L$, we can form the function

$$g_\lambda(x) := (\delta_{(x)}, \lambda)_L$$

and get a map R from L into the space G of all these functions by

$$R(\lambda) = g_\lambda.$$

This map is linear and surjective. If $g_{\lambda_{a,X}} = 0$ for some $\lambda_{a,X} \in L$, then

$$\begin{aligned} 0 &= \sum_{k=1}^{N} a_k g_{\lambda_{a,X}}(x_k) \\ &= \sum_{k=1}^{N} a_k (\delta_{(x_k)}, \lambda_{a,X})_L \\ &= \sum_{k=1}^{N} a_k \sum_{j=1}^{N} a_j K(x_j, x_k) - \sum_{k=1}^{N} a_k \sum_{\ell=1}^{Q} p_\ell(x_k) \sum_{j=1}^{N} a_j K(x_j, \xi_\ell) \\ &= \sum_{k=1}^{N} a_k \sum_{j=1}^{N} a_j K(x_j, x_k) \end{aligned}$$

and $a = 0$ due to conditional positive definiteness. Thus R is injective on L, and we can define an inner product on $G = R(L)$ by

$$(g_\lambda, g_\mu)_G := (\lambda, \mu)_L$$

to make G isometrically isomorphic to L with Riesz map $R : L \to G$.

Now G and L can be completed, and the completion will have the extension of R as a Riesz map $\overline{L} \to \overline{G}$. This proves that $\overline{G}$ is the space of all functions which can be written as

$$g_\lambda(x) := (\delta_{(x)}, \lambda)_{\overline{L}}$$

for all $\lambda \in \overline{L}$, and it is a Hilbert space under the bilinear form which we redefine as

$$(g_\lambda, g_\mu)_{\overline{G}} := (\lambda, \mu)_L \text{ for all } \lambda, \mu \in \overline{L}.$$

This means that each abstract element $g \in \overline{G}$ has a representation

$$g(x) = (R(\delta_{(x)}), g)_G \text{ for all } x \in \Omega, g \in \overline{G}$$

as a function and $\overline{G}$ is a Hilbert space with the reproducing kernel

$$R(\delta_{(x)})(y) = (\delta_{(y)}), \delta_{(x)})_{\overline{G}} = K_\Xi(x,y).$$

But since $\delta_{(\xi_j)} = 0$ holds for all $\xi_j \in \Xi$, we see that $g(\xi_j) = 0$ for all $\xi_j \in \Xi$ and all $g \in \overline{G}$ by this assignment of function values.

Theorem 4.37 *Let K be a $\mathbb{P}$–conditionally positive definite kernel on Ω, and let Ξ be a unisolvent subset of Ω consisting of Q points. Then there is a Hilbert space $\overline{G}$ of functions on Ω vanishing on Ξ, and it has K_Ξ as reproducing kernel.* □

4.6 Native Space

The above construction is nice, but it still has some unnatural dependence on Ξ which should be eliminated. In particular, the Hilbert space $\overline{G}$ contains only functions vanishing on Ξ. Fortunately, values in Ξ are accounted for by functions in $\mathbb{P}$, and we should thus add them back into the scenery.

On the other hand, our pre–native space $N = \mathbb{P} + F$ of (4.27) does not care for function values being zero on Ξ. We thus want to show

$$G + \mathbb{P} = N = F + \mathbb{P} \tag{4.38}$$

and perform a completion process on N later. To prove (4.38) we take a $g_{\lambda_{a,X}} + p \in G + \mathbb{P}$ with $\lambda_{\lambda_{a,X}} \in L$. Then

$$\begin{aligned} g_{\lambda_{a,X}}(x) &= (\delta_{(x)}, \lambda_{a,X})_L \\ &= \delta^t_{(x)} \lambda^s_{a,X} K(t,s) \\ &= \lambda^s_{a,X} K(x,s) - \sum_{\ell=1}^{Q} p_\ell(x) \lambda^s_{a,X} K(\xi_\ell, s) \\ &= f_{a,X}(x) - \sum_{\ell=1}^{Q} p_\ell(x) \lambda^s_{a,X} K(\xi_\ell, s) \\ &= f_{a,X}(x) - (Q_\Xi(f_{a,X}))(x) \\ &\in F + \mathbb{P} = N, \end{aligned}$$

in particular

$$g_{\lambda_{a,X}} = f_{a,X} - Q_\Xi(f_{a,X})$$

leading to

$$G = (Id - Q_\Xi)(F)$$

and proving $G + \mathbb{P} = N = F + \mathbb{P}$.

Clearly, the sum $N = G + \mathbb{P}$ is direct because functions from G vanish on Ξ, and Ξ is $\mathbb{P}$–unisolvent. If we take any $f \in N$ and write it as $f = p + g$ with $p \in \mathbb{P}$ and $g \in G$, then automatically

$$(Id - Q_\Xi)f = (Id - Q_\Xi)(p + g) = (Id - Q_\Xi)g = g \in G.$$

Thus the direct sum $N = \mathbb{P} + G$ can be written via the projectors Q_Ξ and $Id - Q_\Xi$ and the splitting

$$N \ni f = Q_\Xi(f) + (Id - Q_\Xi)(f) \in \mathbb{P} + G.$$

Then we can define a bilinear form on N by

$$(f,g)_\Xi := \sum_{j=1}^{Q} Q_\Xi(f)(\xi_j) Q_\Xi(g)(\xi_j) + (f - Q_\Xi(f), g - Q_\Xi(g))_G \text{ for all } f, g \in N. \tag{4.39}$$

This clearly defines a pre-Hilbert space with $I\!P$ being a Q–dimensional subspace with orthogonal complement $I\!P^\perp = G$. In particular, we have

$$\begin{array}{rcll} (f,g)_\Xi & = & (f,g)_G & \text{for all } f,g \in G \\ (g,p)_\Xi & = & 0 & \text{for all } g \in G,\ p \in I\!P. \end{array}$$

The functions $p_1, \ldots, p_Q$ are orthonormal in this inner product due to their Lagrange property, and consequently

$$(p+s, p_\ell)_\Xi = (p+s, p_\ell + 0)_\Xi = p(\xi_\ell)\ 1 \le \ell \le Q.$$

The canonical projector to $I\!P$ in this definition thus is Q_Ξ due to

$$p+s \mapsto \sum_{\ell=1}^{Q} (p+s,p_\ell)_\Xi p_\ell = \sum_{\ell=1}^{Q} p(\xi_\ell) p_\ell(\cdot) = \sum_{\ell=1}^{Q} (p+s)(\xi_\ell) p_\ell(\cdot) = Q_\Xi(p+s) = p$$

for all $p \in I\!P$ and $s \in G$.

We should relate the new inner product $(.,.)_\Xi$ on N with the semi–inner product $(.,.)_N$ we defined before. If we write arbitrary elements of $N = I\!P + F$ as $f = p + f_{a,X},\ g = q + f_{b,Y}$, we get

$$\begin{array}{rcl} (f,g)_N & = & (p + f_{a,X}, q + f_{b,Y})_N \\ & = & (f_{a,X}, f_{b,Y})_F \\ & = & (\lambda_{a,X}, \lambda_{b,Y})_L \\ & = & (g_{\lambda_{a,X}}, g_{\lambda_{b,Y}})_G \\ & = & (f_{a,X} - Q_\Xi(f_{a,X}), f_{b,Y} - Q_\Xi(f_{b,Y}))_G \\ & = & (f - Q_\Xi(f), g - Q_\Xi(g))_G \\ & = & (f - Q_\Xi(f), g - Q_\Xi(g))_\Xi \end{array} \tag{4.40}$$

proving that the right–hand side is independent of Ξ, and the inner product on N can be rewritten as

$$(f,g)_\Xi = \sum_{j=1}^{Q} Q_\Xi(f)(\xi_j) Q_\Xi(g)(\xi_j) + (f,g)_N \text{ for all } f,g \in N.$$

When going over to the completion, the finite–dimensional orthogonal subspace $I\!P$ stays fixed, and the completion of N under the inner product $(.,.)_\Xi$ must be the orthogonal direct sum of $I\!P$ with the completion $\overline{G}$ of G.

Definition 4.41 *Let K be a $I\!P$–conditionally positive definite kernel on a set Ω, and let Ξ be a unisolvent subset of Ω with Q points, supplying a Lagrange basis $p_1, \ldots, p_Q$ of $I\!P$ on Ξ. Then we define the* **native space** *for K to be the orthogonal direct sum*

$$\mathcal{N}_{K,\Xi} := I\!P + \overline{G} \tag{4.42}$$

under the extension of the above inner product $(.,.)_\Xi$. The canonical projector onto $I\!P$ is Q_Ξ, and each abstract element $p + g_\lambda \in \mathcal{N}_{K,\Xi}$ with $p \in I\!P$ and $g_\lambda \in \overline{G}$ with $\lambda \in \overline{L}$ is a function on Ω with

$$(p+g_\lambda)(x) := p(x) + (\delta_{(x)}, \lambda)_{\overline{L}} \text{ for all } x \in \Omega.$$

Thus the orthogonal decomposition of each $f = p + g_\lambda \in \mathcal{N}_{K,\Xi} = I\!P + \overline{G}$ is

$$f = p + g_\lambda = \underbrace{Q_\Xi(f)}_{=p} + \underbrace{(Id - Q_\Xi)(f)}_{=g_\lambda}.$$

This leads to the generalized reproduction equation

$$\begin{aligned} f(x) &= Q_\Xi(f)(x) + (Id - Q_\Xi)(f)(x) \\ &= Q_\Xi(f)(x) + ((Id - Q_\Xi)(f), K_\Xi(x,\cdot))_{\overline{G}} \\ &= Q_\Xi(f)(x) + (f, K_\Xi(x,\cdot))_\Xi \text{ for all } f \in \mathcal{N}_{K,\Xi},\ x \in \Omega, \end{aligned}$$

where the final step uses the definition of K_Ξ and orthogonality in the sense

$$(Q_\Xi(f), (Id - Q_\Xi)(g))_\Xi = 0 \text{ for all } f, g \in \mathcal{N}_{K,\Xi}.$$

But we can simplify the kernel in the reproduction equation somewhat, using (4.40) and the definition of K_Ξ for

$$\begin{aligned} f(x) - Q_\Xi(f)(x) &= ((Id - Q_\Xi)(f), K_\Xi(x,\cdot))_{\overline{G}} \\ &= ((Id - Q_\Xi)(f), ((Id - Q_\Xi)^t K(t,\cdot))(x))_N \\ &= ((Id - Q_\Xi)(f), K(x,\cdot) - \sum_{\ell=1}^{Q} p_\ell(x) K(\xi_\ell,\cdot))_N \end{aligned}$$

bringing us back to the original kernel K and the earlier semi–inner product.

Readers should be aware of several pitfalls here: the equations

$$\begin{aligned} f(x) &= Q_\Xi(f)(x) + (f, ((Id - Q_\Xi)^t K(t,\cdot))(x))_N \text{ for all } f \in \mathcal{N}_{K,\Xi},\ x \in \Omega \\ f(x) &= Q_\Xi(f)(x) + (f - Q_\Xi f, ((Id - Q_\Xi)^t K(t,\cdot))(x))_F \text{ for all } f \in \mathcal{N}_{K,\Xi},\ x \in \Omega \end{aligned}$$

are incorrect.

The equation (4.40) teaches us that the closure of F under the inner product on F is the same as the closure under the inner product $(.,.)_\Xi$ of $(Id - Q_\Xi)(F)$, which is the closure $\overline{G}$ under the inner product $(.,.)_\Xi$ of $G = (Id - Q_\Xi)(F)$. This proves

Corollary 4.43 *The native space $\mathcal{N}_{K,\Xi}$ can also be written as $I\!P + \overline{F}$, and thus it is independent of Ξ as a linear space of functions.* □

But note that this statement does not carry over to the topology of the native space. The native space consists of functions on Ω and is independent of Ξ as a space of functions on Ω, but it carries at least two topologies:

1. the extension of the semi–inner product on F to $\overline{F}$ induces a Ξ–independent semi–inner product,

2. the inner product $(.,.)_\Xi$ generates a Ξ–dependent Hilbert space structure,

and the two are connected via (4.40). The reproduction formulas are always Ξ–dependent and come in different forms:

$$\begin{aligned} f(x) - Q_\Xi(f)(x) &= ((Id - Q_\Xi)(f), K_\Xi(x,\cdot))_{\overline{G}} \\ &= (f, K_\Xi(x,\cdot))_\Xi \\ &= ((Id - Q_\Xi)(f), K(x,\cdot) - \sum_{\ell=1}^{Q} p_\ell(x) K(\xi_\ell,\cdot))_N \end{aligned}$$

for all $f \in \mathcal{N}_{K,\Xi},\ x \in \Omega$.

As an example, we should look back at the spline case.

Theorem 4.44 *The native space for splines of order $2k$ or degree $2k-1 \geq 1$ on $[a,b] \subset \mathbb{R}$ is the* **Beppo–Levi space** *of functions whose generalized k-th derivative is in $L_2[a,b]$, and it carries the seminorm $(.,.)_k$ defined in (4.1). If the norm is defined via (4.39), the native space is norm–equivalent to the well–known Sobolev space $H_2^k[a,b]$, and it is identical to this space as a set of functions.*

"Proof": We do not want to deal with Lebesgue integration or generalized derivatives here, and thus we present a short and incomplete argument. First, we know that each function in $\mathbb{P}+F$ is of the form in (4.9) where the kernel $K_k(x,t)$ can be written in terms of $(x-t)_+^{2k-1}$ or variations thereof. Its k–th derivative is of the form $(x-t)_+^{k-1}$ and thus still a square–integrable function. Since the square–integrable piecewise continuous functions, when completed under the L_2 norm, lead to the standard Hilbert space $L_2[a,b]$ (we suppress a detailed definition and proof here), one can prove that the completion $\overline{F}$ and $\mathbb{P}$ are in the space of functions with k–th derivative in L_2.

But the main point is that all such functions are reached that way. Thus we now assume we have a function f on $[a,b]$ with a square–integrable k–th derivative, and we assume that it is orthogonal in the seminorm to all functions in F. Given any unisolvent set $X = \{x_1, \ldots, x_N\} \subset [a,b]$ of at least k points and a vector $a \in \mathbb{R}^N$ with the moment conditions $P_X^T a = 0$, we see that

$$\begin{aligned} 0 &= (f, \sum_{j=1}^{N} a_j K_k(x_j, \cdot))_k \\ &= \sum_{j=1}^{N} a_j \left(f(x_j) - (P_k(f))(x_j)\right) \\ &= \sum_{j=1}^{N} a_j f(x_j) \end{aligned}$$

holds. In particular, we now know that all divided differences on k or more points vanish, if applied to f. If we interpolate f on any set of k points and apply the usual error representation involving divided differences on $k+1$ points, we get f must be a polynomial of order at most k. Thus $f \in \mathbb{P}$ with zero seminorm, proving that the Beppo–Levi space is exactly $\mathbb{P} + \overline{F}$.

The Beppo–Levi space carries the k–th order Sobolev seminorm, and it is identical to Sobolev space $H_2^k[a,b]$ as a vector space of functions. Under any other norm extending the seminorm by norming polynomials of order up to k, one gets a topology which is norm–equivalent to Sobolev space. Our construction in (4.39) is just one way to do that. □

5 Practical Observations

.... incomplete...

In particular, I should add all the MATLAB programming hints that I gave for the exercises.

This chapter is from a draft of a book, and it is not yet linked into this text properly. Quite some cross–references are hanging in the air, in particular those leading to methods for solving partial differential equations. Furthermore, this chapter focuses on radial basis functions and ignores more general kernels.

Before we go on with serious theory, we should present some experimental results.

5.1 Lagrange Interpolation

In Figure 10 we have 150 scattered data points in $[-3, 3]^2$ in which we interpolate the MATLAB `peaks` function (top right). The next row shows the interpolant using Gaussians, and the absolute error. The lower row shows MATLAB's standard technique for interpolation of scattered data using the `griddata` command. The results are typical for such problems: radial basis function interpolants recover smooth functions very well from a sample of scattered values, provided that the values are noiseless and the underlying function is smooth.

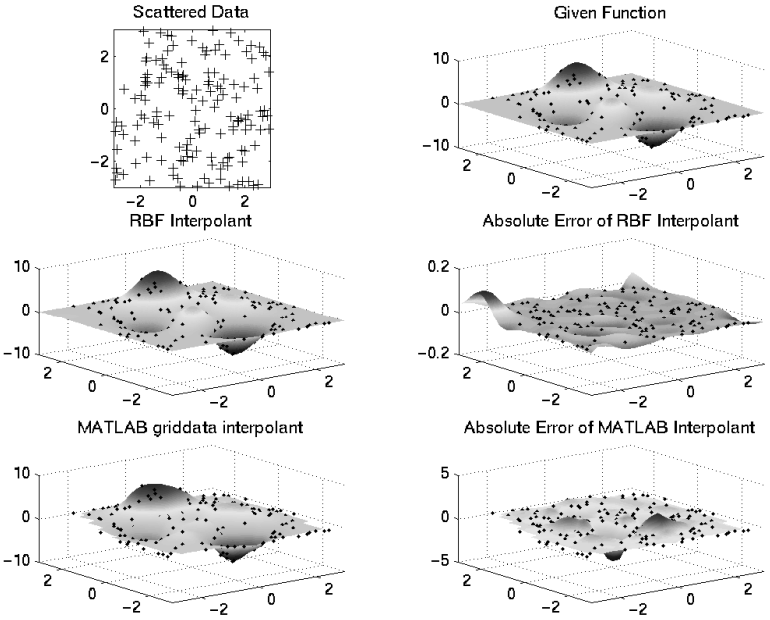

Figure 10: Interpolation by radial basis functions

The ability of radial basis functions to deal with arbitrary point locations in arbitrary dimensions is very useful when geometrical objects have to be constructed, parametrized, or warped, see e.g. [3, 12, 50, 11, 51, 53, 73, 6]. In particular, one can use such transformations to couple incompatible finite element codes [2].

Furthermore, interpolation of functions has quite some impact on methods solving partial differential equations.

Another important issue is the possibility to parametrize spaces of translates of kernels not via coefficients, but via function values at the translation centers. This simplifies meshless methods "*constructing the approximation entirely in terms of nodes*" [5]. Since kernel interpolants approximate higher derivatives well, local function values can be used to provide good estimates for derivative data [69]. This has connections to **p**seudospectral methods [19].

5.2 Interpolation of Mixed Data

It is quite easy to allow much more general data for interpolation by radial basis functions. For example, consider recovery of a multivariate function f from data including the values $\frac{\partial f}{\partial x_2}(z),\ \int_\Omega f(t)dt$. The basic trick, due to Z.M. Wu [75], is to use special trial functions

$$\begin{array}{ll} \dfrac{\partial \phi(\|x-z\|_2)}{\partial x_2} & \text{for} \quad \dfrac{\partial f}{\partial x_2}(z) \\ \displaystyle\int_\Omega \phi(\|x-t\|_2)dt & \text{for} \quad \displaystyle\int_\Omega f(t)dt \end{array}$$

to cope with these requirements. In general, if a linear functional λ defines a data value $\lambda(f)$ for a function f as in the above cases with $\lambda_1(f) = \frac{\partial f}{\partial x_2}(z),\ \lambda_2(f) = \int_\Omega f(t)dt$, the special trial function $u_\lambda(x)$ to be added is

$$u_\lambda(x) := \lambda^t \phi(\|x-t\|_2) \text{ for } \lambda^t(f(t))$$

where the upper index denotes the variable the functional acts on. If $m = n$ functionals $\lambda_1, \ldots, \lambda_m$ are given, the span (2.9) of trial functions is to be replaced by

$$u(x) = \sum_{k=1}^{n} a_k \lambda_k^t \phi(\|x-t\|_2).$$

The interpolation system (2.11) turns into

$$\lambda_j u = \sum_{k=1}^{n} a_k \lambda_k^t \lambda_j^x \phi(\|x-t\|_2),\ 1 \le j \le n \tag{5.1}$$

with a symmetric matrix composed of $\lambda_k^t \lambda_j^x \phi(\|x-t\|_2),\ 1 \le j,k \le n$ which is positive definite if the functionals are linearly independent and ϕ is positive definite.

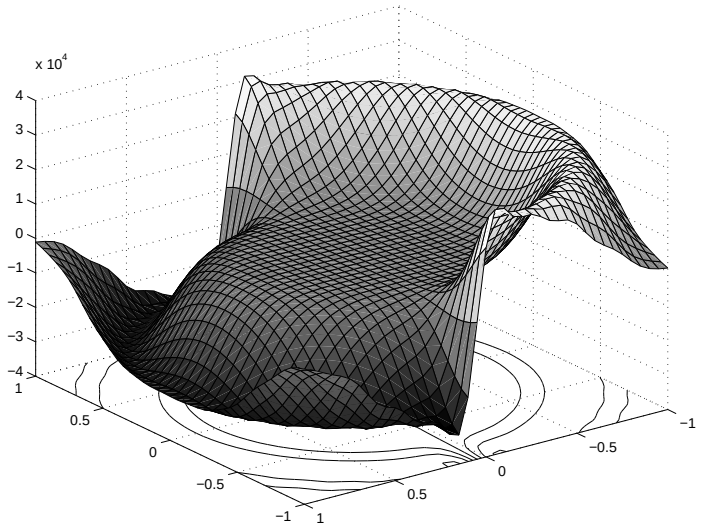

Figure 11: Generalized interpolant to Neumann data

To give an example with general functionals, Figure 11 shows an interpolation to Neumann data +1 and -1 on each half of the unit circle, respectively, in altogether 64 points by linear combinations of properly scaled Gaussians.

In case of conditionally positive definite radial basis functions, the span of (4.21) or (4.24) turns into

$$u(x) := \sum_{k=1}^{n} a_k \lambda_k^t \phi(\|x - t\|_2) + \sum_{\ell=1}^{q} b_\ell p_\ell(x)$$

while the additional condition (4.18) is replaced by

$$\sum_{k=1}^{n} \alpha_k \lambda_k^t p_\ell(t) = 0,\ 1 \le \ell \le q$$

and the interpolation problem is solvable, if the additional condition

$$\lambda_k^t p(t) = 0 \text{ for all } 1 \le k \le n \text{ and } p \in P_{Q-1}^d \text{ implies } p = 0$$

is imposed, replacing (4.18) and $I\!P$–unisolvency.

Another example of recovery from non-Lagrange data is the construction of **Lyapounov basins** from data consisting of **orbital derivatives** [24, 25].

The flexibility to cope with general data is the key to various applications of radial basis functions within methods solving partial differential equations. Collocation techniques, as treated in books on numerical methods for solving partial differential equations, solve partial differential equations numerically by interpolation of values of differential operators and boundary conditions.

Another important aspect is the possibility to implement additional linear conditions or **constraints** like

$$\lambda(u) := \int_\Omega u(x)dx = 1$$

on a trial function. For instance, this allows to handle conservation laws and is inevitable for **finite-volume methods**. A constraint like the one above, when used as additional data, adds another degree of freedom to the trial space by addition of the basis function $u_\lambda(x) := \lambda^t \phi(\|x - t\|_2)$, and at the same time it uses this additional degree of freedom to satisfy the constraint. This technique deserves much more attention in applications.

5.3 Error Behavior

If exact data come from smooth functions f, and if smooth kernels K or radial basis functions ϕ are used for interpolation, users can expect very small interpolation errors. In particular, the error goes to zero when the data samples are getting dense. The actual error behavior is limited by the smoothness of both f and ϕ. Quantitative error bounds can be obtained from the standard literature [10, 71] and recent papers [48]. They are completely *local*, and they are in terms of the **fill distance**

$$h := h(X, \Omega) := \sup_{y \in \Omega} \min_{x \in X} \|x - y\|_2 \tag{5.2}$$

of the discrete set $X = \{x_1, \ldots, x_n\}$ of centers with respect to the domain Ω where the error is measured. The interpolation error converges to zero for $h \to 0$ at a rate dictated by the minimum smoothness of f and ϕ. For infinitely smooth radial basis functions like the Gaussian or multiquadrics, convergence even is exponential [40, 76] like $\exp(-c/h)$. Derivatives are also convergent as far as the smoothness of f and ϕ allows, but at a smaller rate, of course.

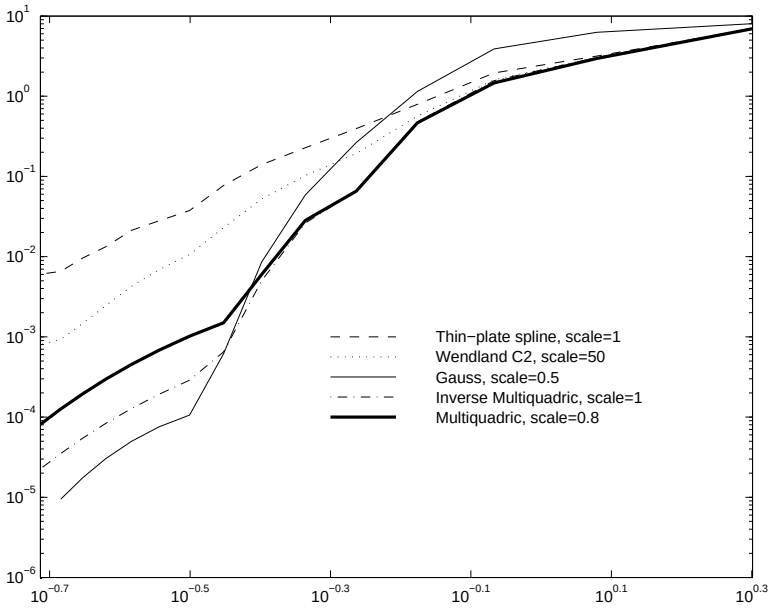

Figure 12: Nonstationary interpolation to a smooth function as a function of fill distance

This is particularly important when applications require good reproductions of derivatives, e.g. velocity fields or stress tensors.

For interpolation of the smooth `peaks` function provided by MATLAB and used already in Figure 10, the error behavior on $[-3,3]^2$ as a function of fill distance h is given by Figure 12. It can be clearly seen that smooth ϕ yield smaller errors with higher convergence rates. In contrast to this, Figure 13 shows interpolation to the nonsmooth function

$$f(x,y) = 0.03 * \max(0, 6 - x^2 - y^2)^2, \tag{5.3}$$

on $[-3,3]^2$, where now the convergence rate is dictated by the smoothness of f instead of ϕ and is thus more or less fixed. Excessive smoothness of ϕ never spoils the error behavior, but induces excessive instability, as we shall see later.

5.4 Stability

But there is a serious drawback when using radial basis functions on dense data sets, i.e. with small fill distance. The condition of the matrices used in (2.11) and (5.1) will get extremely large if the **separation distance**

$$S(X) := \frac{1}{2} \min_{1 \le i < j \le n} \|x_i - x_j\|_2$$

of points of $X = \{x_1, \ldots, x_n\}$ gets small. Figure 14 shows this effect in the situation of Figure 12.

If points are distributed well, the separation distance $S(X)$ will be proportional to the fill distance $h(X, \Omega)$ of (5.2). In fact, since the fill distance is the radius of the largest ball with arbitrary center in the underlying domain Ω without any data point in its interior, the separation distance $S(X)$ is the radius of the smallest ball anywhere without any data point in its interior, but with at least two points of X on the boundary. Thus for convex domains

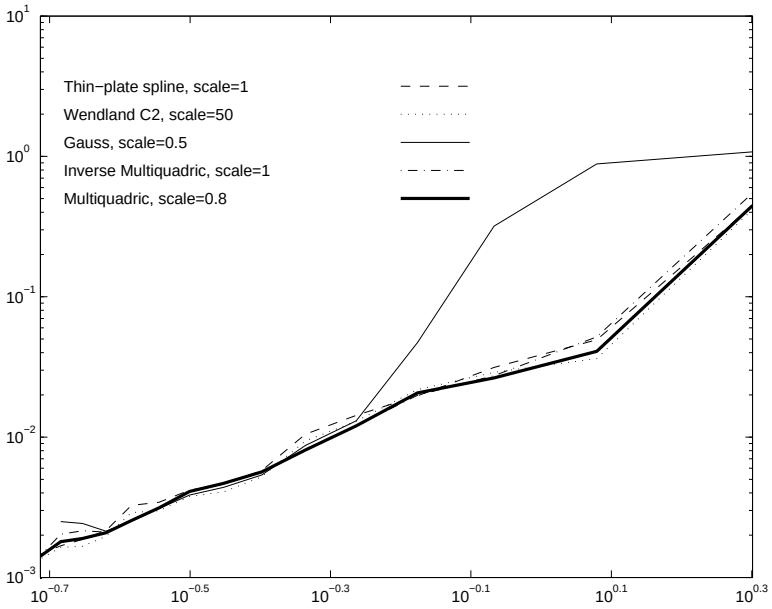

Figure 13: Nonstationary interpolation to a nonsmooth function as a function of fill distance

one always has $S(X) \leq h(X, \Omega)$. But since separation distance only depends on the closest pair of points and ignores the rest, it is reasonable to avoid unusually close points leading to some $S(X)$ which is considerably smaller than $h(X, \Omega)$. Consequently, a distribution of data locations in X is called **quasi–uniform** if there is a positive **uniformity constant** $\gamma \leq 1$ such that

$$\gamma \, h(X, \Omega) \leq S(X) \leq h(X, \Omega). \tag{5.4}$$

To maintain quasi-uniformity, it suffices in most cases to delete "duplicates". Furthermore, there are sophisticated "thinning" techniques [20, 17, 72] to keep fill and separation distance proportional, i.e. to assure quasi-uniformity at multiple scaling levels.

5.5 Uncertainty Principle

Unless radial basis functions are rescaled in a data-dependent way, it can be proven [56] that there is a close link between error and stability, even if fill and separation distance are proportional. In fact, both are tied to the smoothness of ϕ, letting stability become worse and errors become smaller when taking smoother radial basis functions. This is kind of an **Uncertainty Principle**:

> It is impossible to construct radial basis functions which guarantee good stability and small errors at the same time.

We illustrate this by an example. Since [56] proves that the square of the L_∞ error roughly behaves like the smallest eigenvalue of the interpolation matrix, Figure 15 plots the product of the MATLAB condition estimate `condest` with the square of the L_∞ error for the nonstationary interpolation of the MATLAB `peaks` function, used already for Figures 12, 6, and 14 to show the error and condition behavior there. Note that the curves do not vary much if compared to Figure 14.

Thus **smoothness** of radial basis functions must be chosen with some care, and chosen dependent on the smoothness of the function to be approximated. From the point of view of

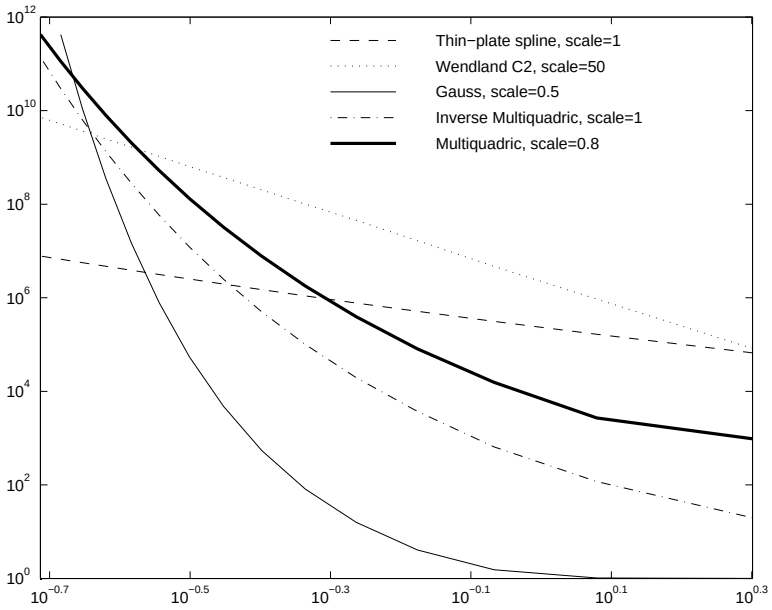

Figure 14: Condition as function of separation distance

reproduction quality, smooth radial basis functions can well recover nonsmooth functions, as proven by papers concerning error bounds [48]. On the other hand, non-smooth radial basis functions will not achieve high convergence rates when approximating smooth functions [62]. This means that using too much smoothness in the chosen radial basis function is not critical for the error, but rather for the stability. But in many practical cases, the choice of smoothness is not as sensible as the choice of scale, as discussed in section 5.6.

5.6 Scaling

If radial basis functions are used directly, without any additional tricks and treats, users will quickly realize that **scaling** is a crucial issue. The literature has two equivalent ways of scaling a given radial basis function ϕ, namely replacing it by either $\phi(\|x-y\|_2/c)$ or by $\phi(\epsilon\|x-y\|_2)$ with c and ϵ being positive constants. Of course, these scalings are equivalent, and the case $\epsilon \to 0,\ c \to \infty$ is called the **flat limit** [18]. In numerical methods for solving differential equations, the **scale parameter** c is preferred, and it is called **shape factor** there. Readers should not be irritated by slightly other ways of scaling, e.g.

$$\phi_c(\|x\|_2) := \sqrt{c^2 + \|x\|_2^2} = c \cdot \sqrt{1 + \frac{\|x\|_2^2}{c^2}} = c \cdot \phi_1\left(\frac{\|x\|_2}{c}\right) \tag{5.5}$$

for multiquadrics, because the outer factor c is irrelevant when forming trial spaces from functions (2.9). Furthermore, it should be kept in mind that only the **polyharmonic spline** and its special case, the **thin-plate spline** generate trial spaces which are scale-invariant.

Like the tradeoff between error and stability when choosing smoothness (see the preceding section), there often is a similar tradeoff induced by scaling: a "wider" scale improves the error behavior but induces instability. Clearly, radial basis functions in the form of sharp spikes will lead to nearly diagonal and thus well-conditioned systems (2.11), but the error behavior is disastrous, because there is no reproduction quality between the spikes. The opposite case of

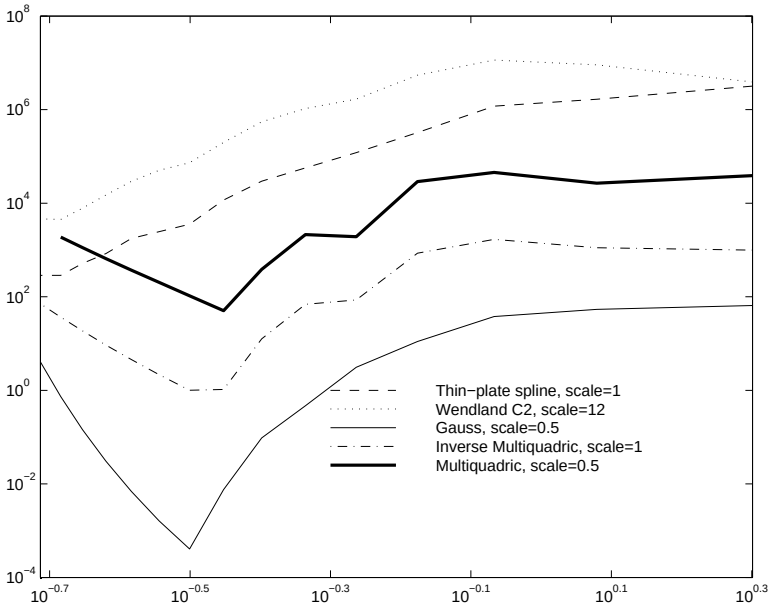

Figure 15: Squared L_∞ error times condition as a function of fill distance

extremely "flat" and locally close to constant radial basis functions leads to nearly constant and thus badly conditioned matrices, while many experiments show that the reproduction quality is even improving when scales are made wider, as far as the systems stay solvable.

For **analytic** radial basis functions (i.e. in C^∞ with an expansion into a power series), this behavior has an explanation: the interpolants often converge towards polynomials in spite of the degeneration of the linear systems [18, 58, 35, 36, 60]. This has implications for many examples in this text which approximate analytic solutions of partial differential equations by analytic radial basis functions like Gaussians or multiquadrics: whatever is calculated is close to a good polynomial approximation to the solution. Users might suggest to use polynomials right away in such circumstances, but the problem is to pick a good polynomial basis. For multivariate problems, choosing a good polynomial basis must be data-dependent, and it is by no means clear how to do that. It is one of the intriguing properties of analytic radial basis functions that they automatically choose good data-dependent polynomial bases when driven to their "flat limit". There are new techniques [34, 22] which circumvent the instability at large scales, but these are still under investigation.

Figure 16 shows the error for interpolation of the smooth MATLAB `peaks` function on a fixed data set, when interpolating radial basis functions ϕ are used with varying scale relative to a ϕ-specific starting scale given in the legend. Only those cases are plotted which have both an error smaller than 1 and a condition not exceeding 10^{12}. Since the data come from a function which has a good approximation by polynomials, the analytic radial basis functions work best at their condition limit. But since the `peaks` function is a superposition of Gaussians of different scales, the Gaussian radial basis function still shows some variation in the error as a function of scale.

Interpolating the nonsmooth function (5.3) shows a different behavior (see Figure 17), because now the analytic radial basis functions have no advantage for large scales. In both cases one can see that the analytic radial basis functions work well only in a rather small scale range, but there they beat the other radial basis functions. Thus it often pays off to select a good scale or to circumvent the disadvantages of large scales [34, 22].

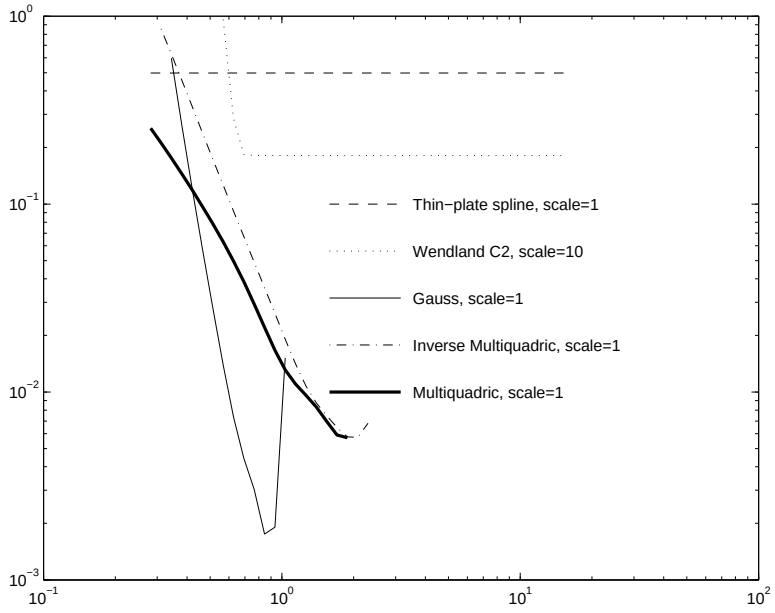

Figure 16: Error as function of relative scale, smooth case

Like in finite element methods, users might want to scale the basis functions in a data-dependent way, making the scale c in the sense of using $\phi(\|x-y\|_2/c)$ proportional to the fill distance h as in (5.2). This is often called a **stationary** setting, e.g. in the context of wavelets and quasi-interpolation. If the scale is fixed, the setting is called **nonstationary**, and this is what we were considering up to this point. Users must be aware that the error and stability analysis, as described in the previous sections, apply to the nonstationary case, while the stationary case will not converge for $h \to 0$ in case of unconditionally positive definite radial basis functions [8, 9]. But there is a way out: users can influence the "relative" scale of c with respect to h in order to achieve a good compromise between error and stability. The positive effect of this can easily be observed [57], and for special situations there is a sound theoretical analysis called **approximate approximation** [41]. Figure 18 shows the stationary error behavior for interpolation of the smooth MATLAB `peaks` function when using different radial basis functions ϕ at different starting scales. It can be clearly seen how the error goes down to a certain small level depending on the smoothness of ϕ, and then stays roughly constant. Using larger starting radii decreases these saturation levels, as Figure 19 shows.

Due to the importance of *relative* scaling, users are strongly advised to always run their programs with an *adjustable* scale of the underlying radial basis functions. Experimenting with small systems at different scales give a feeling of what happens, and users can fix the relative scale of c versus h rather cheaply. Final runs on large data can then use this relative scaling. In many cases, given problems show a certain "intrinsic" preference for a certain scale, as shown in Figure 17, but this is an experimental observation which still is without proper theoretical explanation.

5.7 Practical Rules

If users adjust the smoothness and the scaling of the underlying radial basis function along the lines of the previous sections, chances are good to get away with relatively small and sufficiently stable systems. The rest of the text contains plenty of examples for this observation.

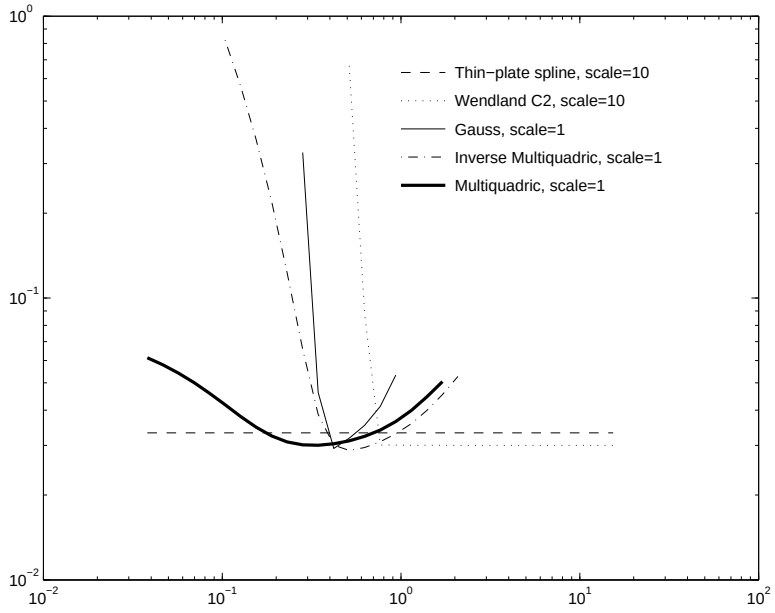

Figure 17: Error as function of relative scale, nonsmooth case

For completeness, we add a few rules for Scientific Computing with radial basis functions, in particular concerning good choices of scale and smoothness. Note that these apply also to methods for solving partial differential equations in later chapters.

- Always allow a scale adjustment.
- If possible, allow different RBFs to choose from.
- Perform some experiments with scaling and choice of RBF before you turn to tough systems for final results.
- If you do not apply iterative solvers, do not worry about large condition numbers, but use a stabilized solver, e.g. based on Singular Value Decomposition (SVD). Remember that unless you apply certain tricks, getting a good reproduction quality will always require bad condition. If you need k decimal digits of final accuracy for an application, do not bother about condition up to 10^{12-k}.
- If you use compactly supported radial basis functions, do not expect them to work well when each support contains less than about 50 neighbors. This means that the bandwidth of large sparse systems should not be below 50. Increasing bandwidth will usually improve the quality of the results at the expense of computational complexity.
- When using either compactly supported or quickly decaying radial basis functions of high smoothness, the theoretical support and the practical support do not coincide. In such cases one should enforce sparsity by chopping the radial basis functions, in spite of losing positive definiteness properties. But this should be done with care, and obeying the "50 neighbors" rule above.
- If systems get large and ill-conditioned, and if change of scale and RBF do not improve the situation, try methods described in the following section.

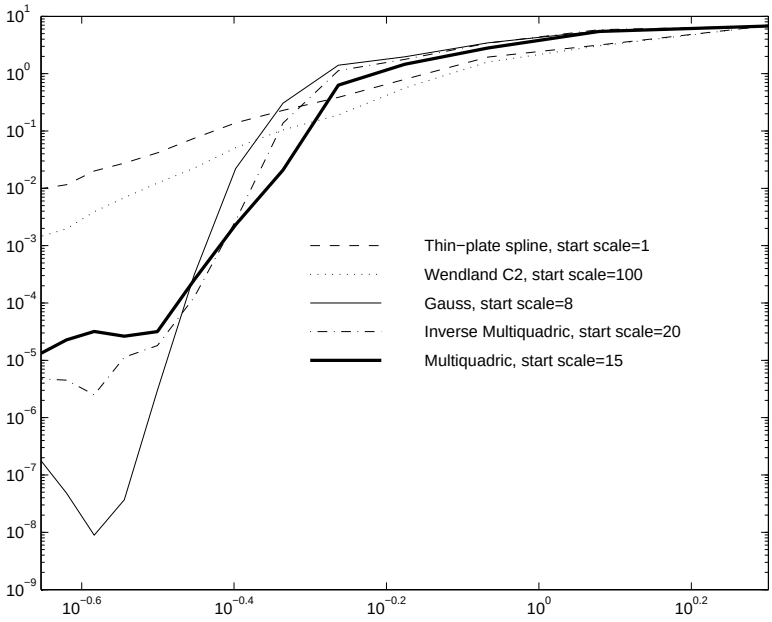

Figure 18: Stationary interpolation to a smooth function at small starting scales

- Use blockwise iteration ("domain decomposition") first, because it is simple and often rather efficient.
- Blockwise iteration can be speeded up by precalculation of LR decompositions of blocks.
- If all of this does not work, try partitions of unity, multilevel methods, or special preconditioning techniques. You are now at current research level, and you should look into the next section.

5.8 Sensitivity to Noise

So far, the discussion focused on noiseless data, with the exception of Figure 7. If users expect **noise** in the data, an interpolatory recovery is not appropriate, because it treats noise as data. In most of the later sections of this text, data are right-hand sides or boundary values for partial differential equations, and they usually are given as noiseless functions which can be evaluated anywhere. Thus the rest of the text does not treat noisy inputs in detail. But at this point, some remarks are appropriate.

In all noisy situations, interpolation should be replaced by approximation. This can be done in various ways leading to **stabilization**.

A primitive, but often quite sufficient technique is to run a smoothing process on the raw data and to recover the unknown function from the smoothed data instead of the raw data.

Another standard trick is to solve (2.11) in the L_2 sense with oversampling, if only $n << m$ trial points x_j are used for m data points y_k. The trial points can then be placed rather freely with a large separation distance, while a small separation distance of data points will not have a dramatic effect on stability any more. However, there is not very much theoretical and practical work done on unsymmetric recovery techniques [59, 61, 60].

A third possibility is the old Levenberg-Marquardt trick of adding a positive value λ into the diagonal of the kernel matrix of (2.11) with entries $\phi(\|x_j - x_k\|_2)$. As is well-known from

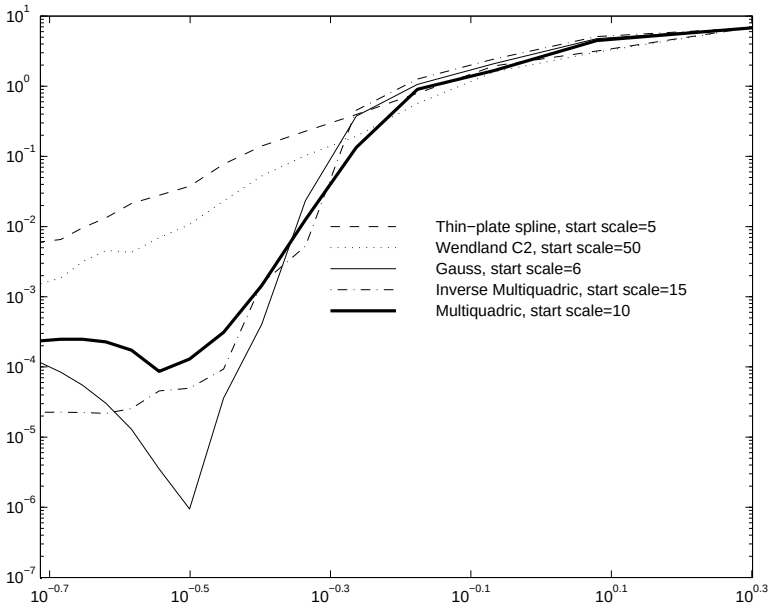

Figure 19: Stationary interpolation to a smooth function at wider starting scales

literature on spline smoothing, this leads to an approximant achieving a tradeoff between smoothness and reproduction quality which can be controlled by λ. If a stochastic background is available, there are methods to estimate λ properly, e.g. by **cross-validation**. However, in most cases users adjust λ experimentally. This technique also helps to fight instability when working on irregularly distributed data [72], because it is able to shift the stability from dependence on the separation distance to dependence on the fill distance (see section 5.4).

A fourth possibility is **regularization**, for example using a singular-value decomposition as described in section 3.4.

In general, one can replace the system (2.11) by an **optimization method** which penalizes the reproduction error on one hand and either a complexity or smoothness criterion on the other, allowing a fair amount of control over the tradeoff between error and stability. Penalties for the discrete reproduction error can be made in various discrete norms, the ℓ_1 and ℓ_∞ case having the advantage to lead to linear optimization restrictions, while the discrete ℓ_2 norm leads to quadratic ones. For radial basis functions of the form (2.9) or (4.21), the quadratic form

$$\|u\|_\phi^2 := \sum_{j,k=1}^{n} \alpha_j \alpha_k \phi(\|x_j - x_k\|_2) \tag{5.6}$$

is a natural candidate for penalizing high derivatives without evaluating any. This is due to the standard fact that the above expression is a squared norm in a **native space** of functions with about half the smoothness of ϕ, penalizing all available derivatives there. For details, we have to refer to basic literature [10, 71] on the theory of radial basis functions. But even though we skip over native spaces here, all users should be aware that they always lure in the theoretical background, and that all methods based on radial basis functions implicitly minimize the above quadratic form under all functions in the native space having the same data. This has a strong **regularization** effect which is the background reason why radial basis function or more general **kernel methods** work well for many **ill-posed** and **inverse problems** [31, 37, 67, 14, 13, 32, 33, 38, 55, 49]. The above strategy of minimizing the quadratic form (5.6) also is central for modern methods of **machine learning**, but we cannot pursue

this subject in detail here [15, 65, 66].

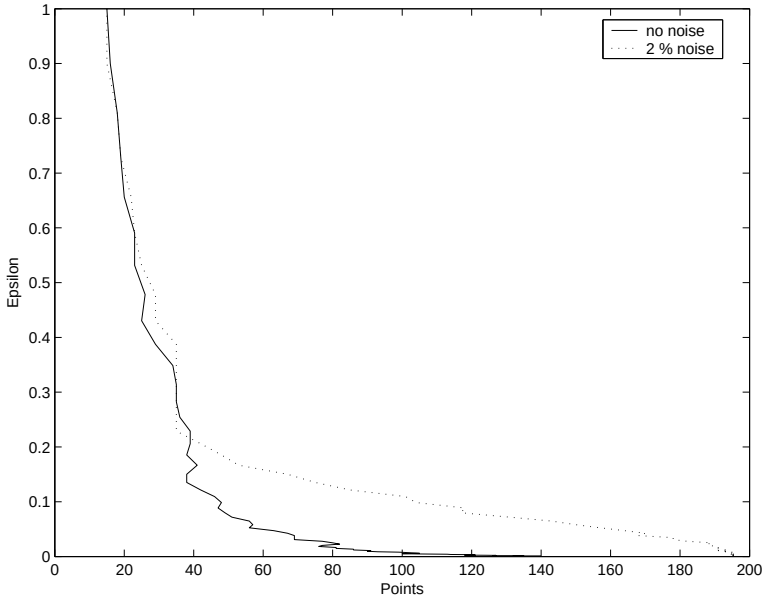

Figure 20: Connection between ϵ and the number $n(\epsilon)$ of necessary points

Let us use minimization of the quadratic form (5.6) to provide an example for the tradeoff between error and complexity. Again, the basic situation is interpolation to the MATLAB `peaks` function, this time in 14×14=196 regularly distributed points in $[-3,3]^2$ by Gaussians of scale 1. The global $L_\infty[-3,3]^2$ error of the exact interpolation on these data is 0.024, evaluated on a fine grid with 121×121=14641 points. But now we minimize the quadratic form (5.6) under the constraints

$$-\epsilon \le \sum_{j=1}^{n} \alpha_j \phi(\|x_j - x_k\|_2) - f(x_k) \le \epsilon,\ 1 \le k \le n \tag{5.7}$$

for positive ϵ. The case of $\epsilon = 0$ is exact interpolation using all 196 data points and trial functions. For positive ϵ, the usual Karush-Kuhn-Tucker conditions imply that only those points x_k are actually used where one of the bounds in (5.7) is attained with equality. The number $n(\epsilon)$ of required points grows up to the maximally possible $n(0) = 196$ when ϵ decreases. Figure 20 shows this for the case of exact and noisy data.

But even more interesting is the behavior of the global $L_\infty[-3,3]^2$ error $E(\epsilon)$ as a function of ϵ. Figure 21 shows that $E(\epsilon)$ roughly follows the behavior of ϵ when plotted as a function of the required points $n(\epsilon)$. Both curves are experimentally available, and one can read off that the optimal choice of ϵ in the noisy case is at the point where the curve takes its L-turn, i.e. at the point of largest curvature around $n = 40$. This can be viewed as an experimental method to determine the noise level. Note that it does not pay off to use more points, and note the similarity to the L-curve technique [30].

But also for exact data, these curves are useful. Since the maximum value of the `peaks` function is about 8.17, one can get a relative global accuracy of 1% using roughly 60 points for exact data. It makes no sense to use the full 196 points, even for exact data, if exact results are not required. Of course, larger noise levels lead to smaller numbers of required points, but a thorough investigation of these tradeoff effects between error and complexity is still a challenging research topic.

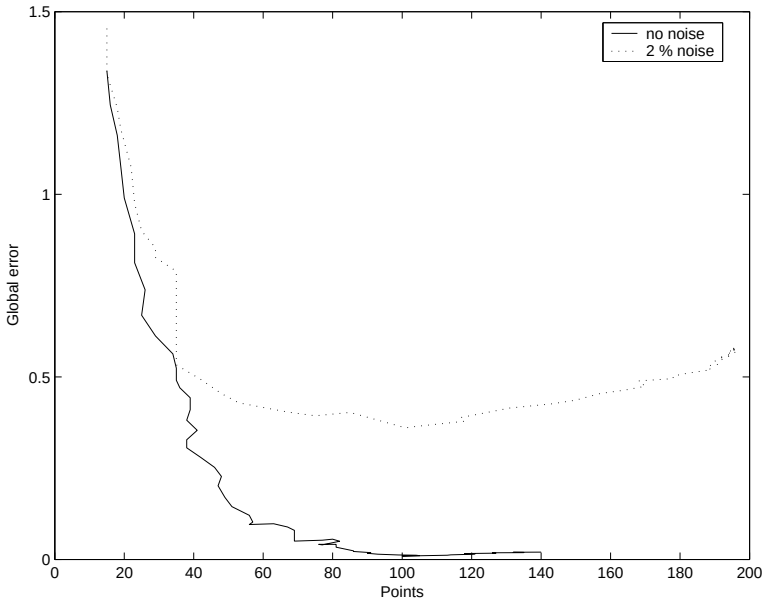

Figure 21: Error $E(\epsilon)$ as a function of the number $n(\epsilon)$ of necessary points

.... incomplete...

Demos on power functions and on point selection

.... incomplete...

6 Error Analysis

This section is the core for any error analysis of interpolation or approximation methods. There are essentially two possible approaches:

1. via upper bounds on the power function and
2. via "sampling inequalities".

These share some common tools. e.g. the concept of "norming sets" and local polynomial approximation. The second alternative is more modern, and thus I chose it. But, as the other one, it is hardcore mathematics and it would take too much time to present a complete rigid proof including all the details. Thus I shall focus on motivating and explaining the important ingredients, but I shall skip over tedious and detailed calculations. In order to start with an easy case, we shall focus on univariate sampling inequalities first.

6.1 Sampling Inequalities

The basic idea of sampling inequalities is the following.

Assume that a function s approximates or interpolates a function f on a discrete subset X of its domain Ω. Then $f - s$ is small or even zero on X. How large can the **error function** $f - s$

be outside of X? If, for instance, any directional derivative of **both** f and s is bounded above by some constant K, we can write

$$|f(x) - s(x)| \leq |f(x_j) - s(x_j)| + 2K \cdot \|x - x_j\|_2$$

if the line connecting x and $x_j \in X$ is in Ω and if we integrate the directional derivative along the line. If we define the **fill distance**

$$h := h(X, \Omega) := \sup_{x \in \Omega} \min_{x_j \in X} \|x - x_j\|_2$$

of X in Ω and if Ω is convex, this yields the simple error bound

$$\|f - s\|_{\infty,\Omega} \leq 2K \cdot h,$$

but we need to have K under control, i.e. in terms of some tricky high–order norm $\|f\|$ of f.

Here is a more general description of the above argument, applied to the error function:

> If a smooth function has a bound on its highest derivatives, and if it is small on a large set of points which "fills" the domain, then it should be small everywhere.

We can apply such a result nicely in our kernel–based methods, because if we have a function f in the native space $\mathcal{N}$, we know that its interpolant $s_{f,X}$ on a set X minimizes the native space (semi–) norm under all other interpolants, and thus $|s_{f,X}|_{\mathcal{N}} \leq |f|_{\mathcal{N}}$ will hold in the native space (semi–) norm. Thus we should be able to infer that $f - s_{f,X}$ is small if the fill distance h of X is small, i.e. we want something like

$$\|f - s_{f,X}\|_{\infty,\Omega} \leq F(h(X, \Omega))|f|_{\mathcal{N}} \tag{6.1}$$

with $F(h) \to 0$ for $h \to 0$. This works along the lines of Theorem (3.10) and Corollary (4.33) if we can bound the power function by $F(h(X, \Omega))$. For about two decades, this technique was the standard way of obtaining error bounds for kernel–based interpolants.

However, a more general bound like

$$\|f\|_{\infty,\Omega} \leq F(h(X, \Omega))|f|_{\mathcal{N}} + C \cdot \|f\|_{\infty,X} \tag{6.2}$$

holding for all f in the native space $\mathcal{N}$ would imply (6.1) with a slightly larger constant, if it is applied to $f - s_{f,X}$ via

$$\begin{aligned} \|f - s_{f,X}\|_{\infty,\Omega} &\leq F(h(X, \Omega))|f - s_{f,X}|_{\mathcal{N}} + C \cdot \|f - s_{f,X}\|_{\infty,X} \\ &\leq F(h(X, \Omega))|f - s_{f,X}|_{\mathcal{N}} \\ &\leq F(h(X, \Omega))(|f|_{\mathcal{N}} + |s_{f,X}|_{\mathcal{N}}) \\ &\leq 2F(h(X, \Omega))|f|_{\mathcal{N}}. \end{aligned}$$

We shall thus go for (6.2) directly, but we shall allow other norms than the ∞–norms, and we shall consider general spaces instead of native spaces for kernels.

Note that (6.2) is of crucial importance for all kinds of discretization processes. If a function f is discretized by taking values in X, users must often know that it does not get large between the data sites. This is exactly provided by (6.2) in case of $F(h) \to 0$ for $h \to 0$.

6.2 Univariate Case

Since the forthcoming mathematical analysis is quite hard, let us first start with the much simpler univariate case. We shall work on a bounded interval $I := [a,b] \subset I\!R$, and there we shall define the (semi–) inner products and (semi–) norms

$$\begin{aligned}
(f,g)_j &:= \int_I f^{(j)}(t) g^{(j)}(t) dt \text{ for all } f,g \in C^j(I) \\
|f|_{\infty,X} &:= \sup_{t\in X} |f(t)| \text{ for all } f \in C(I),\ X \subset I, \\
|f|^2_{2,X} &:= h(X,\Omega) \sum_{x\in X} f^2(x) \text{ for all } f \in C(I),\ X \subset I,\ |X| < \infty.
\end{aligned}$$

Throughout, we shall confine ourselves to subsets $X \subset [a,b]$ with fill distance $h = h(X,[a,b])$.

Lemma 6.3 *Given $g \in C^1(I)$ and a subset X of $[a,b]$ with fill distance h. Then we have*

$$\begin{array}{rcrl}
\|g\|_0 & \leq & h\ |g|_1 & +\sqrt{2}|g|_{2,X}, \\
\|g\|_0 & \leq & h\ |g|_1 & +\sqrt{2(b-a)}|g|_{\infty,X}, \\
\|g\|_{\infty,I} & \leq & h\ \|g'\|_{\infty,I} & +|g|_{\infty,X}, \\
\|\ g\|_{\infty,I} & \leq & \sqrt{h}\ |g|_1 & +|g|_{\infty,X}.
\end{array} \tag{6.4}$$

Proof: For all $x \in I$ we can take its closest neighbor $x_j \in X$ with distance $|x - x_j| \leq h$. Then

$$\begin{aligned}
g(x) &= g(x_j) + \int_{x_j}^x g'(\tau) d\tau \\
|g(x)| &\leq |g(x_j)| + \int_{x_j}^x |g'(\tau)| d\tau \\
\|g\|_{\infty,I} &\leq h\ \|g'\|_{\infty,I} + |g|_{\infty,X}
\end{aligned}$$

for all $x \in I$. With the Cauchy–Schwarz inequality we we get

$$\begin{aligned}
|g(x)| &\leq |g(x_j)| + \left| \int_{x_j}^x 1^2 d\tau \right|^{1/2} \cdot \left| \int_{x_j}^x (g'(\tau))^2 d\tau \right|^{1/2} \qquad (x \in I) \\
&\leq |g(x_j)| + |x - x_j|^{1/2} \cdot \left| \int_{x_j}^x (g'(\tau))^2 d\tau \right|^{1/2}, \\
\|\ g\|_{\infty,I} &\leq \sqrt{h}\ |g|_1 + |g|_{\infty,X}.
\end{aligned}$$

By taking squares and the usual trick

$$(a+b)^2 \leq a^2 + b^2 + 2|ab| \leq 2a^2 + 2b^2$$

we see that

$$g(x)^2 \leq 2g(x_j)^2 + 2|x - x_j| \cdot \int_{x_j}^x (g'(\tau))^2 d\tau$$

and by integration we find

$$\begin{aligned}
\textstyle\int_{x_j}^x g(t)^2 dt &\leq 2|x - x_j| g(x_j)^2 + \int_{x_j}^x 2|t - x_j| \cdot \int_{x_j}^t (g'(\tau))^2 d\tau dt \\
&= 2|x - x_j| g(x_j)^2 + \int_{x_j}^x (g'(\tau))^2 \int_x^\tau 2|t - x_j| dt d\tau \\
&\leq 2|x - x_j| g(x_j)^2 + |x - x_j|^2 \cdot \int_{x_j}^x (g'(\tau))^2 d\tau.
\end{aligned}$$

From here on we superimpose different integrals of this form to a full integral over $[a,b]$ to end up with

$$\begin{aligned} \|g\|_0^2 &\leq h^2\ |g|_1^2 + 2(b-a)|g|_{\infty,X}^2 \\ \|g\|_0 &\leq h\ |g|_1 + \sqrt{2(b-a)}|g|_{\infty,X} \end{aligned}$$

where we used $\sqrt{a^2+b^2} \leq a+b$ for $a,b>0$. If we sum up the discrete values, we arrive at

$$\begin{aligned} \|g\|_0^2 &\leq h^2\ |g|_1^2 + 2h\sum_{j=1}^{N} g(x_j)^2 \\ \|g\|_0 &\leq h\ |g|_1 + \sqrt{2}|g|_{2,X}. \end{aligned}$$

□

6.3 Example: Univariate Splines

In the notation of the text on splines, we have

Theorem 6.5 *Let $f \in \mathcal{C}^k[a,b]$ be interpolated by s^* in $N \geq k$ data with a* fill distance

$$h := \sup_{x\in[a,b]} \min_{x_j} |x - x_j|.$$

Then there is a constant c_k depending only on k and $[a,b]$, but not on f or the data or h, such that

$$\begin{aligned} \|f-s^*\|_{L_2[a,b]} &\leq c_k h^k |f-s^*|_k \leq 2c_k h^k |f|_k, \\ \|f-s^*\|_{L_\infty[a,b]} &\leq c_k h^{k-1/2} |f-s^*|_k \leq 2c_k h^{k-1/2} |f|_k. \end{aligned}$$

Proof: Note that the zeros of $f-s^*$ have a distance of at most $2h$ between each other and of at most $h \leq 2h$ to the boundary. By Rolle's theorem, there are zeros of $(f-s^*)'$ with distance of at most $4h$ between each other and $3h$ to the boundary. This means that we can use the fill distance $4h$ for the zeros of the first derivative. This works up to the derivative of order $k-1$, which has zeros with distance of at most $4^{k-1}h$ between each other and to the boundary. Using induction on the previous Lemma yields

$$\|f-s^*\|_{L_2[a,b]} \leq h \cdot 4h \cdots 4^{k-1}h |f-s^*|_k =: c_k h^k |f-s^*|_k$$

and the left–hand parts of the assertions follow.

For the right-hand parts we use the optimality condition $|s^*|_k \leq |f|_k$. □

If some additional boundary conditions are satisfied, the convergence order doubles.

Theorem 6.6 *If, in addition, $f \in \mathcal{C}^{2k}[a,b]$ and if $(f-s^*)^{(j)}$ vanishes at a and b for $j = 0,\ldots,k-1$, then*

$$\begin{aligned} \|f-s^*\|_{L_2[a,b]} &\leq c_k^2 h^{2k} |f|_{2k}. \\ \|f-s^*\|_{L_\infty[a,b]} &\leq \tilde{c}_k^2 h^{2k-1} |f|_{2k}. \end{aligned}$$

Proof: We can use the orthogonality relation

$$(f-s^*, s^*)_k = 0$$

and do integration by parts via

$$\begin{aligned} |f-s^*|_k^2 &= (f-s^*, f-s^*)_k \\ &= (f-s^*, f)_k \\ &= \int_a^b (f-s^*)^{(k)}(t) f^{(k)}(t) dt \\ &= (-1)^k \int_a^b (f-s^*)^{(0)}(t) f^{(2k)}(t) dt \\ &\leq \|f-s^*\|_{L_2[a,b]} |f|_{2k}. \end{aligned}$$

Then

$$\begin{aligned} \|f-s^*\|_{L_2[a,b]}^2 &\leq c_k^2 h^{2k} |f-s^*|_k^2 \\ &\leq c_k^2 h^{2k} \|f-s^*\|_{L_2[a,b]} |f|_{2k} \\ \|f-s^*\|_{L_2[a,b]} &\leq c_k^2 h^{2k} |f|_{2k}. \end{aligned}$$

Similarly,

$$\begin{aligned} \|f-s^*\|_{L_\infty[a,b]}^2 &\leq c_k^2 h^{2k-1} |f-s^*|_k^2 \\ &\leq c_k^2 h^{2k-1} \|f-s^*\|_{L_2[a,b]} |f|_{2k} \\ &\leq \sqrt{b-a} c_k^2 h^{2k-1} \|f-s^*\|_{L_\infty[a,b]} |f|_{2k} \\ \|f-s^*\|_{L_\infty[a,b]} &\leq \tilde{c}_k^2 h^{2k-1} |f|_{2k}. \end{aligned}$$

Note that the above argument used Rolle's theorem, which does not hold in multivariate settings. Thus we cannot generalize this approach directly to functions of several variables.

6.4 Univariate Polynomial Reproduction

Our goal is to prove a multivariate version of a sampling inequality. But already in the univariate case, a general inequality like

$$\|u\|_{\infty,[a,b]} \leq C \left(h^{k-1/2} |u|_k + |u|_{\infty,X} \right)$$

means that for all polynomials $p \in I\!P_k$ we have

$$\|p\|_{\infty,[a,b]} \leq C |p|_{\infty,X}. \tag{6.7}$$

Then X must be unisolvent, but this is not enough. If we take X to contain exactly k points, an equality of the above form cannot hold. To see this, fix $k-1$ zeros and prescribe 1 at a point which moves close to a zero. The resulting Lagrange basis polynomial will converge to infinity except at the zeros.

But if we take many more than k points, i.e. we do some **oversampling**, chances are better to get something like (6.7). If we extend (6.7) trivially to the right, we get

$$\|p\|_{\infty,[a,b]} \leq C |p|_{\infty,X} \leq C \|p\|_{\infty,[a,b]}$$

and see that X must guarantee norm equivalence of a discrete norm with a "continuous" norm. We could call X a "norming set", but there is a more general definition of that notion, and we provide it later.

To move closer to (6.7), let us fix a polynomial $p \in I\!P_k$ with $\|p\|_{\infty,[a,b]} = 1$. We need to show that such a polynomial cannot be too small on a nontrivial set X, but we want to get away with a smallest possible set X. For simplicity, we take a $t \in [a,b]$ with $|p(t)| = 1$ and ask: How far must we go to let $|p|$ drop below $1/2$? Fortunately, we have a bound on the derivative:

Theorem 6.8 *Any univariate polynomial of degree n satisfies* **Markov's inequality**

$$\|p'\|_{\infty,[-1,1]} \leq n^2 \|p\|_{\infty,[-1,1]}.$$

We skip over the proof, but by norm equivalence there must be an n–dependent constant that does the job. The only problem is to prove that the constant is n^2.

Now we know that our special polynomial has a derivative $\|p'\|_{\infty,[-1,1]} \leq n^2$ if we assume that the interval is $[-1,1]$. Thus, in order to let p go down to $1/2$ we need to go at least a distance $1/2n^2$. If we know that X has a fill distance

$$h \leq \frac{1}{2n^2}$$

we can be sure that we cannot reach a point with absolute value of p smaller than $1/2$ when starting from t. This means that

$$|p|_{\infty,X} \geq \frac{1}{2}$$

and consequently $\|p\|_{\infty,[-1,1]} \leq 2|p|_{\infty,X}$. Thus we have

Theorem 6.9 *If $X \subset [-1,1]$ is a set of fill distance $h \leq \frac{1}{2n^2}$, then*

$$\|p\|_{\infty,[-1,1]} \leq 2|p|_{\infty,X}$$

for all polynomials of degree at most n. □

From now on we assume that the hypothesis of Theorem 6.9 is satisfied. Then X clearly is unisolvent, and we know that we can reproduce all polynomials $p \in I\!P_n$ by a nonunique formula like

$$p(x) = \sum_{x_j \in X} u_j(x) p(x_j). \tag{6.10}$$

In our old notation, this is an under-determined linear system

$$P_X \cdot u(x) = p(x)$$

and one can impose additional conditions. In fact, there are efficient numerical techniques (e.g. **moving least squares** which produce useful admissible solutions.

Theorem 6.11 *Under all possibilities to satisfy (6.10) under the conditions of Theorem 6.9 there is one which has a uniformly bounded* **Lebesgue function**

$$\sum_{x_j \in X} |u_j(x)| \leq 2 \text{ for all } x \in [-1,1].$$

Proof: This statement is not at all evident, and for now we have to do an abstract existence proof. We define the **sampling operator**

$$T_X \ : \ f \mapsto (f(x_1), \ldots, f(x_N))^T \in I\!R^N,$$

which is continuous on $C[-1,1]$ with the $\|.\|_\infty$ norm and invertible on $T(I\!P_n) \subset I\!R^N$. It has a bounded inverse

$$S \ : \ T(I\!P_n) \to I\!P_n \subset C[-1,1].$$

For each vector $y \in T(I\!P_n) \subset I\!R^N$ there is a unique $p \in I\!P_n$ with such $y = (p(x_1), \ldots, p(x_N))^T$, and thus for each $x \in [-1,1]$ we have a linear functional

$$\lambda_x \ : \ y = (p(x_1), \ldots, p(x_N))^T \mapsto p(x)$$

on $T(I\!P_n)$. By the **Hahn–Banach theorem** it has a **norm–preserving** extension to all of $I\!R^N$, and this is the abstract and non-constructive part of the argument. As a functional on all of $I\!R^N$ it can be written as

$$\lambda_x(y) := \sum_{j=1}^{N} u_j(x) y_j$$

with certain real values $u_j(x)$, and its norm must be

$$\|\lambda_x\|_1 := \sum_{x_j \in X} |u_j(x)| = \sup_{y \in I\!R^N \setminus \{0\}} \frac{\left|\sum_{x_j \in X} u_j(x) y_j\right|}{\|y\|_\infty}$$

because the dual of $I\!R^N$ under the $\|.\|_\infty$ norm is $I\!R^N$ with the L_1 norm. But since the extension is norm–preserving, this norm is equal to the norm of the functional on the subspace $T(I\!P_n)$. There it has the form $\lambda_x = \delta_x \circ S$, and thus

$$\|\lambda_x\|_1 \leq \|\delta_x\| \cdot \|S\|$$

with operator norms

$$\|\delta_x\| = \sup_{f \in C[-1,1] \setminus \{0\}} \frac{|f(x)}{\|f\|_\infty} \leq 1$$

and

$$\|S\| := \sup_{T(p) \in T(I\!P_n) \setminus \{0\}} \frac{|p(x)|}{\|T(p)\|_\infty} \leq 2.$$

Thus we get the assertion. □

It is an easy task to find functions $u_j(x)$ for which the Lebesgue function is pointwise minimal, while a certain polynomial reproduction of a degree n is required. It boils down to an L_1 optimization problem, because we can write each $u_j(x)$ as $u_j(x) = a_j - b_j$ with nonnegative variables to arrive at the linear optimization problem

$$\text{Minimize } 1_N^T(a+b) \text{ under } P_X(a-b) = p(x)$$

in normal form with nonnegative variables $a, b \in I\!R^N$. This can be solved pointwise, but the solution is rather strange, see figures 23 and 22. The theory of linear optimization implies that at a certain point x there can be at most $2(n+1)$ nonzero $u_j(x)$, but the functions u_j are not localized around x_j, as Figure 24 shows.

Thus it is an additional problem to find a stable solution which is **localized** in the sense that $u_j(x)$ vanishes if x is "far" from x_j. But this can be done by localizing the above argument. However, we shall not do this at this point. Instead, we start with the general multivariate case and care for both localization and a bounded Lebesgue function.

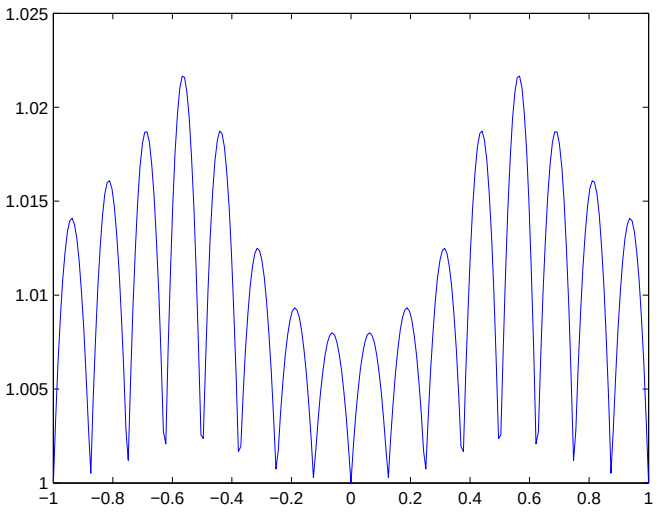

Figure 22: Minimal Lebesgue function for 17 equidistant points in $[-1,1]$ and polynomial degree $n=3$

6.5 Norming Sets

As a little digression, we generalize the above construction, following an idea of Jetter, Stöckler, and Ward.

... incomplete here.....

6.6 Multivariate Polynomial Reproduction

In the multivariate setting, we should go for sampling inequalities of the form

$$\begin{aligned} \|u\|_{W_2^m(\Omega)} &\leq C\left(h^{M-m}|u|_{W_2^M(\Omega)} + h^{-m}\|u\|_{\infty,X_h}\right) \\ \|D^\alpha u\|_{L_\infty(\Omega)} &\leq C\left(h^{M-|\alpha|-d/2}|u|_{W_2^M(\Omega)} + h^{-m}\|u\|_{\infty,X_h}\right) \end{aligned} \tag{6.12}$$

for all functions in the Sobolev space $W_2^M(\Omega)$ with the inner product

$$\begin{aligned} (f,g)_{W_2^M(\Omega)} &:= \sum_{j=0}^{M}(f,g)_j \\ (f,g)_j &:= \sum_{|\alpha|=j}\int_\Omega D^\alpha f D^\alpha g \end{aligned}$$

and (semi–)norms

$$\begin{aligned} \|f\|^2_{W_2^M(\Omega)} &:= (f,f)_{W_2^M(\Omega)} \\ |f|^2_{W_2^M(\Omega)} &:= (f,f)_M = \sum_{|\alpha|=j}\int_\Omega |D^\alpha f|^2 \end{aligned}$$

where we use standard multivariate notation. If u is a polynomial of order at most M, then $|u|^2_{W^M(\Omega)} = |u|_M = 0$ and our equations take a special form

$$\begin{aligned} \|p\|_{L_2(\Omega)} &\leq C\|p\|_{\infty,X_h} \\ \|p\|_{L_\infty(\Omega)} &\leq C\|p\|_{\infty,X_h} \end{aligned}$$

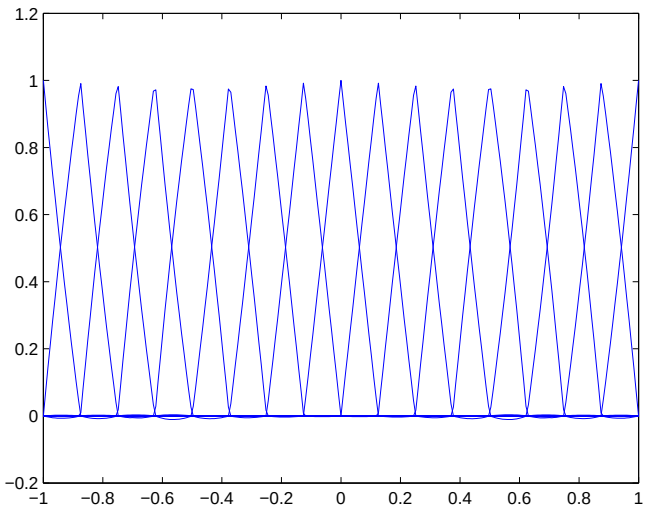

Figure 23: Functions u_j for the same case as in the previous figure

for all $p \in I\!P_M$, i.e. there is stable polynomial reproduction in the sense of (6.7). But we also want this reproduction to be local and it should be guaranteed via Lagrange–type functions u_j. We thus formulate the following goal:

Definition 6.13 *Let $\Omega \subset I\!R^d$ be a domain and fix a number $k \in I\!N$. If there are positive numbers h_0, c_1, c_2 depending on k and Ω such that for each finite subset $X = \{x_1, \ldots, x_N\} \subset \Omega$ with fill distance $h(X, \Omega) \le h_0$ there are functions $u_1^X, \ldots, u_N^X$ on Ω such that*

$$\begin{array}{rcll} \sum_{j=1}^{N} u_j^X(x) p(x_j) & = & p(x) & \text{for all } p \in I\!P_k,\ x \in \Omega \\ \sum_{j=1}^{N} |u_j^X(x)| & \le & c_1 & \text{for all } x \in \Omega \\ \|x - x_j\|_2 > c_2 h & \Rightarrow & u_j^X(x) = 0 & \text{for all } x \in \Omega,\ 1 \le j \le N \end{array} \tag{6.14}$$

then we say that the quasi–interpolation process

$$f \mapsto Q_X(f)(\cdot) := \sum_{j=1}^{N} u_j^X(\cdot) f(x_j) \text{ for all } f : \Omega \to I\!R$$

defined for all such X provides **stable local polynomial reproduction** *of order k.*

Then we have a rather simple local error bound:

Theorem 6.15 *Assume that $\Omega \subset I\!R^d$ is bounded and admits stable local polynomial reproduction of order k. Then there is a constant C with*

$$|f(x) - Q_X(f)(x)| \le C h^{k+1} |f|_{k+1,\Omega^*}$$

for all $x \in \Omega$ and all $f \in C^{k+1}(\Omega^*)$ on the extended domain

$$\Omega^* := \bigcup_{x \in \Omega} \overline{B(x, c_2 h_0)}$$

where $B(x, r)$ is the open ball with center at x and radius r.

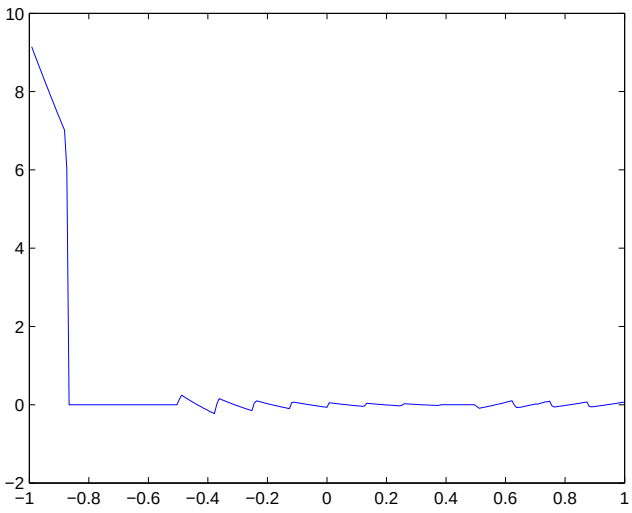

Figure 24: Derivative of u_1 for the same case as in the previous figure

Proof: See Wendland's book [71], p. 25/26. The basic trick is to introduce the local Taylor expansion at x which has the same error bound and is reproduced by the quasi–interpolant. Then use the boundedness of the Lebesgue function to show that the error of the local Taylor expansion carries over to the quasi–interpolant. □.

Inspection of the proof shows that the extended domain is not necessary for all forms of local bounds.

To proceed towards an existence proof of stable local polynomial reproduction on nondegenerate domains, we repeat our argument from the univariate case, but we can focus on $k > 1$ because stable polynomial reproduction by constants is trivial, using the nearest neighbor of X to each $x \in \Omega$. Another choice for stable local approximation of order one is **Shepard approximation**, which we shall describe in the next section.

It does not suffice to use $I\!P_k$–unisolvent subsets, because they will not have bounded Lebesgue functions. We thus first aim at a proof of an inequality like

$$\|p\|_{\infty,\Omega} \leq c_2 \|p\|_{\infty,X} \text{ for all } p \in I\!P_k$$

for suitable domains Ω and finite sets $X \subset \Omega$. We start with a polynomial $p \in I\!P_k$ with $p(x) = \|p\|_{\infty,\Omega} = 1$ and see how fast it can go down when moving away from x. On a ray going from x to some other point z, the polynomial

$$q(t) := p\left(x + t(z - x)\right),\ t \in [0, 1]$$

is univariate and of order at most k. We have

$$|p\left(x + t(z - x)\right) - p(x)| = |q(t) - q(0)| = \left| \int_0^t q'(t)dt \right| \leq t\|q'\|_{\infty,[-1,1]}$$

and use Markov's inequality to get

$$|p\left(x + t(z - x)\right) - p(x)| \leq t(k-1)^2 \|q\|_{\infty,[-1,1]} \leq t(k-1)^2 \|p\|_{\infty,\Omega}$$

for all $t \in [0,1]$. But for our choice of $1 = p(x) = \|p\|_{\infty,\Omega}$ we get

$$p\left(\underbrace{x + t(z-x)}_{=:y_t}\right) = q(t) \geq 1 - 2t(k-1)^2$$

leading to $p(y_t) \geq 1/2$ for all y_t on the ray with $t = \|x - y_t\|_2 \leq \frac{1}{4(k-1)^2}$. For all finite sets X with

$$\min_{z \in X} \|x - z\|_2 \leq \frac{1}{4(k-1)^2}$$

this implies

$$|p|_{\infty,X} \geq \frac{1}{2} = \frac{1}{2}p(x).$$

This argument needs that the ray from x to z is contained in the domain we want to focus on.

Definition 6.16 *A set $\Omega \subset \mathbb{R}^d$ is* **convex***, if for all points $x, y \in \Omega$ the line*

$$[x,y] := \{\lambda x + (1-\lambda)y \ : \ \lambda \in [0,1]\}$$

consisting of all **convex combinations** *of x and y belongs to Ω.*

Thus the argument works for all closed bounded convex sets Ω and we have

Theorem 6.17 *If X is a finite subset with fill distance*

$$h \leq \frac{1}{4(k-1)^2} =: h_1$$

in a closed bounded convex set $\Omega \subset \mathbb{R}^d$, then the inequality

$$\|p\|_{\infty,\Omega} \leq 2\|p\|_{\infty,X}$$

holds for all polynomials $p \in \mathbb{P}_k$. □

Note that this result is independent of the size of Ω, but it requires convexity, because we need the rays from any point $x \in \Omega$ to any point $z \in X$.

To proceed towards the argument providing the functions u_j^X, we look at the **sampling operator**

$$T_X(f) := (f(x_1), \ldots, f(x_N))^T$$

mapping functions on Ω into $\mathbb{R}^N$ for each set $X = \{x_1, \ldots, x_N\} \subset \Omega$. Under the assumptions of Theorem 6.17, this map is injective on $V := \mathbb{P}_k$, and we can proceed exactly as in the univariate case to get

Theorem 6.18 *If X is a finite subset with fill distance*

$$h \leq \frac{1}{4(k-1)^2} =: h_1$$

in a closed bounded convex set $\Omega \subset \mathbb{R}^d$, then there are functions u_j^X on Ω that realize stable polynomial reproduction in the sense of the first two equations of (6.14) with $c_2 = 2$. □

The remaining problem is localization together with elimination of convexity. But the problem is that, for instance with gridded data, there may be convex subdomains which contain no point of X at all, and a fortiori there are convex subdomains where a set X with fill distance h_1 with respect to Ω has a local fill distance larger than h_1. We thus have to focus on domains where we have positive constants $h_0 \leq h_1$ and c_1 such that each point of Ω belongs to a convex subset $\tilde{\Omega}$ of Ω in which any discrete set with fill distance $h \leq h_0$ with respect to Ω still has a fill distance $\tilde{h} \leq c_1 h_0 \leq h_1$ with respect to $\tilde{\Omega}$. In this case, given a point x, we just work on the subdomain $\tilde{\Omega}$ containing x and restrict ourselves to points in $X \cap \tilde{\Omega}$ to prove (6.14). This will turn out to work for domains satisfying

Definition 6.19 *A domain $\Omega \subset I\!R^d$ satisfies an* **interior cone condition** *with angle α and radius r such that for each $x \in \Omega$ there is a normalized* **cone axis** *z_x with $\|z_x\|_2 = 1$ such that the cone*

$$\{x + \lambda y \ : \ \lambda \in [0, r], \ \|y\|_2 = 1, y^T z_x \geq \cos(\alpha)\}$$

of height $r > 0$ and opening angle $2\alpha > 0$ is still contained in Ω.

... missing picture...

We restrict ourselves to domains with $0 < \alpha < \pi$ and $r \leq 1$ in order to avoid difficulties. In fact, if a domain satisfies a cone condition with angle $\alpha > 0$ and $r > 0$, it satisfies a condition also for all smaller positive α and r. We shall cover the domain by cones of the above form, and thus we only have to prove that such cones have the property we mentioned, i.e. any discrete set with fill distance $h \leq h_0$ with respect to Ω still has a fill distance $\tilde{h} \leq c_1 h_0 \leq h_1$ with respect to such a cone, where we can define h_0 and c_1 in terms of α and r.

In such a cone, the point x has distance $z = \frac{r}{1+\sin\alpha}$ from a ball of radius $\frac{r\sin\alpha}{1+\sin\alpha}$ which still is in the cone.

... missing picture...

If

$$h_0 \leq \frac{r\sin\alpha}{1+\sin\alpha}$$

we have at least one point of X in the ball. Since the maximal distance of x to this point is r, we get that X has fill distance at most

$$r \geq \frac{1+\sin\alpha}{\sin\alpha} h_0$$

with respect to that cone. But we have to aim at a fill distance $\tilde{h} \leq c_1 h_0 \leq h_1$, and we can get away with

$$\begin{aligned} c_1 &:= \frac{1+\sin\alpha}{\sin\alpha} \geq 1 \\ h_0 &:= \frac{1}{c_1}\min(r, h_1). \end{aligned}$$

In fact, this implies $c_1 h_0 \leq h_1$, and any set X with fill distance $h \leq h_0$ with respect to Ω will have a fill distance at most $c_1 h_0$ with respect to any of our cones, because we can use $\tilde{r} = c_1 h_0 \leq r$ in the interior cone condition.

Theorem 6.20 *In bounded domains in R^d with interior cone condition with angle $0 < \alpha < \pi$ and radius $r \leq 1$ there is stable local polynomial reproduction of order k with the constants given above.* □

6.7 Moving Least Squares

The above theory does not provide a practical way to construct functions u_j^X with the required properties. But there is a constructive way to generate stable local polynomial reproduction constructively.

The easiest case is **Shepard approximation**. Take a nonnegative nontrivial kernel in translation–invariant form, e.g. $K(x,y) = \Phi(x-y)$ with $\Phi : \mathbb{R}^d \to \mathbb{R}$ and consider the function

$$S_X(f)(x) := \sum_{x_j \in X} f(x_j) \frac{\Phi(x-x_j)}{\sum_{x_k \in X} \Phi(x-x_k)}$$

for any finite set X and any function f. If the quotient is undefined for certain exceptional cases, e.g. when the support of Φ is very small and the set X has large fill distance, the quotient can be defined to be zero. This approximant preserves constant, because it uses a **partition of unity**, i.e. a set of nonnegative functions that sum up to one. It clearly is stable with Lebesgue function bounded by one, and it can be made local if the support of the kernel is proportional to the fill distance h of the set X.

Motivated by this case, we introduce a scaling into the kernel by defining

$$\Phi_\delta(x) := \Phi(x/\delta) \text{ for all } x \in \mathbb{R}^d,\ \delta > 0$$

and take a nonnegative kernel with precise support in the unit ball $B(0,1)$, i.e.

$$\Phi(x) = 0 \text{ for } x \in \mathbb{R}^d \text{ if and only if } \|x\|_2 \geq 1.$$

Then the kernel Φ_δ has support in the ball $B(0,\delta)$ with center 0 and radius δ.

When constructing an approximation at some point x based on function values $f(x_j)$ at certain points x_j of a finite set X, we use the kernel as a **weight function** to let the points $x_j \in X$ closer to x have more importance than those further away. If Φ is smooth, there is some hope that the resulting function of x can be defined to be continuous or even differentiable, because the weights of the used points are varying smoothly. Furthermore, one can hope to get a fully local method, if the support radius δ is scaled like $\delta = c_1 h$ and thus connected to the fill distance h of X. Finally, polynomial reproduction should be built into the method by adding equations like (6.10) as constraints and allowing enough oversampling to let them be satisfied up to a certain order.

Put together, all of this suggests the following pointwise definition of a **moving least–squares** approximation:

$$\text{Minimize} \sum_{x_j \in X_{x,\delta}} (f(x_j) - p(x_j))^2 \Phi_\delta(x-x_j) \text{ over all } p \in \mathbb{P}_k$$

for fixed x and sets $X := \{x_1, \ldots, x_N\}$, and call the resulting function value $MLS(x) := MLS(f, X\Phi_\delta)(x) := p^*(x)$ when p^* is the optimal polynomial. Note that we restricted the sum in the objective function to

$$\begin{array}{rcl} X_{x,\delta} & := & \{x_k \in X \ : \ \|x-x_k\|_2 < \delta\} \\ J_{x,\delta} & := & \{j \ : \ x_j \in X_{x,\delta}\} \end{array} \tag{6.21}$$

describing the points of X close enough to x to enter into the calculation.

For convenience, we use the shorthand notation $J := J_{x,\delta}$, and we should rewrite the problem in terms of a polynomial basis and in matrix–vector form. Then it is

$$\text{Minimize } \|\sqrt{W}(F - Pb)\|_2^2 \text{ over } b \in I\!R^Q$$

where

$$\begin{aligned} F &:= (f(x_j),\ j \in J)^T \in I\!R^{|X_{x,\delta}|} \\ P &:= (p_\ell(x_j))_{j\in J,\ 1\le\ell\le Q} \\ b &:= (b_1, \ldots, b_Q)^T \in I\!R^Q \\ W &:= (\delta_{jk}\Phi_\delta(x - x_k))_{j,k\in J} \end{aligned}$$

and it is a standard least–squares problem approximationg $\sqrt{W}f_X$ by $\sqrt{W}P_X b$ for $b \in I\!R^Q$. In theory, the solution satisfies the **normal equations**

$$P_X^T W F = P_X^T W P_X b$$

and is unique if the rank of $P_X^T W P_X$ is Q. This requires $Q \le |X_{x,\delta}| \le N$ and $I\!P_k$–unisolvence of $X_{x,\delta}$. With this assumption, we can calculate the unique solution of the problem by standard least–squares techniques, but we have no information about stability or polynomial reproduction. To this end, one can try to rewrite the problem as one in the localized quasi–interpolant form

$$MLS(x) = \sum_{j\in J} a_j^*(x) f(x_j)$$

where we already inserted the optimal values $a_j^*(x)$ for fixed x without knowing how to set up an equivalent optimization problem. Anyway, we should impose the polynomial reproduction constraints

$$p(x) = \sum_{j\in J} a_j(x) p(x_j) \text{ for all } p \in I\!P_k \tag{6.22}$$

restricting the variables $a_j(x)$ for $j \in J$. But we still have nothing to optimize. Clearly, we should make sure that $a_j^*(x)$ gets small if x_j is just about to leave the influence region for x, i.e. if $\|x - x_j\|_2$ is close to δ, or if $\Phi_\delta(x - x_j)$ is small. This suggests to minimize

$$\sum_{j\in J} a_j^2(x) \frac{1}{\Phi_\delta(x - x_j)}. \tag{6.23}$$

Theorem 6.24 *If the set $X_{x,\delta}$ of (6.21) is $I\!P_k$–unisolvent, the moving least–squares problem has a unique solution. It coincides with the solution of the minimization of (6.23) under the constraints (6.22) and thus has polynomial reproduction of order k.*

Proof: We already have the first part of the theorem. If we take the optimal solution vector $b^* \in I\!R^Q$ of the first form of the problem, we can write the optimal polynomial

$$p^*(x) := \sum_{\ell=1}^{Q} b_\ell^* p_\ell(x)$$

in terms of the basis $p_1, \ldots, p_Q$ of $I\!P_k$ we used in defining the matrix P_X. Since we know that we have a unisolvent set, we can rewrite the polynomial at arbitrary points $y \in I\!R^d$ as

$$p^*(y) = \sum_{\ell=1}^{Q} b_\ell^* p_\ell(y) = \sum_{j\in J} \hat{a}_j(y) p^*(x_j)$$

with certain nonunique coefficients $\hat{a}_j(y)$. Thus the constraints (6.22) can be satisfied, but our choice of the $\hat{a}_j$ may not be the optimal ones for minimization of erefeqMLSobjfun. The diagonalized quadratic form of (6.23) is positive definite, and thus it attains its unique minimum on all affine subspaces like the one defined by (6.22). Thus there is a solution vector $a_j^*(x)$, and we have to prove that the equation

$$\sum_{j\in J}^{N} a_j^*(x) f(x_j) = p^*(x) = \sum_{\ell=1}^{Q} b_\ell^* p_\ell(x) = \sum_{j\in J} \hat{a}_j(x) p^*(x_j)$$

holds. The new problem takes the form

$$\text{Minimize } \|\sqrt{W^{-1}}a(x)\|_2^2 = a(x)^T W^{-1} a(x) \text{ under } P_X^T a(x) = p(x)$$

with $p(x) := (p_1(x), \ldots, p_Q(x))^T$ as in (4.20). By the standard "parabola" argument for solutions of quadratic problems with affine–linear constraints, we know that $a(x)^T W^{-1} c = 0$ for all $c \in I\!R^{|J|}$ with $P_X^T c = 0$, and by the factorization lemma 9.10 we get **Lagrange multipliers** $\lambda(x) \in I\!R^Q$ with $a(x)^T W^{-1} c = \lambda^T(x) P_X^T c$ for all $c \in I\!R^{|J|}$. This proves that $a^*(x) = W P_X \lambda(x)$ and from $P_X^T a^*(x) = p(x)$ we get $p(x) = P_X^T W P_X \lambda(x)$. But then

$$p^*(x) = p^T(x) b^* = \lambda^T(x) P_X^T W P_X b^* = \lambda^T(x) P_X^T W F = F^T a^*(x)$$

proves the assertion. □

For further analysis, we note some results of the above argument. First, the equation $a^*(x) = W P_X \lambda(x)$ is

$$a_j^*(x) = \Phi_\delta(x - x_j) \sum_{k=1}^{Q} \lambda_k(x) p_k(x_j),\ j \in J,$$

and $p(x) = P_X^T W P_X \lambda(x)$ shows that the $\lambda_k(x)$ satisfy the system

$$p_\ell(x) = \sum_{k=1}^{Q} \lambda_k(x) \sum_{j\in J} p_k(x_j) \Phi_\delta(x - x_j) p_\ell(x_j),\ 1 \le \ell \le Q.$$

This is another way to calculate the solution, and it only requires a $Q \times Q$ system whose entries can be calculated with complexity $|J|$ each.

From the first equation we get that the a_j^* are as smooth as the functions Φ_δ and λ_k allow. But the second system can be written as

$$p_\ell(x) = \sum_{k=1}^{Q} \lambda_k(x) \sum_{j=1}^{N} p_k(x_j) \Phi_\delta(x - x_j) p_\ell(x_j),\ 1 \le \ell \le Q,$$

and thus we have

Corollary 6.25 *If all sets $X_{x,\delta}$ for arbitrary $x \in \Omega$ and fixed δ are $I\!P_k$–unisolvent, then the solution of the moving least–squares approcimation is as smooth as the kernel Φ_δ.*

Proof: Due to global $I\!P_k$–unisolvence of all sets $X_{x,\delta}$, the coefficient matrix is globally nonsingular and has a determinant as smooth as the kernel itself. □

To align moving least–squares with our previous theory of stable local polynomial reproduction, we should fix δ to be $c_1 h$ for a fixed set X with fill distance h. Then we have unisolvence of each set $X_{x,\delta}$ and local polynomial reproduction. The main problem is stability, and for this we shall need **quasi–uniformity** in the sense that the separation distance q and the fill distance h are related by $qc_3 \geq h$ for some positive constant c_3.

For stability, we bound the factors of

$$\left(\sum_{j\in J} |a_j^*(x)|\right)^2 \leq \left(\sum_{j\in J} \frac{|a_j^*(x)|^2}{\Phi_\delta(x-x_j)}\right)\left(\sum_{j\in J} \Phi_\delta(x-x_j)\right)$$

separately. The first is the objective function of the second form of our minimization, and thus it can be bounded by any stable solution $u_j(x)$ we have from the previous theory. To have some leeway, we make δ larger, taking it as $2c_1 h$, while we use the u_j for $c_1 h$. Thus

$$\begin{aligned}
\sum_{j\in J} \frac{|a_j^*(x)|^2}{\Phi_\delta(x-x_j)} &\leq \sum_{j\in J} \frac{|u_j(x)|^2}{\Phi_\delta(x-x_j)} \\
&\leq \frac{1}{\inf_{z\in B(0,1/2)} \Phi(z)} \sum_{j\in J} |u_j(x)|^2 \\
&\leq C\left(\sum_{j\in J} |u_j(x)|\right)^2 \\
&\leq 4C.
\end{aligned}$$

The second factor can be dealt with a counting argument, since it is bounded by a constant times the number $|J|$. Each point of $X_{x,\delta}$ has a ball of radius $q/2$ around it with no other point of X. Since all of these disjoint balls lie in the ball $B(x, \delta + q/2)$, we have

$$|J| vol(B(x_j, q/2)) \leq vol(B(x, \delta + q/2))$$

and

$$|J| \frac{q^d}{2^d} \leq (\delta + q/2)^d \leq (2c_1 h + q/2)^d \leq (2c_1 c_3 q + q/2)^d$$

leading to

$$|J| \leq (4c_1 c_3 + 1)^d.$$

Altogether, we see that moving least–squares realize stable local polynomial reproduction.

6.8 Bramble–Hilbert Lemma

We now leave the stable local polynomial reproduction part and go back to (6.12). We now have to care for the part varying with h, but we already know something about stable local polynomial recovery, i.e. we have

$$\|p\|_{\infty,\Omega} \leq C \|p\|_{\infty,X_h}$$

for all $p \in I\!P_M$ and all set X_h with $h \leq h_0$. For convenience, we restrict our attention to the $L\infty$ norm, and focus on the remaining part

$$\|u\|_{L_\infty(\Omega)} \leq C h^{M-d/2} |u|_{W_2^M(\Omega)}.$$

Clearly, an inequality like this cannot hold unless the function u is replaced by something like $u - p^*$ for some polynomial $p^* \in I\!P_M$, because if the right–hand side is zero, the left–hand side must be zero. Thus we go for something like

$$\|u - p^*\|_{L_\infty(\Omega)} \leq Ch^{M-d/2}|u|_{W_2^M(\Omega)}$$

and inequalities like this are well-known in simple cases like the univariate ones. We already did that for $M = d = 1$, and it also is easy for univariate cases of higher order when p^* is the Taylor polynomial and if the basis interval is of length h.

This observation is the clue for what we are doing next. Let us consider a simple case first, which is a variation of a **Poincaré–Friedrichs** inequality.

Lemma 6.26 *Let Ω be a bounded cube in $I\!R^d$ of maximal sidelength s. Then for each function u in $W_2^1(\Omega)$ there is a constant $\gamma(u)$ such that*

$$\|u - \gamma(u)\|_{L_2(\Omega)} \leq s|u|_{W_2^1(\Omega)} \tag{6.27}$$

and the constant can be taken as the mean value of u over Ω.

Proof: We first prove the assertion for smooth functions, and then we go to the completion limit. There is a point x where $u(x) = \gamma(u)$. We set $v := u - \gamma(u)$. As in the univariate case we now integrate from x to any z in the cube, but we first integrate along the first coordinate only, i.e. we take $z = x + \tau e_1$ and get

$$\begin{aligned} v(z) &= \int_0^\tau \frac{\partial v}{\partial x_1}(x + te_1)dt \\ |v(z)|^2 &\leq \left|\tau \int_0^\tau \left(\frac{\partial v}{\partial x_1}(x + te_1)\right)^2 dt\right| \leq s\left|\int_0^\tau \left(\frac{\partial v}{\partial x_1}(x + te_1)\right)^2 dt\right|. \end{aligned}$$

We integrate this over the full line L of length s through x and z along the first coordinate to get

$$\int_L |v(y)|^2 dy \leq s^2 \int_L \left(\frac{\partial v}{\partial x_1}(y)\right)^2 dy.$$

because the right–hand side is independent of z and the length $|\tau|$ of integration towards z cannot be greater than s. If we integrate both sides over the other dimensions as well, we get

$$\|v\|^2_{L_2(\Omega)} \leq s^2 \int_\Omega \left(\frac{\partial v}{\partial x_1}(y)\right)^2 dy = s^2|v|_1^2.$$

Now the assertion follows when inserting $v = u - \gamma(u)$ and when going over to the Hilbert space closure. □

A more standard and classical version of this, named after Poincaré and Friedrichs, does the same thing without $\gamma(u)$, but with the assumption that u vanishes somewhere on the boundary. The proof is the same.

Unfortunately, we cannot sum up the inequalities (6.27) when combining a larger domain from cube subdomains, because the constants will be different in each subdomain. But we can proceed on cube subdomains Ω_s like

$$\begin{aligned} \|u\|_{L_2(\Omega_s)} &\leq \|u-\gamma(u)\|_{L_2(\Omega_s)} + \|\gamma(u)\|_{L_2(\Omega_s)} \\ &\leq s|u|_{W_2^1(\Omega_s)} + |\gamma_s(u)|\sqrt{vol(\Omega_s)} \\ \|u\|^2_{L_2(\Omega_s)} &\leq 2s^2|u|^2_{W_2^1(\Omega_s)} + 2|\gamma_s(u)|^2 vol(\Omega_s) \end{aligned}$$

and we can sum this up for a domain Ω composed of such subdomains. The result is

$$\begin{aligned} \|u\|^2_{L_2(\Omega)} &\leq 2s^2|u|^2_{W_2^1(\Omega)} + 2vol(\Omega)\sum_{\Omega_s}|\gamma_s(u)|^2 \\ \|u\|_{L_2(\Omega)} &\leq \sqrt{2}s|u|_{W_2^1(\Omega)} + \sqrt{2vol(\Omega)}\sqrt{\sum_{\Omega_s}|\gamma_s(u)|^2} \end{aligned}$$

and can be viewed as our first full–size sampling inequality.

Having understood the basic logic, readers can now imagine that a generalization of Lemma 6.26 is

Lemma 6.28 *("Local" Bramble–Hilbert Lemma)*
Let Ω be a nice domain of diameter s, e.g. a cube, a ball, or a convex or a star–shaped set. Then there is a constant C such that for all functions $u \in W_2^M(\Omega)$ with $M \geq 1$ there is a polynomial $p(u) \in I\!P_M$ such that

$$\|u - p(u)\|_{L_2(\Omega)} \leq Cs^M|u|_{W_2^M(\Omega)}.$$

The polynomial can be chosen as an averaged Taylor polynomial, and the constant is only dependent on the dimension d and the type of "nice" domain.

We just had the case $M = 1$ for cubes, but we want to avoid a full proof (see [7] for the star–shaped case, and certain papers for other cases). The main argument first works on a domain of diameter 1 and bounds the error of the averaged Taylor polynomials uniformly by

$$\|u - p(u)\|_{L_2(\Omega)} \leq C|u|_{W_2^M(\Omega)} \text{ for all } u \in W_2^M(\Omega).$$

Roughly, this is a result of the factorization lemma 9.10, because the operator $Id - T_M$ with T_M being the Taylor projector of order M, vanishes on the kernel $I\!P_M$ of the linear map

$$L_M \ : \ u \mapsto (D^\alpha u, \ |\alpha| = M)$$

and thus must be factorizable over the range of L_M, implying that it can be bounded by the derivatives $D^\alpha u$ with $|\alpha| = M$ alone.

The next step in this rough proof sketch is a plain scaling argument. We now take $u \in W_2^M(\Omega_s)$ and define $v(x) := u(x \cdot s)$ to get some $v \in W_2^M(\Omega_1)$. Then

$$s^{-d}\|u - p(u)\|^2_{L_2(\Omega_s)} = \|v - p(v)\|^2_{L_2(\Omega)} \leq C^2|v|^2_{W_2^M(\Omega)} = C^2 s^{2M} s^{-d}|u|^2_{W_2^M(\Omega_s)}$$

does the job, provided that we also have the scale invariance

$$p(v)(x) = p^t(u(t \cdot s))(x) = p(u)(s \cdot x).$$

But this holds for standard Taylor polynomials at zero

$$\begin{aligned} T_M(v)(x) &= \sum_{j=0}^{M-1} \sum_{|\alpha|=j} \frac{v^\alpha(0)}{\alpha!} x^\alpha \\ &= \sum_{j=0}^{M-1} \sum_{|\alpha|=j} \frac{u^\alpha(0) s^{|\alpha|}}{\alpha!} x^\alpha \\ &= \sum_{j=0}^{M-1} \sum_{|\alpha|=j} \frac{u^\alpha(0)}{\alpha!} (xs)^\alpha \\ &= T_M(u)(xs) \end{aligned}$$

and carries over to the averaged ones.

For the L_∞ norm, the correspondent result is

$$\|u - T_M(u)\|_{L_\infty(\Omega)} \le C s^{M-d/2} |u|_{W_2^M(\Omega)} \tag{6.29}$$

under the restriction $M > d/2$ because otherwise we have no continuous point evaluation. here, we wrote the averaged Taylor projector T_M. We can also sketch the idea that leads to (6.29). By a factorization argument, one can get a result like

$$\|u - T_M(u)\|_{L_\infty(\Omega)} \le C |u|_{W_2^M(\Omega)}$$

for a "nice" domain of diameter one. This time, one can also work with the standard Taylor projector. The next step is again a scaling argument like the one above, using $v(x) = u(x \cdot s)$ connecting v on Ω_1 with u on Ω_s with diameter s. The scaling now gives

$$\|u - T_M(u)\|^2_{L_\infty(\Omega_s)} = \|v - T_M(v)\|^2_{L_\infty(\Omega)} \le C |v|^2_{W_2^M(\Omega)} = C s^{2M-d} |u|^2_{W_2^M(\Omega_s)}$$

which is what we need. Note that this argument is a local proof of Sobolev's imbedding inequality, since it implies that $C(\Omega_s)$ is continuously embedded in $W_2^M(\Omega_s)$ for $M > d/2$.

6.9 Globalization

From (6.29) and stable polynomial reproduction $f \mapsto Q_M(f)$ of order M from values on a set X we can proceed to a local sampling inequality on "nice" domains of diameter s. This starts from bounding the Taylor operator in terms of data on X via

$$\begin{aligned} \|T_M u\|_{\infty,\Omega} &\le C \|T_M u\|_{\infty,X} \\ &\le C \left(\|T_M u - u\|_{\infty,X} + \|u\|_{\infty,X} \right) \\ &\le C \left(\|T_M u - u\|_{\infty,\Omega} + \|u\|_{\infty,X} \right) \\ &\le C \left(C s^{M-d/2} |u|_{W_2^M(\Omega)} + \|u\|_{\infty,X} \right) \end{aligned}$$

and proceeds via

$$\begin{aligned} \|u\|_{L_\infty(\Omega)} &\le \|u - T_M u\|_{L_\infty(\Omega)} + \|T_M u\|_{L_\infty(\Omega)} \\ &\le C s^{M-d/2} |u|_{W_2^M(\Omega)} + C \left(C s^{M-d/2} |u|_{W_2^M(\Omega)} + \|u\|_{\infty,X} \right) \\ &\le C \left(s^{M-d/2} |u|_{W_2^M(\Omega)} + \|u\|_{\infty,X} \right) \end{aligned}$$

with generic constants. This is perfectly fine for "nice" domains of diameter s proportional to h such that a global set X restricted to Ω still has fill distance h. In fact, this can be done at the expense of changing the constants, and it can be done uniformly for arbitrary domains with a fixed cone condition. We do not want to do this in full detail, because it is rather technical and provides no new insights.

But we state the final results for sampling inequalities, as they are provided now by the literature. In all cases, the domain $\Omega \subset \mathbb{R}^d$ has to be bounded with a Lipschitz boundary and an interior cone condition, and the order m has to be fixed beforehand, together with real numbers $1 \leq p, q \leq \infty$. Then there exist constants $C, h_0 > 0$ such that the following inequalities hold for all functions u in m–th order Sobolev space and all discrete sets $X_h \subset \Omega$ with fill distance $h \leq h_0$:

- Narcowich, Ward, & Wendland MC 2005 [48]
  $|u|_{W_q^{|\alpha|}} \leq c h^{m-|\alpha|-d(1/p-1/q)_+} |u|_{W_p^m},\ u(X_h) = \{0\}$
  for $0 \leq |\alpha| \leq m > d/p$,
- Wendland & Rieger Num. Math. 2005 [72]
  $|u|_{W_q^{|\alpha|}} \leq C \left(h^{m-|\alpha|-d(1/p-1/q)_+} |u|_{W_p^m} + h^{-|\alpha|} \|u\|_{\infty, X_h} \right)$
  for $0 \leq |\alpha| \leq m > d/p$,
- Madych JAT 2006 [39]
  $\|u\|_{L_q} \leq C \left(h^{m-d(1/p-1/q)_+} |u|_{W_p^m} + h^{d/max(p.q)} \|u\|_{\ell_p, X_h} \right)$
  for $0 \leq m > d/p$.

Finally, there is a more sophisticated inequality due to Christian Rieger and Barbara Zwicknagl [54]. It holds on Hilbert spaces $\mathcal{H}$ of functions on bounded Lipschitz domains $\Omega \subset \mathbb{R}^d$ with interior cone condition provided that the Hilbert spaces are uniformly and continuously embedded in all Sobolev spaces $W_2^m(\Omega)$ for all $m \in \mathbb{N}$, i.e.

$$\|u\|_{W_2^m(\Omega)} \leq C \|u\|_{\mathcal{H}} \text{ for all } u \in \mathcal{H}.$$

Then for all $1 \leq q \leq \infty$ and $m \geq 0$ there exist constants $C, h_0 > 0$ such that

$$\|u\|_{W_q^m(\Omega)} \leq C \left(\exp\left(c_1 \frac{\log(c_2 h)}{\sqrt{h}} \right) \|u\|_{\mathcal{H}} + h^{-|\alpha|} \|u\|_{\ell_q(X_h)} \right) \tag{6.30}$$

holds for all functions u in $\mathcal{H}$, all discrete sets $X_h \subset \Omega$ with fill distance $h \leq h_0$. The proof of this is based in the Wendland–Rieger form of the fixed–order sampling inequality, but tracks the constants carefully in terms of the order m. Then the used m is connected to h in such a way that the new sampling inequality is obtained.

6.10 Error Bounds

We now can use the sampling inequalities for error bounds concerning kernel interpolation in Sobolev spaces. As we pointed out before, we only need that the native space $\mathcal{N}_K$ for a kernel K is continuously embedded in some Sobolev space of order m, i.e.

$$\|u\|_{W_2^m(\Omega)} \leq C \|u\|_{\mathcal{N}_K} \text{ for all } u \in \mathcal{N}_K. \tag{6.31}$$

If we take a set X_h with fill distance $h \leq h_0$ in one of the above situations, we can use the minimum norm property of the interpolation operator Q_{X_h} to get

$$\|Q_{X_h}u\|_{W_2^m(\Omega)} \leq C\|Q_{X_h}u\|_{\mathcal{N}_K} \leq C\|u\|_{\mathcal{N}_K} \text{ for all } u \in \mathcal{N}_K,$$

and we use Sobolev embedding from (6.31) in one of the sampling inequalities of 6.9 when applying them to the difference $u - Q_{X_h}u$. This yields in the first case

$$\begin{aligned} |u - Q_{X_h}u|_{W_q^{|\alpha|}} &\leq ch^{m-|\alpha|-d(1/2-1/q)_+}|u - Q_{X_h}u|_{W_p^m} \\ &\leq ch^{m-|\alpha|-d(1/2-1/q)_+}\|u - Q_{X_h}u\|_{\mathcal{N}_K} \\ &\leq Ch^{m-|\alpha|-d(1/2-1/q)_+}\|u\|_{\mathcal{N}_K} \end{aligned}$$

for all $0 \leq |\alpha| \leq m > d/2,\ 1 \leq q \leq \infty$ and thus also in the full Sobolev norm

$$\|u - Q_{X_h}u\|_{W_q^\mu} \leq Ch^{m-\mu-d(1/2-1/q)_+}\|u\|_{\mathcal{N}}$$

for all $0 \leq \mu \leq m > d/2,\ 1 \leq q \leq \infty$.

Using Madych's form we get

$$\begin{aligned} \|u - Q_{X_h}u\|_{L_q(\Omega)} &\leq Ch^{m-d(1/2-1/q)_+}|u - Q_{X_h}u|_{W_2^m} \\ &\leq Ch^{m-d(1/2-1/q)_+}\|u\|_{W_2^m} \\ &\leq Ch^{m-d(1/2-1/q)_+}\|u\|_{\mathcal{N}_K} \end{aligned}$$

for $0 \leq m > d/2,\ 1 \leq q \leq \infty$.

In the situation of the refined inequality (6.30), the correspondent error baound

$$\|u - Q_{X_h}u\|_{W_q^m(\Omega)} \leq C\exp\left(c_1\frac{\log(c_2h)}{\sqrt{h}}\right)\|u\|_{\mathcal{H}}$$

yields **exponential** convergence of the error.

7 Construction of Kernels

For this section, we only present some additional material. The standard procedure will be like the one in the book [71] of Holger Wendland, but with several omittances. This theory heavily relies on Fourier transforms, the essentials of which are covered by an appendix in section 10.2.

7.1 General Construction Techniques

This section is planned to give an overview of methods for the construction of new kernels from existing ones. For the time being, we restrict ourselves to translation-invariant cases in $\mathbb{R}^d$.

7.1.1 Elementary Operations

It is very easy to see that (conditionally) positive (semi-) definite functions on Ω form a **cone** in the space of all functions on $\Omega \times \Omega$. In particular, if Φ and Ψ are (conditionally) positive (semi-) definite, so are $\alpha\Phi + \beta\Psi$ for $\alpha, \beta > 0$. Furthermore, if a family Φ_ζ of (conditionally) positive (semi-) definite functions can be integrated against a positive function $w(\zeta)$, the result

$$\Phi(x, y) := \int w(\zeta)\Phi_\zeta(x, y)d\zeta$$

will again be (conditionally) positive (semi-) definite.

7.1.2 Autocorrelation Method

If we cannot start with a (conditionally) positive (semi-) definite function but have an arbitrary function $\Psi \in L_2(\mathbb{R}^d)$, we can form the **autocorrelation function**

$$\Phi(x,y) := \int_{\mathbb{R}^d} \Psi(x-z)\Psi(y-z)dz.$$

This always yields a symmetric positive semidefinite function which even is positive definite, if all translates $\Phi(x_j - \cdot)$ for different points x_j are linearly independent in $L_2(\mathbb{R}^d)$.

7.1.3 Integration Method

The previous method easily generalizes for any Ω. For any function Ψ on $\Omega \times \Pi$ one can formally consider

$$\Phi(x,y) := \int_{\Pi} \Psi(x,\zeta)\Psi(y,\zeta)w(\zeta)d\zeta$$

with a positive weight function w on Π. If the integral is well-defined, the result will be a symmetric positive semidefinite function on Ω.

7.2 Special Kernels on $\mathbb{R}^d$

In Machine Learning, the **polynomial kernels**

$$\begin{aligned} K_n(x,y) &= (x^T y)^n \text{ for all } n \geq 0,\ x,y \in \mathbb{R}^d \\ \text{or } K_n(x,y) &= (1+x^T y)^n \text{ for all } n \geq 0,\ x,y \in \mathbb{R}^d \end{aligned}$$

are often used. Due to Theorem 2.15, they are positive semidefinite when the kernel $K(x,y) = x^T y$ is, but this is easy to see.

Clearly, their translates generate polynomials of degree at most n of d variables, such that the native space of the kernels must be a subspace of this polynomial space. However, the geometry of Ω will determine the native space.

For illustration, consider the kernel $K_1(x,y) = x^T y$. It generates functions

$$g(y) := \left(\sum_{j=1}^{N} a_j x_j \right)^T y,\ y \in \mathbb{R}^d.$$

Each such function lies in the subspace

$$N_\Omega := \{ f_a \ :\ y \to a^T y \ :\ a \in LH(\Omega) \} \tag{7.1}$$

of $(R^N)^*$ where $LH(\Omega)$ is the **linear hull** of Ω, i.e.

$$LH(\Omega) := \text{ span } \{x \in \Omega\}.$$

If Ω lies in a k–dimensional subspace of $\mathbb{R}^d$, the space N_Ω cannot be more than k–dimensional. The inner product in the native space is defined as usual, and it turns out easily that it coincides with the usual dual inner product on $(\mathbb{R}^N)^*$ in the notation of (7.1) as

$$(f_a, f_b) := a^T b \text{ for all } a,b \in \mathbb{R}^N.$$

It is now an interesting exercise to see what happens if we solve a system with the usual kernel matrix for K_1 on any choice of N points, but we skip over details.

To Do: Insert details

The kernel K_1 has an analogon in the periodic case, i.e. the kernel $K(x,y) = \cos(x-y)$. It is an easy exercise to see that it is positive semidefinite. This works similarly for the kernels $K_\alpha(x,y) := \cos(\alpha(x-y))$.

To Do: Insert details

Inspired by the previous example, we can consider kernels in polar coordinates (r,φ) in $\mathbb{R}^2$. If we describe two variables in R^2 via two polar coordinates (r,φ) and (s,ψ), the kernels

$$K_\alpha((r,\varphi),(s,\psi)) := r^\alpha s^\alpha \cos(\alpha(\varphi-\psi))$$

are positive semidefinite and harmonic, i.e. they satisfy the homogeneous Laplace equation $\Delta u = 0$ in both arguments.

To Do: Insert details

These examples are closely related to the complex–valued case

$$K_n(z,u) := (z\overline{u})^n,\ z,u \in \mathbb{C},\ n \in \mathbb{N}_0.$$

To Do: Insert details

From these kernels, we can generate new kernels by additive superposition. Let us do a simple example by taking

$$K(x,y) := \sum_{n=0}^{n} \frac{(x^T y)^n}{n!} = \exp(x^T y),\ x,y \in \mathbb{R}^d.$$

It is well–defined since the series is absolutely convergent everywhere, and it is positive semidefinite due to Theorem 2.15. By an easy additional argument this proves that the Gaussian kernel

$$G(x,y) := \exp(-\|x-y\|_2^2),\ x,y \in \mathbb{R}^d$$

is positive semidefinite.

To Do: Insert details

7.3 Translation–Invariant Kernels on $\mathbb{R}^d$

We now go back to section 2.3 and define kernels on $\Omega := \mathbb{R}^d$ by the feature map

$$\Phi(x) := \exp(-ix^T \cdot) \text{ for all } x \in \mathbb{R}^d$$

into a weighted L_2 feature space

$$\mathcal{F}_c := \{g\ :\ \mathbb{R}^d \to \mathbb{C},\ (2\pi)^{-d/2} \int_{\mathbb{R}^d} |g(\omega)|^2 c(\omega) d\omega < \infty\}$$

for a nonnegative and integrable weight function c on $\mathbb{R}^d$. This defines a kernel K_c in translation–invariant form via

$$K_c(x-y) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} \exp(-i(x-y)^T\omega) c(\omega) d\omega$$

and this coincides with $c^\wedge(x-y)$ since the Fourier transform $c^\wedge$ of c exists pointwise under the above assumption. If c is even in the sense $c(\omega) = c(-\omega)$, the Fourier transform and the kernel are real–valued.

Theorem 7.2 *If c is a nonnegative even and integrable function on $\mathbb{R}^d$, its Fourier transform is a real–valued symmetric translation–invariant positive semidefinite kernel K_c on $\mathbb{R}^d$.* □

This is the easiest approach to translation–invariant kernels on $\mathbb{R}^d$, and it is rather close to the general situation due to the famous but difficult

Theorem 7.3 *(Bochner)*
A continuous complex–valued translation–invariant kernel on $\mathbb{R}^d$ is positive semidefinite if and only if it is the Fourier transform of a nonnegative Borel measure μ on $\mathbb{R}^d$, i.e.

$$K(x,y) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} \exp(i(x-y)^T\omega) d\mu(\omega).$$

Since we omitted measure theory in this text, we do not want to prove Bochner's theorem, but the reader should be aware that the connection between the above theorems is via the case that the measure μ has density c, i.e. $d\mu(\omega) = c(\omega)d\omega$.

The above construction immediately implies that the Gaussian is positive semidefinite on all $\mathbb{R}^d$. It even is positive definite, but we shall prove this soon in more generality.

In fact, we should take a general nonnegative even and integrable function c on Ω and ask for sufficient conditions to make the kernel K_c positive definite. As always, we consider the quadratic form

$$\begin{aligned} & \sum_{j,k=1}^N a_j \overline{a}_k K(x_j, x_k) \\ = \; & (2\pi)^{-d/2} \int_{\mathbb{R}^d} \sum_{j,k=1}^N a_j \overline{a}_k e^{i(x_j-x_k)^T\omega} c(\omega) d\omega \\ = \; & (2\pi)^{-d/2} \int_{\mathbb{R}^d} \left| \sum_{j=1}^N a_j e^{ix_j^T\omega} \right|^2 c(\omega) d\omega \geq 0 \end{aligned}$$

and assume that it vanishes. Then the product of the generalized trigonometric polynomial

$$p(\omega) := p_{a,X}(\omega) := \sum_{j=1}^N a_j e^{ix_j^T\omega} \tag{7.4}$$

with c vanishes almost everywhere. But we can expect that such polynomials cannot vanish on reasonable sets without being identically zero and having zero coefficients. More precisely:

Lemma 7.5 *If a generalized trigonometric polynomial of the above form vanishes on an open set in $\mathbb{R}^d$, it has zero coefficients.*

Proof: By a simple shift (which multiplies each coefficient with a nonzero complex number) we can assume that the open set contains the origin in its interior. Then all derivatives of p at zero must vanish. This implies that all complex numbers

$$D^\beta p(0) = \sum_{j=1}^N a_j (ix_j)^\beta,\ \beta \in I\!N_0^d$$

vanish, and this means that all

$$\sum_{j=1}^N a_j x_j^\beta,\ \beta \in I\!N_0^d$$

are zero. If we pick a single index $j,\ 1 \le j \le N$, we can find a multivariate polynomial $q_j(x)$ with the Lagrange property $q_j(x_k) = \delta_{jk},\ 1 \le j,k \le N$, for instance

$$q_j(x) := \sum_{k \neq j} \frac{\|x - x_k\|_2^2}{\|x_j - x_k\|_2^2} =: \sum_\beta b_\beta^{(j)} x^\beta$$

for finitely many nonzero coefficients $b_\beta^{(j)}$. Then we get

$$\begin{aligned} a_j &= \sum_{k=1}^N a_k q_j(x_k) \\ &= \sum_{k=1}^N a_k \sum_\beta b_\beta^{(j)} x_k^\beta \\ &= \sum_\beta b_\beta^{(j)} \sum_{k=1}^N a_k x_k^\beta \\ &= 0 \end{aligned}$$

for all $j,\ 1 \le j \le N$. □

Theorem 7.6 *If c is a nonnegative even and integrable function on $I\!R^d$ which is positive on an open set, its Fourier transform is a real–valued symmetric translation–invariant positive definite kernel K_c on $I\!R^d$.* □

This now implies that the Gaussian is positive definite, because its inverse Fourier transform is never zero.

Furthermore, since both the kernel and its inverse Fourier transform are real–valued and symmetric, we can ignore the difference between inverse and forward Fourier transform in this context.

We can turn the above theorem upside–down to get

Theorem 7.7 *Let K be a translation–invariant symmetric real–valued kernel on $I\!R^d$ whose Fourier transform exists, is even and nonnegative and integrable on $I\!R^d$ and positive on an open set. Then K is positive definite.* □

This gives us plenty of leeway to construct positive definite kernels. However, we are interested in explicitly known kernels only, and then we have to check their Fourier transforms. For instance, Wendland's kernel

$$K(x-y) = (1 - \|x-y\|_2)_+^4 (1 + 4\|x-y\|_2),\ x,y \in I\!R^d$$

has positive Fourier transform in $I\!R^d$ for $d \le 3$, but this is not at all clear. We postpone such kernels for a while.

7.4 Global Sobolev Kernels on $\mathbb{R}^d$

Clearly, we should look for the reproducing kernel of global Sobolev space $W_2^m(\mathbb{R}^d)$. This is defined as the space of functions with generalized derivatives up to order m being in $L_2(\mathbb{R}^d)$, and we can cast this into a condition on Fourier transforms. If f is a smooth function on $\mathbb{R}^d$, we know that the Fourier transform of the derivative $D^\alpha f$ is the function $\omega \mapsto (i\omega)^\alpha \hat{f}(\omega)$, and this is in $L_2(\mathbb{R}^d)$ if the integral

$$\int_{R^d} |D^\alpha f|^2(x)dx = \int_{\mathbb{R}^d} |(i\omega)^\alpha \hat{f}(\omega)|^2 d\omega = \int_{\mathbb{R}^d} |\omega^\alpha|^2 |\hat{f}(\omega)|^2 d\omega$$

is finite. Thus Sobolev space $W_2^m(\mathbb{R}^d)$ can be defined via the inner product

$$\begin{aligned}(f,g)_{W_2^m(\mathbb{R}^d)} &:= \sum_{j=0}^m \binom{m}{j} \sum_{|\alpha|=j} \binom{j}{\alpha} \int_{R^d} D^\alpha f(x) \overline{D^\alpha g}(x)dx \\ &= \sum_{j=0}^m \binom{m}{j} \sum_{|\alpha|=j} \binom{j}{\alpha} \int_{R^d} |\omega^\alpha|^2 \hat{f}(\omega)\overline{\hat{g}(\omega)}d\omega \\ &= \int_{R^d} \left(1+\|\omega\|_2^2\right)^m \hat{f}(\omega)\overline{\hat{g}(\omega)}d\omega\end{aligned}$$

and consists of all functions f on $\mathbb{R}^d$ with

$$\|f\|^2_{W_2^m(\mathbb{R}^d)} = \int_{R^d} \left(1+\|\omega\|_2^2\right)^m |\hat{f}(\omega)|^2 d\omega < \infty.$$

We now look for the kernel K which will be reproducing in Sobolev space $W_2^m(\mathbb{R}^d)$. We write it in difference form right away, and we need the relation

$$f(x) = (f, K(x-\cdot))_{W_2^m(\mathbb{R}^d)} \text{ for all } x \in \mathbb{R}^d,\ f \in W_2^m(\mathbb{R}^d).$$

We formally see that

$$(\hat{K}(x-\cdot))(\omega) = e^{-ix^T\omega}\hat{K}(\omega)$$

and find that we have to satisfy

$$\begin{aligned}f(x) &= (f, K(x-\cdot))_{W_2^m(\mathbb{R}^d)} \\ &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{f}(\omega) \left(1+\|\omega\|_2^2\right)^m e^{+ix^T\omega}\overline{\hat{K}(\omega)}d\omega\end{aligned}$$

which works if we can set

$$\hat{K}(\omega) = \left(1+\|\omega\|_2^2\right)^{-m}$$

and if K and f are inverse Fourier–transformable.

Since we now have an idea what the kernel should be, we define it as

$$K(x-y) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} \left(1+\|\omega\|_2^2\right)^{-m} e^{i(x-y)^T\omega}d\omega \tag{7.8}$$

This integral is well–defined if $-2m < -d$ or $m > d/2$, which is the usual sufficient condition for an embedding of $C(\Omega)$ into $W_2^m(\Omega)$. Thus the kernel exists pointwise, and we have to check whether $K(x-\cdot)$ lies in $W_2^m(\Omega)$. To this end, we check the Fourier transform condition

$$\begin{aligned}&(2\pi)^{-d/2} \int_{R^d} \left(1+\|\omega\|_2^2\right)^m |(\hat{K}(x-\cdot))(\omega)|^2 d\omega \\ = &(2\pi)^{-d/2} \int_{R^d} \left(1+\|\omega\|_2^2\right)^m \hat{K}(\omega)^2 d\omega \\ = &(2\pi)^{-d/2} \int_{R^d} \left(1+\|\omega\|_2^2\right)^{-m} d\omega\end{aligned}$$

which is finite and equal to $K(0)$, again due to the condition $m > d/2$.

What is left is the inverse Fourier transformability of f, since we can form the right–hand side of the reproduction equation, and it is

$$(2\pi)^{-d/2}\int_{R^d} e^{ix^T\omega}\hat{f}(\omega)d\omega.$$

This integral is classically integrable because of

$$\begin{aligned}
& \int_{R^d} |\hat{f}(\omega)|d\omega \\
= & \int_{R^d} \left(1+\|\omega\|_2^2\right)^{m/2} |\hat{f}(\omega)| \left(1+\|\omega\|_2^2\right)^{-m/2} d\omega \\
\leq & \sqrt{\int_{R^d} (1+\|\omega\|_2^2)^m |\hat{f}(\omega)|^2 d\omega}\sqrt{\int_{R^d} (1+\|\omega\|_2^2)^{-m} d\omega} \\
\leq & C\|f\|_{W_2^m(\Omega)}K(0)
\end{aligned}$$

and thus it represents $f(x)$. We have

Theorem 7.9 *The reproducing kernel for Sobolev space $W_2^m(\mathbb{R}^d)$ for $m > d/2$ is given by (7.8) and turns out to have the explicit radial representation*

$$\frac{2^{1-m}}{(m-1)!}\|x-y\|_2^{m-d/2}K_{-m+d/2}(\|x-y\|_2) \tag{7.10}$$

where K_ν is the modified Bessel function of order ν.

We postpone the explicit calculation ending with the above formula, but in Figure 25 we include a plot of the kernels $r^\nu K_\nu(r)$ after a rescaling to attain 1 at zero. In section 10.6.8 we provide some properties of these functions. In particular, they decrease monotonically away from zero, and they have exponential decay towards infinity. At zero, they have limited smoothness.

7.5 Native Spaces of Translation–Invariant Kernels

After we have seen the special case of a kernel that directly leads to gloabel Sobolev space, we now go back to the more general situation of a translation–invariant kernel K_c generated from an even,l nonnegative, and summable Fourier transform $c = \hat{K}_c$. We want to calculate the native space of the kernel, but in order to be aligned with our error analysis, we have to do this on a bounded domain $\Omega \subset \mathbb{R}^d$. We can drop c and work with K and $\hat{K}$ directly.

The inner product in the native space is defined for typical functions $f_{a,X}$ as in (2.18) in section 2.4. But we can now use Fourier transforms on it and get first

$$\begin{aligned}
(\hat{f}_{a,X})(\omega) &= (2\pi)^{-d/2}\int_{\mathbb{R}^d} f_{a,X}(x)e^{-ix^T\omega}dx \\
&= (2\pi)^{-d/2}\sum_{j=1}^M a_j e^{-ix_j^T\omega}\int_{\mathbb{R}^d} K(x-x_j)e^{i(x_j-x)^T\omega}dx \\
&= \hat{K}(\omega)\sum_{j=1}^M a_j e^{-ix_j^T\omega} \\
&=: \hat{K}(\omega)\, p_{a,X}(\omega)
\end{aligned}$$

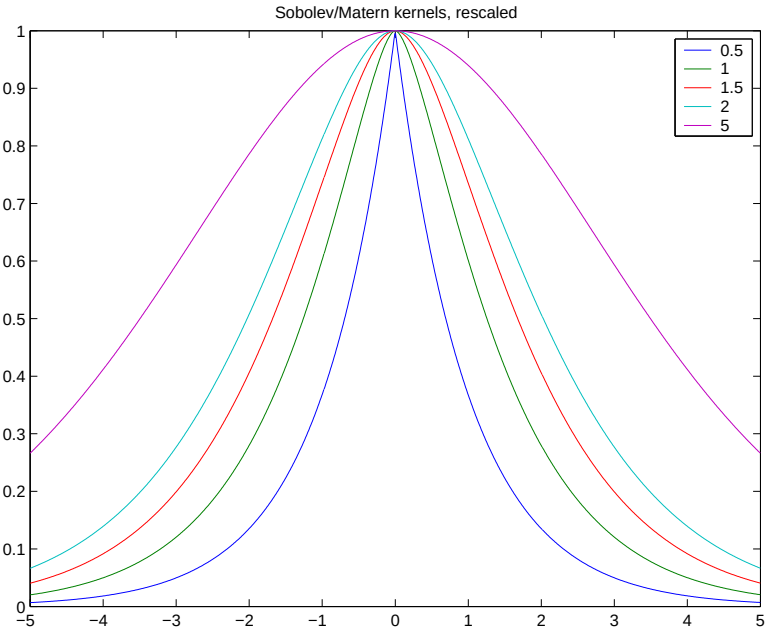

Figure 25: The Matern/Sobolev kernels $r^\nu K_\nu(r)$

and then

$$\begin{aligned}
(f_{a,X}, f_{b,Y})_K &= \sum_{j=1}^{M}\sum_{k=1}^{N} a_j b_k K(y_k, x_j) \\
&= (2\pi)^{-d/2} \sum_{j=1}^{M}\sum_{k=1}^{N} a_j b_k \int_{I\!R^d} \hat{K}(\omega) e^{i(y_k - x_j)^T\omega} d\omega \\
&= (2\pi)^{-d/2} \int_{I\!R^d} \hat{K}(\omega) \sum_{j=1}^{M} a_j e^{-ix_j^T\omega} \sum_{k=1}^{N} b_k e^{iy_k^T\omega} d\omega \\
&= (2\pi)^{-d/2} \int_{I\!R^d} \frac{\hat{f}_{a,X}(\omega)\overline{\hat{f}_{b,Y}(\omega)}}{\hat{K}(\omega)} d\omega.
\end{aligned}$$

Thus we can read off the right inner product of the native space. We define

$$\mathcal{F}_K := \{f \in L_2(I\!R^d) \ : \ \int_{I\!R^d} \frac{|\hat{f}(\omega)|^2}{\hat{K}(\omega)} d\omega < \infty\}$$

and see that this space contains the native space for K because it contains it as a set and has the same topology. Since, by definition as a closure, the native space for K is closed, it is a closed subspace of $\mathcal{F}_K$. We now look at its orthogonal complement. For this, we take any $f \in \mathcal{F}_K$ and evaluate the inner product

$$\begin{aligned}
(f, f_{a,X})_K &= (2\pi)^{-d/2} \int_{I\!R^d} \frac{\hat{f}(\omega)\overline{\hat{f}_{a,X}(\omega)}}{\hat{K}(\omega)} d\omega \\
&= (2\pi)^{-d/2} \int_{I\!R^d} \hat{f}(\omega)\overline{p_{a,X}(\omega)} d\omega \\
&= (2\pi)^{-d/2} \int_{I\!R^d} \hat{f}(\omega) \sum_{j=1}^{M} a_j e^{ix_j^T\omega} d\omega \\
&= \sum_{j=1}^{M} a_j f(x_j)
\end{aligned}$$

which implies that K is the reproducing kernel in $\mathcal{F}_K$ on the full domain $\mathbb{R}^d$.

If f is orthogonal to all $f_{a,X}$ with $X \subset \Omega$ for a bounded domain $\Omega \subset \mathbb{R}^d$, we see that $f(\Omega) = \{0\}$, and conversely.

Theorem 7.11 *The native space for a general translation–invariant symmetric positive definite kernel K on a domain Ω is the orthogonal subspace of the space of functions in $\mathcal{F}_K$ vanishing on Ω, where orthogonality is understood in $\mathcal{F}_K$.* □

We now check for which m we have a continuous embedding of the native space $\mathcal{N}_K$ of K into $W_2^m(\Omega)$. We take an $f \in \mathcal{N}_K$ and note first that it is in $\mathcal{F}_K$, which means that it has a global extension and satisfies

$$\int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{\hat{K}(\omega)} d\omega < \infty.$$

We now check if we can prove $f \in W_2^m(\mathbb{R}^d)$. This would work if we get

$$\begin{aligned} & \int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 (1 + \|\omega\|_2^2)^m d\omega \\ = & \int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{\hat{K}(\omega)} \hat{K}(\omega)(1 + \|\omega\|_2^2)^m d\omega \\ \leq & \left(\sup_{\omega \in \mathbb{R}^d} \hat{K}(\omega)(1 + \|\omega\|_2^2)^m \right) \cdot \int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{\hat{K}(\omega)} d\omega < \infty \end{aligned}$$

under the hypothesis

$$\sup_{\omega \in \mathbb{R}^d} \hat{K}(\omega)(1 + \|\omega\|_2^2)^m \leq C < \infty$$

or

$$\hat{K}(\omega) \leq C(1 + \|\omega\|_2^2)^{-m} \text{ for all } \omega \in \mathbb{R}^d. \tag{7.12}$$

Theorem 7.13 *If a translation–invariant symmetric positive definite kernel K on $\mathbb{R}^d$ satisfies (7.12) for some $m > d/2$, then its native space $\mathcal{N}_K$ is continuously embedded in $W_2^m(\mathbb{R}^d)$, and its restriction to a domain Ω is in $W_2^m(\Omega)$. Furthermore, interpolation on subsets X of Ω with fill distance $h \leq h_0(m, \Omega)$ has convergence order $h^{m-d/2}$ for $h \to 0$ in the L_∞ norm on Ω.* □

Note that the condition (7.12) is closely related to the smoothness of the kernel K in the global $L_2(\mathbb{R}^d)$ sense. Roughly spoken, it means that the kernel itself is in Sobolev space $W_2^n(\mathbb{R}^d)$ for all $n < 2m - d/2$.

7.6 Construction of Positive Definite Radial Functions on $\mathbb{R}^d$

This subsection contains tools from [74] as generalized in [64] for the construction of positive definite radial functions on $\mathbb{R}^d$. We start with the standard reduction of d-variate Fourier transforms of radial functions to Hankel transforms of univariate functions. Introducing $t = r^2/2$ as a new variable, two such transforms for different space dimensions are related to each other by a simple univariate differential or integral operator that preserves compact supports. This fundamental trick of Z. Wu then opens up the way for the easy derivation of various compactly supported radial basis functions.

7.6.1 Hankel Transforms

We assume a radial function $\Phi(\cdot) = \phi(\|\cdot\|_2)$ to be given such that $\phi : \mathbb{R}_{>0} \to \mathbb{R}$ has some decay towards infinity that we are going to quantify later. Let us formally look at the Fourier transform formula and simplify it, using radiality, and introducing polar coordinates for x:

$$\begin{aligned}\widehat{\Phi}(\omega) &= (2\pi)^{-d/2}\int_{\mathbb{R}^d}\Phi(x)e^{-ix\cdot\omega}dx \\ &= (2\pi)^{-d/2}\int_{\mathbb{R}^d}\phi(\|x\|_2)e^{-ix\cdot\omega}dx \\ &= (2\pi)^{-d/2}\int_0^\infty \phi(r)r^{d-1}\int_{\|y\|_2=1} e^{-ir\|\omega\|_2 y\cdot\frac{\omega}{\|\omega\|_2}}dydr.\end{aligned}$$

This contains the function $F(r\|\omega\|_2, d)$ defined in (10.28) by the integral

$$F(t,d) := \int_{\|y\|_2=1} e^{-ity\cdot z}dy$$

for $t \geq 0$ and some $\|z\|_2 = 1,\ z \in \mathbb{R}^d$. Using its representation (10.30) via a Bessel function, we get the very important equation

$$\begin{aligned}\widehat{\Phi}(\omega) &= (2\pi)^{-d/2}\sigma_{d-2}\int_0^\infty \phi(r)r^{d-1}\frac{\Gamma(\frac{d-1}{2})\Gamma(\frac{1}{2})}{(r\|\omega\|_2/2)^{(d-2)/2}}J_{(d-2)/2}(r\|\omega\|_2)dr \\ &= \|\omega\|_2^{-(d-2)/2}\int_0^\infty \phi(r)r^{d/2}J_{(d-2)/2}(r\|\omega\|_2)dr.\end{aligned} \tag{7.14}$$

that allows the Fourier transform of a radial function to be written as a univariate **Hankel transform**. Equation (7.14) implies that the Fourier transform of a radial function Φ is again a radial function. It holds also for $d = 1$, as can be proven by direct calculation and

$$\sqrt{\frac{\pi}{2z}}J_{-1/2}(z) = \frac{\cos z}{z}. \tag{7.15}$$

This equation is not directly compatible with (10.29), because the latter does not exist for $\nu = -1/2$. But we can use the usual power series representation (10.31) of Bessel functions to get (7.15) from (10.33).

7.6.2 Bessel Kernels

We apply the Hankel transform for evaluating the Fourier transform of the characteristic function χ_1 of the unit ball in $\mathbb{R}^d$. This is needed in the proof of a theorem tn stability theory, but it also yields useful new kernels.

In particular, we apply (10.37) and get

$$\begin{aligned}\widehat{\chi_1}(\omega) &= \|\omega\|_2^{-(d-2)/2}\int_0^1 r^{d/2}J_{(d-2)/2}(r\|\omega\|_2)dr \\ &= \|\omega\|_2^{-d/2}J_{d/2}(\|\omega\|_2).\end{aligned} \tag{7.16}$$

Considered as a univariate radial function, this is an entire analytic function of exponential type that we shall meet again later. Figure 26 shows the kernels $r^{-\nu}J_\nu(r)$ for various ν after rescaling to have value 1 at zero. For $\nu \in \mathbb{Z}/2$ they are positive definite on $\mathbb{R}^d$ for $d \leq 2\nu$, since they are positive definite on $\mathbb{R}^d$ for $d = 2\nu$ and all smaller dimensions. Since their Fourier transform is compactly supported, they are **band–limited** and they generalize the usual *sinc* function.

Theorem 7.17 *The Bessel kernel $r^{-\nu}J_\nu(r)$ when acting as a radial kernel $K(x,y) := \|x-y\|_2^\nu J_\nu(\|x-y\|_2)$ on $\mathbb{R}^d$ is positive definite if $2\mathbb{Z} \ni \nu \geq d/2$. The kernel $r^{-d/2}J_{d/2}(r)$ generates the space of* **bandlimited functions** *on $\mathbb{R}^d$ with L_2 Fourier transforms supported in the unit ball of $\mathbb{R}^d$.* □

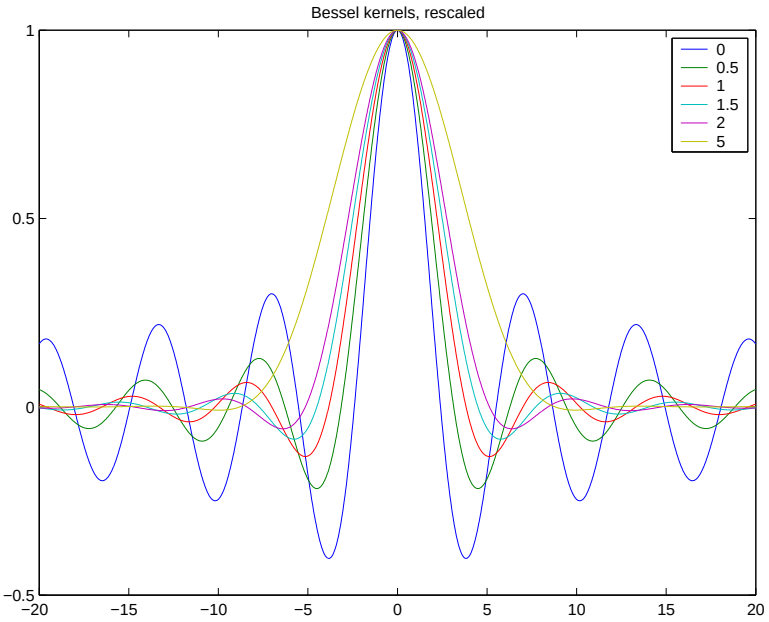

Figure 26: The Bessel kernels $r^{-\nu}J_\nu(r)$

From (10.30) we know that $F(r,d)$ behaves like $r^{-\nu}J_\nu(r)$ for $\nu = (d-2)/2$ for $d > 1$. Its Fourier transform on $\mathbb{R}^d$ is not positive on an open set, and thus we have to invest some work in order to prove positive definiteness on $\mathbb{R}^d$, while Theorem 7.17 guarantees positive definiteness only on $\mathbb{R}^k$ for $0 \leq k \leq d-2$. This was first observed in [21].

Theorem 7.18 *The Bessel kernel $r^{-\nu}J_\nu(r)$ for $\nu = (d-2)/2$ is positive definite on $\mathbb{R}^d$ for $d > 1$.*

Proof: With our standard argument we have to prove that a generalized polynomial $p = p_{a,X}$ of the form (7.4) has zero coefficients, if it vanishes on the sphere. If we single out any two coordinates of Ω, we have an analytic function which vanishes on a circle, thus it vanishes for all arguments. Repeating this for all combinations of two variables, we see that the polynomial must vanish on all of $\mathbb{R}^d$, and then we can proceed as before to prove positive definiteness of the kernel. □

From [21] we also take the interesting observation that the kernel $(cr)^{-(d-2)/2}J_{(d-2)/2}(rc)$ satisfies the Laplace eigenvalue equation (or the **Helmholtz equation**) $\Delta u + c^2 u = 0$ for $d > 1$ dimensions.

A second application of the Hankel transform is the proof of (7.10) in Theorem 7.9. A more explicit and direct proof is in [71] on pages 76–77, but we cite (10.50) with

$$\nu = \frac{d-2}{2},\ \mu = m-1,\ t = r,\ a = \|\omega\|_2,\ z = c$$

to get

$$\int_0^\infty \frac{r^{d/2} J_{(d-2)/2}(r\|\omega\|_2)}{(r^2+c^2)^m} dr = \frac{\|\omega\|_2^{m-1} c^{-m+d/2}}{2^{m-1}\Gamma(m)} K_{-m+d/2}(c\|\omega\|_2).$$

Combining with (7.14) applied to $\phi(r) = (r^2+c^2)^{-m}$ this yields

$$\begin{aligned} \hat{\Phi}(\omega) &= \frac{\|\omega\|_2^{m-1-(d-2)/2} c^{-m+d/2}}{2^{m-1}\Gamma(m)} K_{-m+d/2}(c\|\omega\|_2) \\ &= \left(\frac{\|\omega\|_2}{c}\right)^{m-d/2} \frac{2^{1-m}}{(m-1)!} K_{-m+d/2}(c\|\omega\|_2). \end{aligned}$$

7.6.3 Change of Variables

We now introduce $t = r^2/2$ as a new variable, writing a radial basis function Φ as

$$\Phi(\cdot) = \phi(\|\cdot\|_2) = f(\|\cdot\|_2^2/2), \tag{7.19}$$

and we shall use Latin characters $f, g, \ldots$ to distinguish the transformed functions from the original ones ϕ, ψ, etc. Note that going over from Φ to ϕ and further to f loses the information on the dimension of the space that we want to work on. But we can take advantage of this loss and write dimension-dependent operations like Fourier transforms as univariate operations with a scalar parameter d.

We keep the dimension d in mind and rewrite the d-variate Fourier transform equation (7.14) in terms of the transformed function f to get

$$\begin{aligned} \widehat{\Phi}(\omega) &= \|\omega\|_2^{-\frac{d-2}{2}} \int_0^\infty f(s^2/2) s^{d/2} J_{\frac{d-2}{2}}(s\cdot\|\omega\|_2) ds \\ &= \int_0^\infty f\left(\frac{s^2}{2}\right)\left(\frac{s^2}{2}\right)^{\frac{d-2}{2}} \left(\frac{s\cdot\|\omega\|_2}{2}\right)^{-\frac{d-2}{2}} J_{\frac{d-2}{2}}(s\cdot\|\omega\|_2) s\, ds \\ &= \int_0^\infty f\left(\frac{s^2}{2}\right)\left(\frac{s^2}{2}\right)^{\frac{d-2}{2}} H_{\frac{d-2}{2}}\left(\frac{s^2}{2}\cdot\frac{\|\omega\|_2^2}{2}\right) s\, ds \end{aligned}$$

with the functions J_ν and H_ν defined by

$$\left(\frac{z}{2}\right)^{-\nu} J_\nu(z) = H_\nu(z^2/4) = \sum_{k=0}^\infty \frac{(-z^2/4)^k}{k!\Gamma(k+\nu+1)}$$

for $\nu \in \mathbb{C}$ as in (10.32). If we substitute $t = s^2/2$, we find

$$\begin{aligned} \widehat{\Phi}(\omega) &= \int_0^\infty f(t) t^{\frac{d-2}{2}} H_{\frac{d-2}{2}}\left(t\cdot\frac{\|\omega\|^2}{2}\right) dt \\ &=: \left(F_{\frac{d-2}{2}} f\right)(\|\omega\|^2/2) \end{aligned} \tag{7.20}$$

with the general operator

$$(F_\nu f)(r) := \int_0^\infty f(t) t^\nu H_\nu(tr) dt. \tag{7.21}$$

Theorem 7.22 *The d–variate Fourier transform of a radial kernel ϕ with $\phi(r) = f(r^2/2)$ is given by*

$$F_{(d-2)/2}(f)(\|\omega\|_2^2/2).$$

The operator F_ν is formally defined for all $\nu > -1$ and sufficiently nice functions f, but we can extend it to all $\nu \in \mathbb{R}$, if we omit terms in the series of H_ν that have a singularity of the Gamma function in their denominator. However, we want to check its domain of definition with respect to functions f on $\mathbb{R}_{>0}$ for $\nu > -1$. Near zero, the function $f(t)t^\nu$ should be absolutely integrable, because the analyticity of H_ν causes no problems at zero. For large ν this allows a moderate singularity of f at zero. Near infinity we have to check the decay of H_ν. But since the Bessel functions J_ν have a $\mathcal{O}(t^{-1/2})$ behaviour for $t \to \infty$ due to (10.41), we see that $H_\nu(t)$ decays like $t^{-\nu/2-1/4}$. Thus we require integrability of $f(t)t^{\nu/2-1/4}$ at infinity for $\nu > -1$. Since we do not need the weakest conditions, we can simply assume

$$f(t)t^\nu \in L_1(\mathbb{R}_{>0}). \tag{7.23}$$

Note that both F_ν and H_ν generalize to arbitrary $\nu \in \mathbb{R}$, provided that certain restrictions on f like (7.23) hold. Furthermore, by symmetry of radial functions and our definition of Fourier transforms we have

$$F_{\frac{d-2}{2}}^{-1} = F_{\frac{d-2}{2}} \qquad \text{for } d \in \mathbb{N}$$

on sufficiently smooth functions with sufficient decay. We shall see later that this generalizes to $F_\nu^{-1} = F_\nu$ for all $\nu \in \mathbb{R}$, wherever both operators are defined. Please keep in mind that the parameter ν is related to the space dimension d via $\nu = (d-2)/2$. We shall work with ν instead of d for notational simplification. Furthermore, we consider a space $\mathcal{S}_{rad}$ of **tempered radial functions**. It could be defined as a subspace of the space $\mathcal{S}$ of d-variate tempered test functions, comprising all radial test functions after introducing $\|x\|_2^2/2$ as a new variable. However, we prefer to define it as the space of real-valued functions on $[0,\infty)$ that are infinitely differentiable such that all derivatives vanish faster than any polynomial at infinity. Taking derivatives of (7.19), one can easily see that this yields a subspace of radial test functions on $\mathbb{R}^d$ for all space dimensions d. Conversely, any radial test function Φ in the form (7.19) yields a function f that is in $\mathcal{S}_{rad}$. To see this one can proceed inductively, using

$$\frac{\partial^m}{\partial \omega_j^m}\Phi(\omega) = f^{(m)}(\|\omega\|_2^2/2)\omega_j^m + \text{ lower derivatives with polynomial factors.}$$

Thus the two notions of $\mathcal{S}$ coincide, and each radial function which yields a test function for a specific space dimension will provide a test function for any dimension. Thus $\mathcal{S}_{rad}$ is the proper space to define the operators F_ν on, and it clearly contains e^{-r}, which can easily proven to be a fixed point of any F_ν, using the definitions (10.32) of H_ν and (10.24) of the Gamma function.

7.6.4 Calculus on the Halfline

In the space $\mathcal{S}_{rad}$ we can introduce a quite useful generalization of the classical calculus operations. We start with the family of operators

$$I_\alpha(f)(r) := \int_0^\infty f(s)\frac{(s-r)_+^{\alpha-1}}{\Gamma(\alpha)}ds \tag{7.24}$$

on $\mathcal{S}_{rad}$ for all $\alpha > 0$. The simplest special case is

$$I_1(f)(r) := \int_r^\infty f(s)ds$$

with the inverse

$$I_{-1}(f)(r) := -f'(r).$$

Note that this operation implies that the resulting function vanishes at infinity, and thus there is no additive constant in the integration. Furthermore, the identity

$$Id = I_1^n \circ I_{-1}^n$$

is Taylor's formula at infinity, as follows from (7.24). The identity (10.25) allows a direct proof of the property

$$I_\alpha \circ I_\beta = I_{\alpha+\beta} \tag{7.25}$$

for all $\alpha,\ \beta > 0$ by application of Fubini's theorem. Differentiation and integration by parts imply

$$\begin{array}{rcll} I_{-1}^n \circ I_\alpha & = & I_{\alpha-n} & 0 < \alpha < n \\ I_{n+\alpha} \circ I_{-1}^n & = & I_\alpha & \alpha > 0, n > 0. \end{array}$$

By $I_\alpha = I_\alpha \circ I_n \circ I_{-1}^n = I_n \circ I_\alpha \circ I_{-1}^n$ we get

$$I_{-1}^n \circ I_\alpha = I_\alpha \circ I_{-1}^n,$$

and this suffices to prove that (7.25) holds for all $\alpha, \beta \in I\!R$ if we define

$$\begin{array}{rcl} I_0 & := & Id \\ I_{-n} & := & I_{-1}^n,\ n > 0 \\ I_\alpha & := & I_{\alpha-\lfloor\alpha\rfloor} \circ I_{\lfloor\alpha\rfloor} \end{array}$$

for the remaining cases of α. Altogether, we have

Theorem 7.26 *The operators I_α on $\mathcal{S}_{rad}$ form an abelian group under composition which is isomorphic to $I\!R$ under "+" via $\alpha \mapsto I_\alpha$.*

Proof: The remaining things are easy to prove using the above rules. □

Let us do some simple examples of differentiation and integration of fractional order. The independent variable will be denoted by t, and we indicate the domain of validity of the different cases, because we do not restrict ourselves to tempered radial functions.

$$\begin{array}{rcll} I_\alpha(f(t+x))(r) & = & I_\alpha(f(t))(r+x) & \alpha \in I\!R,\ x \geq 0 \\ I_\alpha(f(tx))(r) & = & x^{-\alpha} I_\alpha(f(t))(rx) & \alpha \in I\!R,\ x \geq 0 \\ I_\alpha(e^{-st})(r) & = & s^{-\alpha} e^{-sr} & \alpha \in I\!R,\ s > 0 \\ I_\alpha(t^{-\beta}\Gamma(\beta))(r) & = & r^{-(\beta-\alpha)}\Gamma(\beta-\alpha) & \beta > 0,\ \alpha < \beta \\ I_\alpha((x+t)^{-\beta}\Gamma(\beta))(r) & = & (x+r)^{-(\beta-\alpha)}\Gamma(\beta-\alpha) & \beta > 0,\ \alpha < \beta,\ x > 0 \\ I_\alpha\left(\dfrac{(s-t)_+^{\beta-1}}{\Gamma(\beta)}\right)(r) & = & \dfrac{(s-r)_+^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)} & \beta > 0,\ \alpha+\beta > 0 \end{array}$$

We shall make specific use of the "semi-integration" operator and its inverse, the "semi-differentiation", as given by

$$\begin{aligned} I_{1/2}(f)(r) &= \int_r^\infty \frac{f(s)}{\sqrt{\pi(s-r)}} ds \\ I_{-1/2}(f)(r) &= -\int_r^\infty \frac{f'(s)}{\sqrt{\pi(s-r)}} ds \\ &= I_{1/2} \circ I_{-1}(f)(r), \end{aligned} \tag{7.27}$$

that are inverses of each other.

A very simple representation of the operators I_α is possible via the **Laplace transform**

$$L(\varphi)(r) := \int_0^\infty \varphi(s) e^{-rs} ds \tag{7.28}$$

which exists classically for any continuous function φ on $[0, \infty)$ that grows at most polynomially towards infinity. For the time being, we ignore the more general definitions of Laplace transforms and observe that the action of I_α can be written down as

$$I_\alpha(L(\varphi)(\cdot)) := L(\varphi(\cdot)(\cdot)^{-\alpha}),$$

where all real α are formally possible (provided that φ behaves nicely enough).

7.6.5 Basic Transitions

The main advantage of $\mathcal{S}_{rad}$ and the definition (7.21) of the radial Fourier transform using (7.20) is that we can compare Fourier transforms for various dimensions, while working on a simple space of univariate functions. But the most surprising fact, as discovered by Wu, shows up when we simply take the derivative of $F_\nu(f)(r)$. We use (10.34) to get

$$\begin{aligned} -\frac{d}{dr} F_\nu(f)(r) &= (I_{-1} \circ F_\nu)(f)(r) \\ &= -\frac{d}{dr} \int_0^\infty f(t) t^\nu H_\nu(rt) dt \\ &= -\int_0^\infty f(t) t^\nu \frac{d}{dr} H_\nu(rt) dt \\ &= \int_0^\infty f(t) t^{\nu+1} H_{\nu+1}(rt) dt \\ &= F_{\nu+1}(f)(r). \end{aligned} \tag{7.29}$$

Going back to $\nu = (d-2)/2$, we see that the $(d+2)$-variate Fourier transform of a radial function after the substitution (7.19) is nothing else than the negative univariate derivative of the d-variate Fourier transform after (7.19). We shall generalize the above identity later to $I_\alpha \circ F_\nu = F_{\nu-\alpha}$ on $\mathcal{R}$, but we already know that $I_1 \circ F_\nu = F_{\nu+1}$ allows to proceed from $(d+2)$-variate radial Fourier transforms to d-variate Fourier transforms by univariate integration.

Let us apply (10.35) to get another identity on tempered functions:

$$\begin{aligned} F_\nu(-f')(r) &= \int_0^\infty -f'(s)s^\nu H_\nu(sr)ds \\ &= \int_0^\infty f(s)s^{\nu-1}H_{\nu-1}(sr)dsdt \\ &= = F_{\nu-1}(f)(r). \end{aligned} \tag{7.30}$$

This will generalize to $F_\nu \circ I_\alpha = F_{\nu+\alpha}$ and is a trivial consequence of $I_\alpha \circ F_{\nu+\alpha} = F_\nu$ and $F_\mu^2 = Id$, if the latter holds in general.

Note that in both cases we have operators that preserve compact supports. The integral operator even preserves nonegativity (it is a **monotone operator**). The explicit construction of compactly supported radial functions relies heavily on these features. But we also want to proceed from d-variate Fourier transforms to $(d+1)$- or $(d-1)$-variate Fourier transforms. This will be achieved by the operator $I_{1/2}$ and its inverse from (7.27). We shall treat this problem in general, comparing two arbitrary instances F_ν and F_μ.

7.6.6 Identities for Transforms, First Version

We can easily evaluate the action of the Fourier operator on the Laplace transform as

$$\begin{aligned} F_\nu(L(\varphi))(r) &= \int_0^\infty s^\nu H_\nu(sr) \int_0^\infty \varphi(t)e^{-st}dtds \\ &= \int_0^\infty \varphi(t) \int_0^\infty s^\nu H_\nu(sr)e^{-st}dsdt \\ &= \int_0^\infty \varphi(t)t^{-\nu-1} \int_0^\infty x^\nu H_\nu(xr/t)e^{-x}dxdt \\ &= \int_0^\infty \varphi(t)t^{-\nu-1}e^{-r/t}dt \\ &= \int_0^\infty \varphi(1/s)s^{\nu-1}e^{-sr}ds \\ &= L\left(\varphi(1/\cdot)(\cdot)^{\nu-1}\right). \end{aligned}$$

Then, again as formal operations,

$$\begin{aligned} F_\nu(L(\varphi(\cdot))) &= L\left(\varphi(1/\cdot)(\cdot)^{\nu-1}\right) \\ &= I_{\mu-\nu}L\left(\varphi(1/\cdot)(\cdot)^{\mu-1}\right) \\ &= I_{\mu-\nu}F_\mu(L(\varphi(\cdot)), \end{aligned}$$

$$\begin{aligned} F_\nu(F_\mu(L(\varphi(\cdot)))) &= F_\nu\left(L\left(\varphi(1/\cdot)(\cdot)^{\mu-1}\right)\right) \\ &= L\left(\varphi(\cdot)(\cdot)^{-\mu+1}(\cdot)^{\nu-1}\right) \\ &= I_{\mu-\nu}(L(\varphi(\cdot))), \end{aligned}$$

as expected. Note that this implies $F_\nu^2 = Id$ for all ν. All of these identities are valid at least on Laplace transforms of functions φ that vanish faster than any polynomial at zero and at infinity, but continuity arguments can be used to enlarge the scopes.

7.6.7 Identities for Transforms, Second Version

The previous section showed that the identity

$$F_\nu \circ F_\mu = I_{\mu-\nu}$$

holds for all $\mu, \nu \in \mathbb{R}$ on a small space of functions, and where I_α is an operator that roughly does α-fold integration for $\alpha \in \mathbb{R}$. We now want to make this more precise and explicit. In particular, we assert $F_\nu^2 = Id$ for all ν, which we only know for $\nu \in \frac{1}{2}\mathbb{Z}_{>-2}$. Furthermore, we want to use our explicit representations for the operators I_α.

To proceed towards inversion of the operator F_ν, let us start calculating the Fourier transform of the simplest compactly supported function, i.e.: a truncated power. The outcome is somewhat surprising, because we run into the function H_ν again:

Lemma 7.31 *For $\nu > \mu > -1$ and all $s, r \geq 0$ we have*

$$F_\mu \left(\frac{s^{-\nu}(s-\cdot)_+^{\nu-\mu-1}}{\Gamma(\nu-\mu)} \right)(r) \quad = \quad H_\nu(rs).$$

Proof: We directly calculate the assertion and use (10.36) from page 153. In detail,

$$\begin{aligned}
& F_\mu \left(\frac{s^{-\nu}(s-\cdot)_+^{\nu-\mu-1}}{\Gamma(\nu-\mu)} \right)(r) \\
= & \int_0^\infty t^\mu \frac{s^{-\nu}(s-t)_+^{\nu-\mu-1}}{\Gamma(\nu-\mu)} H_\mu(tr)dt \\
= & \frac{s^{-\nu}}{\Gamma(\nu-\mu)} \int_0^s t^\mu (s-t)^{\nu-\mu-1} H_\mu(tr)dt \\
= & \frac{s^{-\nu}}{\Gamma(\nu-\mu)} \int_0^s t^\mu (s-t)^{\nu-\mu-1} J_\mu(2\sqrt{rt})(rt)^{-\mu/2}dt,
\end{aligned}$$

and by substitution $t = su^2$, we get

$$\begin{aligned}
= & \frac{s^{-\nu}}{\Gamma(\nu-\mu)} \int_0^1 s^\mu u^{2\mu} s^{\nu-\mu-1}(1-u^2)^{\nu-\mu-1} J_\mu(2\sqrt{rs}u)(rsu^2)^{-\mu/2} 2sudu \\
= & \frac{2(rs)^{-\mu/2}}{\Gamma(\nu-\mu)} \int_0^1 u^{\mu+1}(1-u^2)^{\nu-\mu-1} J_\mu(2\sqrt{rs}u)du \\
= & \frac{2(rs)^{-\mu/2}}{\Gamma(\nu-\mu)} \frac{2^{\nu-\mu-1}\Gamma(\nu-\mu)}{(2\sqrt{rs})^{-\nu-\mu}} J_\nu(2\sqrt{rs}) \\
= & (\sqrt{rs})^{-\nu} J_\nu(2\sqrt{rs}) \\
= & H_\nu(rs).
\end{aligned}$$

□

We would like to invert the Fourier transform in the above assertion, but the decay of H_ν is not sufficient to see directly that F_μ is applicable at all. However, we can resort to specific tools from Special Functions to get

Lemma 7.32 *For $\nu > \mu > -1$ and all $r, s > 0$ we have*

$$(F_\mu H_\nu(s\cdot))(r) = \frac{s^{-\nu}(s-r)_+^{\nu-\mu-1}}{\Gamma(\nu-\mu)} \ .$$

Proof: The assertion is a consequence of the **Weber–Schafheitlin integral** (see (10.44) or [1] p. 487, 11.4.41) after substitutions of the type $t = s^2/2$. In detail, we have

$$\begin{aligned}
& \left(F_\mu H_\nu\left(\frac{u^2}{2}\,\cdot\right)\right)\left(\frac{r^2}{2}\right) \\
= & \int_0^\infty t^\mu H_\mu\left(\frac{r^2}{2}\,t\right) H_\nu\left(\frac{u^2}{2}\,t\right) dt \\
= & \int_0^\infty \left(\frac{s^2}{2}\right)^\mu \cdot s \cdot H_\mu\left(\frac{r^2}{2}\,\cdot\frac{s^2}{2}\right) H_\nu\left(\frac{u^2}{2}\,\cdot\frac{s^2}{2}\right) ds \\
= & \int_0^\infty 2^{-\mu} s^{2\mu+1} \left(\frac{rs}{2}\right)^{-\mu} \left(\frac{us}{2}\right)^{-\nu} J_\nu(us) ds \\
= & \ 2^\nu r^{-\mu} r^{-\nu} \int_0^\infty s^{\mu-\nu+1} J_\mu(rs) J_\nu(us) ds \\
= & \ \frac{2^\nu r^{-\mu} u^{-\nu} 2^{\mu-\nu+1} r^\mu (u^2-r^2)_+^{\nu-\mu-1}}{u^\nu \Gamma(\nu-\mu)} \\
= & \ \frac{1}{\Gamma(\nu-\mu)} \left(\frac{u^2}{2}\right)^{-\nu} \left(\frac{u^2}{2} - \frac{r^2}{2}\right)_+^{\nu-\mu-1} .
\end{aligned}$$

□

The above result can be used to derive the d–variate Fourier transform of the kernel

$$K(x,y) := H_\nu\left(\frac{c^2\|x-y\|_2^2}{4}\right) = \left(\frac{c\|x-y\|_2}{4}\right)^{-\nu} J_\nu\left(\frac{c\|x-y\|_2}{2}\right) \tag{7.33}$$

We have to rewrite this kernel as $\phi(r) = f(r^2/2)$ and get

$$f(t) = H_\nu\left(\frac{c^2}{2}t\right).$$

Then the above lemma yields

$$\left(F_{(d-2)/2} H_\nu\left(\frac{c^2}{2}\cdot\right)\right)\left(\frac{\|\omega\|_2^2}{2}\right) = \frac{\left(\frac{c^2}{2}\right)^{-\nu}\left(\frac{c^2}{2} - \frac{\|\omega\|_2^2}{2}\right)_+^{\nu-d/2}}{\Gamma(\nu-d/2+1)} \tag{7.34}$$

proving

Theorem 7.35 *If $2\nu + 2 > d$ holds, the scaled Bessel kernel (7.33) is positive definite on $\mathbb{R}^d$ and has the compactly supported Fourier transform (7.34) due to Theorems 7.22 and 7.7.*

We now know that $F_\nu \circ F_\nu = Id$ holds on Laplace transforms, on truncated powers, and on functions of the form $H_\mu(s\cdot)$. But before we generalize this to a larger class of functions, we generalize it to other F_μ operators:

Theorem 7.36 *Let $\nu > \mu > -1$. Then for all tempered radial test functions $f \in \mathcal{S}_{rad}$ we have*

$$F_\mu \circ F_\nu = I_{\nu-\mu} \tag{7.37}$$

where the integral operator I_α is given by

$$(I_\alpha f)(r) = \int_0^\infty f(s) \frac{(s-r)_+^{\alpha-1}}{\Gamma(\alpha)}\, ds, \qquad r > 0,\ \alpha > 0.$$

Proof: For any tempered radial test function $f \in \mathcal{S}_{rad}$ we evaluate $(F_\mu \cdot F_\nu)f(r)$ by means of Lemma 7.32 to obtain

$$\begin{aligned} & \int_0^\infty H_\mu(tr) t^\mu \int_0^\infty H_\nu(st) s^\nu f(s) ds dt \\ = & \int_0^\infty s^\nu f(s) \int_0^\infty t^\mu H_\mu(tr) H_\nu(ts) dt\ ds \\ = & \int_0^\infty s^\nu f(s) \cdot F_\mu(H_\nu(s\cdot))(r) ds \\ = & \int_0^\infty f(s) \frac{(s-r)_+^{\nu-\mu-1}}{\Gamma(\nu-\mu)}\, ds = (I_{\nu-\mu} f)(r). \end{aligned}$$

□

By the above theorems it is easy to see that

$$I_\alpha H_\nu = H_{\nu-\alpha}$$

for all $\alpha < \nu + 1$, generalizing (10.34).

7.6.8 Wendland's Functions

Due to a result of Askey [4] the radial truncated power function

$$A_\mu(\cdot) := (1 - \|\cdot\|_2)_+^\mu$$

is positive definite on $\mathbb{R}^d$ for $\mu \geq \lfloor d/2 \rfloor + 1$, because it has a strictly positive radial Fourier transform in this case.

Incomplete: add proof see [71]

Its radial form after substitution is $(1 - \sqrt{2r})_+^\mu$, and due to its finite support we can apply any F_ν operator for $\nu > -1$. We use the identity $F_{\nu+\alpha} = F_\nu \circ I_\alpha$ from (7.30) for this function and get

$$F_{\nu+k} A_\mu = F_\nu(I_k(A_\mu)),\ k \in \mathbb{N},$$

where the left-hand side is strictly positive whenever

$$\mu \geq \lfloor d/2 \rfloor + 1 + k. \tag{7.38}$$

Thus the function $I_k(A_\mu)$ is positive definite on $\mathbb{R}^d$ for the same range of parameters. Since the I_k operators preserve compact supports, the resulting functions

$$\psi_{\mu,k}(r) := I_k(A_\mu(r^2/2))$$

lead to compactly supported positive definite functions

$$\Psi_{\mu,k}(\cdot) = \psi_{\mu,k}(\|\cdot\|_2) = I_k(A_\mu(\|\cdot\|_2^2/2))$$

on $\mathbb{R}^d$ under the condition (7.38). Let us do a straightforward evaluation. This yields

$$\begin{aligned} I_k A_\mu(r) &= \int_0^\infty (1-\sqrt{2s})_+^\mu \frac{(s-r)_+^{k-1}}{(k-1)!} \\ &= \int_{\sqrt{2r}}^1 t(1-t)^\mu \frac{(t^2/2-r)_+^{k-1}}{(k-1)!} \\ &= \int_x^1 t(1-t)^\mu \frac{(t^2-x^2)_+^{k-1}}{(k-1)!2^{k-1}} \end{aligned} \tag{7.39}$$

for $0 \le r \le 1/2$ or $0 \le x = \sqrt{2r} \le 1$. If μ is an integer, the resulting function is a single polynomial of degree $\mu + 2k$ in the variable $x = \|\cdot\|_2$ on its support. The case $k = 1$ is particularly simple. We get the explicit representation

$$\begin{aligned} I_1 A_\mu(x^2/2) &= \int_x^1 t(1-t)^\mu dt \\ &= \frac{x(1-x)^{\mu+1}}{\mu+1} + \frac{(1-x)^{\mu+2}}{(\mu+1)(\mu+2)} \\ &= \frac{(1-x)_+^{\mu+1}}{(\mu+1)(\mu+2)} \left(1 + (\mu+1)x\right). \end{aligned}$$

The smallest possible integer μ for $d \le 3$ and $k = 1$ is $\mu = 3$, whence

$$I_1 A_3(x^2/2) = \frac{1}{20}(1-x)_+^4(1+4x).$$

In addition to $A_{k,\mu} := I_k A_\mu$ let us define

$$B_{k,\mu} := \int_x^1 (1-t)^\mu \frac{(t^2-x^2)_+^{k-1}}{(k-1)!2^{k-1}}$$

and split the integral defining $A_{k,\mu}$ via $t = (t-1)+1$ into

$$A_{k,\mu} = -B_{k,\mu+1} + B_{k,\mu}.$$

Then do integration by parts for $B_{k,\mu}$ and $k > 1$ to get

$$B_{k,\mu} = \frac{1}{\mu+1} A_{k-1,\mu+1}.$$

Thus we have the recurrence relation

$$A_{k,\mu} = -\frac{1}{\mu+2} A_{k-1,\mu+2} + \frac{1}{\mu+1} A_{k-1,\mu+1}.$$

Looking at our result for $k = 1$ we see that we can assume

$$A_{k,\mu}(x^2/2) = (1-x)^{\mu+k} C_{k,\mu}(x)$$

with the recursion

$$C_{k,\mu}(x) = \frac{(x-1)}{\mu+2} C_{k-1,\mu+2}(x) + \frac{1}{\mu+1} C_{k-1,\mu+1}(x),$$

for $k \geq 1$, starting with

$$C_{0,\mu}(x) = 1.$$

Thus the polynomials $C_{k,\mu}$ have degree k with a positive leading coefficient. The number of continuous derivatives of $A_{k,\mu}(x^2/2)$ at $x = 1$ thus is $\mu + k - 1 \geq 2k + \lfloor d/2 \rfloor \geq 2k$. To get the number of derivatives at zero we apply the binomial theorem to the last factor in the integrand. Then

$$\begin{aligned}
A_{k,\mu}(x^2/2) &= \sum_{j=0}^{k-1} \left(\begin{array}{c} k-1 \\ j \end{array} \right) \frac{(-1)^j x^{2j}}{(k-1)!} \int_x^1 t(1-t)^\mu t^{2k-2-2j} dt \\
q_{\mu,k-j}(x) &:= \int_x^1 t(1-t)^\mu t^{2k-2-2j} dt \\
&= q_{\mu,k-j}(1) - \int_0^x t(1-t)^\mu t^{2k-2-2j} dt \\
&= q_{\mu,k-j}(1) - \frac{x^{2k-2j}}{2k-2j} + \text{ higher-order terms}
\end{aligned}$$

shows that the first odd monomial occurring in $A_{k,\mu}(x^2/2)$ cannot have an exponent smaller than $2k+1$. Thus the function has $2k$ continuous derivatives at zero, and we get $2n-1 = 2k+1$ in the context of Wendland's functions. In terms of continuity requirements, we get overall C^{2k} continuity at a minimal degree $\mu + 2k = \lfloor d/2 \rfloor + 3k + 1$, and Wendland proves in [70] that this degree is minimal, if we ask for a single polynomial piece on $[0, 1]$ that induces a positive definite radial function which is C^{2k} and positive definite on $\mathbb{R}^d$. Note that the order of smoothness at the boundary of the support is $\lfloor d/2 \rfloor$ larger than the smoothness at zero, which has a positive effect on the visual appearance of the reproduced functions.

We end this by giving the C^4 case for all dimensions d, where $\mu = \lfloor d/2 \rfloor + 3$:

$$\begin{aligned}
& A_{2,\mu}(x^2/2) \\
= \; & \frac{(1-x)_+^{\mu+2}}{(\mu+1)(\mu+2)(\mu+3)(\mu+4)} (x^2(\mu+1)(\mu+3) + 3x(\mu+2) + 3)
\end{aligned}$$

and the most frequent case for $d \leq 3$ is

$$A_{2,4}(x^2/2) = \frac{(1-x)_+^6}{1680}(35x^2 + 18x + 3).$$

The Fourier transforms are

$$F_\nu I_k A_\mu = F_{\nu+k} A_\mu$$

and thus for $r = x^2/2$ of the form

$$\begin{aligned}
F_{\nu+k} A_\mu(r) &= \int_0^{1/2} (1 - \sqrt{2s})^\mu s^{\nu+k} H_{\nu+k}(rs) ds \\
&= \frac{x^{-\nu-k}}{2^{\nu+k}} \int_0^1 (1-t)^\mu t^{\nu+k+1} J_{\nu+k}(xt) dt \\
&= \frac{x^{-\mu-2\nu-2k-2}}{2^{\nu+k}} \int_0^x (x-u)^\mu u^{\nu+k+1} J_{\nu+k}(u) du.
\end{aligned}$$

Due to a result of Gasper [23], the above integral can be written as a positive sum of squares of Bessel functions, at least in the odd-dimensional case $d = 2n - 1$ with $\mu = n + k + 1$, which leads to $\nu = m - 1/2$ and $\mu = m + 1$ for $m = n + k \geq n$. Results of Wendland [70] then imply the asymptotic behaviour

$$F_\nu I_k A_\mu(r^2/2) = F_{\nu+k} A_\mu(r^2/2) \geq c r^{-d-2k-1}$$

for the necessary values of μ from (7.38).

7.7 Conditionally Positive Definite Kernels

We now go over to the treatment of general unconditionally positive definite kernels. To do this, we shall introduce Fourier transforms in a somewhat more general way that will later save us quite some work. The direct attack is impossible, because some of the most important conditionally positive definite functions on $\mathbb{R}^d$ are radial functions $\Phi(\cdot) = \phi(\|\cdot\|_2)$ that grow towards infinity, e.g.: thin-plate splines $\phi(r) = r^2 \log r$ or multiquadrics $\phi(r) = \sqrt{r^2 + c^2}$. These do not have classical Fourier transforms, but since they grow at most polynomially, they induce functionals on the Schwartz space $\mathcal{S}$. Thus they have generalized Fourier transforms defined via the Fourier transforms of the functionals that they induce on $\mathcal{S}$. These generalized Fourier transforms are not straightforward to handle and require quite some machinery from distribution theory.

We go a different way by picking a very specific set of assumptions to start with, and then we can work our way without distributions. We do not even assume Φ to be a conditionally positive definite function; this will be a consequence of our assumptions and lead to an important technique to prove conditional positive definiteness for specific examples.

In what follows, recall the notation used in section 4.3, but here we fix the space $\mathbb{P}$ to be the space $\mathbb{P}_m^d$ of d–variate polynomials of order at most m. Furthermore, we use the notion of Fourier transforms of functionals as provided in section 10.5.

Assumption 7.40 *Let $\Phi : \mathbb{R}^d \to \mathbb{R}$ be even and continuous. Furthermore, let there be a continuous nonnegative function*

$$\widehat{\Phi} : \mathbb{R}^d \setminus \{0\} \to \mathbb{R}$$

which is positive on at least an open set. It may possibly have an algebraic singularity

$$\widehat{\Phi}(\omega) = \mathcal{O}(\|\omega\|^{-d-\beta_0}) \tag{7.41}$$

with some real value β_0 for ω near zero, and it must have the behavior

$$\widehat{\Phi} \in L_1 \text{ near infinity.} \tag{7.42}$$

Then define $m := \max(0, \lfloor \beta_0 \rfloor) \geq 0$ to get the restriction

$$\beta_0 < 2m \tag{7.43}$$

that will often occur later. Finally, let the usual bilinear form on L be representable by

$$(\lambda_{a,X}, \lambda_{b,Y})_\Phi = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{\Phi}(\omega) \sum_{j=1}^{M} \sum_{k=1}^{N} a_j \overline{b_k} e^{i(x_j - y_k)\cdot\omega} d\omega, \tag{7.44}$$

where the functionals $\lambda_{a,X} \in L$ satisfy the **moment conditions** *(4.18) in the form*

$$\lambda_{a,X}(I\!P_m^d) = \{0\}, \tag{7.45}$$

and thus we may use the notation $(I\!P_m^d)_{I\!R^d}^{\perp}$ for L.

Lemma 7.46 *The functionals $\lambda_{a,X} \in L$ have Fourier transforms*

$$\hat{\lambda}_{a,X}(\omega) = p_{a,X}(\omega) = \sum_{j=1}^{N} a_j e^{-ix_j^T\omega}$$

with zeros of order at least m in the origin.

Proof: Since we have (7.45), we can use Example 10.21 to get our result. □

Theorem 7.47 *Under the above assumptions the function $\Phi(x-y)$ is conditionally positive definite of order $\geq m$ on $I\!R^d$.*

Proof: From the previous lemma we know that the functionals $\lambda_{a,X} \in L$ have Fourier transforms with zeros of order at least m in the origin. Thus the integrand in (7.44) is of order $\mathcal{O}(\|\omega\|^{2m-d-\beta_0})$ near zero, and the integral is well-defined due to (7.43) and (7.42). Nonnegativity of $\widehat{\Phi}$ proves that the bilinear form is positive semidefinite. The rest is as in the proofs of Theorems 7.6 and 10.7. □

The reader should be aware that we did not assume $\widehat{\Phi}$ to be the usual Fourier transform. We thus cannot use equations (10.6) or (10.11), but we have the general identity

$$\sum_{j=1}^{M}\sum_{k=1}^{N} a_j b_k \Phi(x_j - y_k) = (2\pi)^{-d/2} \int_{I\!R^d} \widehat{\Phi}(\omega) \sum_{j=1}^{M}\sum_{k=1}^{N} a_j \overline{b_k} e^{i(x_j-y_k)\cdot\omega} d\omega.$$

that is identical to (7.44) and is valid for all functionals in L due to Assumption 7.40. It will nicely serve as a substitute for (10.11), but note that it does not allow single point-evaluation functionals in the left-hand side.

7.8 Examples

We now present special cases of (7.44) for radial kernels

$$K(x,y) = \Phi(x-y) = \phi(\|x-y\|_2),\ x,y \in I\!R^d$$

where we get a resulting generalized d–variate Fourier transform in radial form which we denote by $\hat{\phi}$.

The first example generalizes the inverse multiquadrics to general **multiquadrics**. If we set

$$\phi(r) := (c^2 + r^2)^{\beta/2},\ r \geq 0,\ c > 0,\ \beta \in I\!R \setminus 2I\!N_0$$

we get the function

$$\hat{\phi}(s) = \frac{2^{1+\beta/2}}{\Gamma(-\beta/2)} \left(\frac{s}{c}\right)^{-\frac{\beta+d}{2}} K_{\frac{\beta+d}{2}}(cs),\ s \in I\!R$$

while the order of conditional positive definiteness turns out to be

$$m = \max(0, \lceil \beta/2 \rceil).$$

Note that for positive β the denominator has the sign $(-1)^{\lceil \beta/2 \rceil}$. Thus we have to multiply ϕ for positive β with this factor to get a conditionally **positive** definite function.

The proof idea is quite nice. Each side of the standard Fourier transform identity (7.44), including the quadratic form and holding first for negative β is proven to be an analytic function of β. Under the additional moment conditions, both sides also make sense for general β, and they can be connected by analytic continuation with the case for negative β by a detour over complex β avoiding passing through the origin. Thus the Fourier transform equation also holds for the other β.

The next example concerns the power functions, and this is the limit of the previous case for $c \to 0$. If we set

$$\phi(r) := (-1)^{\lceil \beta/2 \rceil} r^\beta, \ r \geq 0, \ \beta \in I\!R_{>0} \setminus 2I\!N$$

we get the positive function

$$\hat{\phi}(s) = \frac{2^{\beta+d/2}\Gamma((\beta+d)/2)}{(-1)^{\lceil \beta/2 \rceil}\Gamma(-\beta/2)} s^{-\beta-d}, \ s \in I\!R$$

while the order of conditional positive definiteness turns out to be

$$m = \lceil \beta/2 \rceil.$$

This proof works from the previous case for positive β by letting c tend to zero, checking carefully how the Bessel function interacts with the premultiplied rational function.

The final case is connected to β being an even integer. If we set

$$\phi(r) := (-1)^{k+1} r^{2k} \log r, \ r \geq 0, \ k \in I\!N$$

we get

$$\hat{\phi}(s) = 2^{2k-1+d/2} s^{-2k-d}, \ s \in I\!R$$

while the order of conditional positive definiteness turns out to be

$$m = k + 1.$$

The last two cases are called **polyharmonic**, because they are homogeneous solutions of a power of the Laplacian. This is due to the fact that their generalized Fourier transform is a plain negative power. The last case is called the **thin–plate spline**.

7.9 Connection to $L_2(I\!R^d)$

We now go back to Definition 4.41 of the native space via (4.42) and Corollary 4.43as

$$\mathcal{N}_{K,\Xi} := I\!P^d_m + \overline{G} = I\!P^d_m + \overline{F}.$$

This definition is very general, and we want to re–express the native space via Fourier transforms. We do this using a detour over weighted L_2 spaces.

The space L of section 4.3 consists of functionals $\lambda_{a,X}$ with the moment condition $(a, X) \in M$. These functionals have Fourier transforms $\hat{\lambda}_{a,X}$ with the property

$$\lambda_{a,X}(f) = (2\pi)^{-d/2}(\hat{f}, \hat{\lambda}_{a,X})_{L_2(\mathbb{R}^d)}, \quad (\hat{\lambda}_{a,X})(\omega) = \sum_{j=1}^{N} a_j e^{-ix_j^T\omega}.$$

Assumption 7.40 makes sure that the mapping

$$\mathcal{L} \ : \ \lambda \mapsto \widehat{\lambda}\sqrt{\widehat{\Phi}}, \ \ L = (\mathbb{P}_m^d)_{\mathbb{R}^d}^{\perp} \to L_2(\mathbb{R}^d)$$

is well-defined. Indeed, the function $\mathcal{L}(\lambda)$ is in L_2 near infinity due to (7.42), and it is continuous around zero due to (7.43), since $\widehat{\lambda}$ has a zero of order at least m at the origin.

With the results of the previous section, (7.44) takes the form

$$(\lambda_{a,X}, \lambda_{b,Y})_{\Phi} = (2\pi)^{-d/2}(\mathcal{L}\lambda_{a,X}, \mathcal{L}\lambda_{b,Y})_{L_2(\mathbb{R}^d)}. \tag{7.48}$$

Theorem 7.49 *Let Assumption 7.40 be satisfied, and let m be minimal with respect to (7.43). Then the map $\mathcal{L}$ extends by continuity to clos(L), and it yields an isometry between clos(L) and* all *of $L_2(\mathbb{R}^d)$.*

Proof: It is evident from (7.44) that $\mathcal{L}$ is isometric, and thus $\mathcal{L}$ extends to clos (L) by continuity. But the density of $\mathcal{L}($ clos $(L))$ in $L_2(\mathbb{R}^d)$ does not follow from abstract Hilbert space arguments. We thus need an additional analytic argument. We first prove the assertion for continuous $\widehat{\Phi}$ with $\widehat{\Phi} > 0$ on $\mathbb{R}^d \setminus \{0\}$.

Let some function $f \in L_2(\mathbb{R}^d)$ and some $\varepsilon > 0$ be given. Then there is a compactly supported C^∞ function $g \in L_2(\mathbb{R}^d)$ such that $\|f - g\|_2 \leq \varepsilon$ due to Lemma 10.4. Now define $\widehat{u} := g/\sqrt{\widehat{\Phi}}$ on $\mathbb{R}^d$, where the (possible) singularity of $\widehat{\Phi}$ at zero does no harm. Clearly $\hat{u}$ is continuous and compactly supported, thus in $L_2(\mathbb{R}^d)$ and u is band–limited, of exponential type, and in $L_2(\mathbb{R}^d)$. We now invoke the multivariate sampling theorem to recover u exactly from its function values on a grid in $\mathbb{R}^d$ with spacing h, where h is sufficiently small and related to the support of $\widehat{u}$.

Thus we have

$$u(x) = \sum_{j \in \mathbb{Z}^d} u(jh) \ \mathrm{Sinc}_d\left(\frac{x - jh}{h}\right), \quad x \in \mathbb{R}^d$$

where

$$\mathrm{Sinc}_d(x_1, \ldots, x_d) = \prod_{j=1}^{d} \frac{\sin \pi x_j}{\pi x_j},$$

and

$$\widehat{u}(\omega) = \sum_{j \in \mathbb{Z}_d} u(jh) e^{ihj\cdot\omega}, \qquad \omega \in \mathbb{R}^d$$

has the form $\widehat{u} = \widehat{\lambda_u}$ for the functional

$$\lambda_u(v) = \sum_{j \in \mathbb{Z}^d} v(jh)u(jh).$$

We now have to make sure that $\lambda_u \in \mathrm{clos}\,(L)$. If this is done, we are finished because of $\mathcal{L}(\lambda_u) = g$ and

$$\|f - \sqrt{\widehat{\Phi}}\widehat{\lambda u}\|_2 = \|f - g\|_2 \leq \varepsilon.$$

For all $p \in I\!P_m^d$ we have to show that $\lambda_u(p) = 0$. By a standard argument in Fourier analysis this requires a zero of order at least m of $\widehat{u}$ at zero. But our assumption (7.41) on $\widehat{\Phi}$ and the minimality of m in (7.43) imply that $\widehat{u}$ has a zero of order at least

$$\frac{1}{2}(d + \beta_0) > \frac{1}{2}(d + 2m - 2) = m - 1 + \frac{d}{2},$$

thus of order $\geq m$.

We then evaluate the norm formally as

$$\|\lambda_u\|_\Phi^2 = \|\sqrt{\widehat{\Phi}} \cdot \widehat{\lambda_u}\|_2^2 = \|\sqrt{\widehat{\Phi}}\widehat{u}\|_2^2 = \|g\|_2^2 < \infty.$$

Now we can proceed to prove that λ_u lies in $\mathrm{clos}\,(L)$ by defining the function

$$f_{\lambda_u}(x) := (\lambda_u, \delta_{x,\Xi})_\Phi, \ x \in I\!R^d$$

via the explicit form of the inner product, and using the finiteness of the norm $\|\lambda_u\|_\Phi$ to show that the definition is valid. Then for all $\lambda_{Y,N,\beta} \in L$ we get

$$\lambda_{Y,N,\beta}(f_{\lambda_u}) = (\lambda_u, \lambda_{Y,N,\beta})_\Phi$$

and this proves that $f_{\lambda_u} \in \mathcal{F}$. Finally, we get $\lambda_u = F^{-1}(f_{\lambda_u})$ by checking

$$\begin{aligned}(\lambda_u, \lambda_{Y,N,\beta})_\Phi &= \lambda_{Y,N,\beta}(f_{\lambda_u}) \\ &= (\lambda_{Y,N,\beta}F^{-1}f_{\lambda_u}))_\Phi\end{aligned}$$

for all $\lambda_{Y,N,\beta} \in L$, and this concludes the proof in case of $\widehat{\Phi} > 0$.

Now let $\widehat{\Phi}$ be positive up to a set of Lebesgue measure zero. We cover the set of zeros by intervals I_k, where k varies over some index set K and the total area $\sum_k |I_k|$ is less than some given δ. Now let $\widehat{\Phi}_\delta(\omega) \geq \widehat{\Phi}(\omega)$ be a strictly positive continuous function that differs from $\widehat{\Phi}$ only on the I_k. Then $\widehat{\Phi}_\delta$ will also satisfy our assumptions, and we can use (7.44) in the form

$$(\mu, \lambda)_{\Phi_\delta} := (2\pi)^{-d/2} \int_{I\!R^d} \widehat{\Phi}_\delta(\omega)\widehat{\lambda}(\omega)\overline{\widehat{\mu}(\omega)}d\omega$$

as a definition of an inner product, but we do not need Φ_δ explicitly.

Now we approximate a given $f \in L_2(I\!R^d)$ by some $\sqrt{\widehat{\Phi}_\delta} \cdot \widehat{\lambda}$ up to $\varepsilon/2$ in the L_2 norm, picking a suitable λ for each δ and ε. Then

$$\|f - \sqrt{\widehat{\Phi}}\widehat{\lambda}\|_2 \leq \|f - \widehat{\lambda}\sqrt{\widehat{\Phi}_\delta}\|_2 + \|\widehat{\lambda}(\sqrt{\widehat{\Phi}_\delta} - \sqrt{\widehat{\Phi}})\|_2$$

and

$$\begin{aligned}\|\widehat{\lambda}(\sqrt{\widehat{\Phi}_\delta}-\sqrt{\widehat{\Phi}})\|_2^2 &= \|\widehat{\lambda}\cdot\sqrt{\widehat{\Phi}_\delta}(1-\sqrt{\widehat{\Phi}/\widehat{\Phi}_\delta})\|_2^2 \\ &\le \sum_k \int_{I_k} |\widehat{\lambda}(\omega)|^2\widehat{\Phi}_\delta(\omega)d\omega.\end{aligned}$$

The full integral

$$\int_{I\!R^d} |\widehat{\lambda}(\omega)|^2\widehat{\Phi}_\delta(\omega)d\omega = \|\widehat{\lambda}\cdot\sqrt{\widehat{\Phi}_\delta}\|_2^2$$

can be bounded independent of δ, because it approximates $\|f\|_2^2$. Thus we are able to pick δ small enough to guarantee

$$\sum_k \int_{I_k} |\widehat{\lambda}(\omega)|^2\widehat{\Phi}_\delta(\omega)d\omega \le \varepsilon/2$$

yielding an overall bound $\|f-\sqrt{\widehat{\Phi}}\widehat{\lambda}\|_2 \le \varepsilon$. □

7.10 Characterization of Native Spaces

We now can re–express the native space $\mathcal{N}_{K,\Xi} := I\!P^d_m + \overline{G} = I\!P^d_m + \overline{F}$. via Fourier transforms.

Theorem 7.50 *The native space $\mathcal{N}_{K,\Xi} := I\!P^d_m + \overline{G}$ for a conditionally positive definite function of order m on $I\!R^d$ satisfying Assumption 7.40 coincides with the space of all functions f on $I\!R^d$ that can be written as*

$$f_h(x) = (2\pi)^{-d/2}\int_{I\!R^d} \widehat{h}(\omega)\sqrt{\widehat{\Phi}(\omega)}\left(e^{ix\cdot\omega} - \sum_{j=1}^{Q} p_j(x)e^{i\xi_j\cdot\omega}\right)d\omega \tag{7.51}$$

plus polynomials from $I\!P^d_m$ and where $\hat{h} \in L_2(I\!R^d)$. The above functions are spanning the space $\overline{G}$. The bilinear form on $\overline{G}$ can be rewritten as

$$(f_g, f_h)_\Phi = (2\pi)^{-d/2}(g,h)_{L_2(I\!R^d)}. \tag{7.52}$$

Proof: We first focus on (7.51). Starting with an arbitrary $h \in L_2(I\!R^d)$ and a fixed $I\!P^d_m$-unisolvent set $\Xi \subset I\!R^d$, we mimic the technique of Riesz maps to define a function

$$f_h(x) := (\widehat{h}, \mathcal{L}\delta_{(x)})_{L_2(I\!R^d)}. \tag{7.53}$$

This is (7.51). Since

$$\lambda f_h = (\widehat{h}, \mathcal{L}\lambda)_{L_2(I\!R^d)}$$

follows easily from (7.52) for all $\lambda \in L$, we can transform this equation further into

$$\begin{aligned}\lambda f_h &= (\widehat{h}, \mathcal{L}\lambda)_{L_2(I\!R^d)} \\ &= (\mathcal{L}^{-1}\widehat{h}, \lambda)_\Phi.\end{aligned}$$

By the previous section, (7.48) with Theorem 7.49 yields that $\mathcal{L}^{-1}$ maps $L_2(I\!R^d)$ isometrically back to $\overline{L}$. But $\overline{L}$ is isometric to $\overline{G}$ via the extension $\mathcal{R}$ of the Riesz map $R : L \to G$ we had in section 4.5. Thus the above identity can be extended to

$$\begin{aligned}\lambda f_h &= (\widehat{h}, \mathcal{L}\lambda)_{L_2(I\!R^d)} \\ &= (\mathcal{L}^{-1}\widehat{h}, \lambda)_\Phi \\ &= (\mathcal{R}\mathcal{L}^{-1}\widehat{h}, \mathcal{R}\lambda)_\Phi \text{ for all } \lambda \in L\end{aligned}$$

proving

$$f_h = \mathcal{R}\mathcal{L}^{-1}\widehat{h} \in \overline{G}.$$

□

By (??) we also get

Corollary 7.54 *The mapping*

$$\mathcal{F} = \mathcal{R}\mathcal{L}^{-1} \ : \ h \mapsto f_h$$

is isometric between $L_2(\mathbb{R}^d)$ *and* $\overline{G}$. □

Note that we avoided to use the Fourier transform of f_h. In case that $\sqrt{\hat{\Phi}}\hat{h} =: g_h$ is an absolutely integrable function, the right–hand side of (7.51) is

$$f_h(x) = g_h^\vee(x) - \sum_{j=1}^{Q} p_j(x) g_h^\vee(\xi_j)$$

such that we see that a polynomial variation of f_h has a Fourier transform which is $\sqrt{\hat{\Phi}}\hat{h}$.

But we can also work via the $\overline{F}$ part of the native space. It is the closure of all functions

$$f_{a,X}(x) := \lambda_{a,X}^t \Phi(x - t),$$

and if the functional is such that Fourier tyransforms can be taken, we get

$$\hat{f}_{a,X} = \hat{\Phi}\hat{\lambda}_{a,X} = \sqrt{\hat{\Phi}}\mathcal{L}\lambda_{a,X}$$

such that

$$\frac{\hat{f}_{a,X}}{\sqrt{\hat{\Phi}}} = \mathcal{L}\lambda_{a,X} \in L_2(\mathbb{R}^d).$$

This can also be written as

$$\mathcal{R}(\lambda)^\wedge = \hat{\lambda} = \sqrt{\hat{\Phi}}\mathcal{L}(\lambda)$$

if all transforms exist, and this is a third reason to define

$$\hat{f}_h := \sqrt{\hat{\Phi}}\hat{h}$$

as a generalized Fourier transform of f_h, but the use of standard Fourier transform equations is forbidden without additional argumants along the above lines.

7.11 Connection to Sobolev Spaces

To make error bounds applicable, we need inclusion theorems for native spaces in Sobolev spaces. Since polynomials are not contained in global Sobolev spaces, we can only expect the $\overline{G}$ part of the native space to be contained in a global Sobolev space $W_2^\tau(\mathbb{R}^d)$, while polynomials are always contained in local Sobolev spaces $W_2^\tau(\Omega)$ for bounded domains. Thus we cannot work as easily as in the unconditionally positive definite case.

Let us check the differentiability of the functions from (7.51). Under sufficient regularity of $\sqrt{\hat{\Phi}}\hat{h} =: g_h$, we take a derivative D^α of f_h of order α with $|\alpha| \geq m$. It will have Fourier transform $(i\omega)^\alpha \sqrt{\hat{\Phi}}(\omega)\hat{h}(\omega)$ and we check when it is well–defined and globally in L_2. This is the case when $\|\omega\|^{2|\alpha|}\hat{\Phi}(\omega)$ is globally bounded. Since we only use L_2 arguments for this result, we can ignore the additional regularity assumptions on $\sqrt{\hat{\Phi}}\hat{h} =: g_h$ by an additional density argument.

Near zero, the boundedness of $\|\omega\|^{2|\alpha|}\hat{\Phi}(\omega)$ follows for

$$2|\alpha| \geq d + \beta_0$$

because

$$\|\omega\|^{2|\alpha|}\hat{\Phi}(\omega) \leq \mathcal{O}(\|\omega\|^{2|\alpha|-d-\beta_0}).$$

Near infinity, we have not yet made any assumptions about the behavior of $\hat{\Phi}$. For simplicity, we mimic (7.41) as

$$\hat{\Phi}(\omega) \leq \mathcal{O}(\|\omega\|^{-d-\beta_\infty}) \text{ near } \infty. \tag{7.55}$$

Then a sufficient condition for boundedness at infinity is

$$2|\alpha| \leq d + \beta_\infty.$$

Altogether, we get that the derivative $D^\alpha f_h$ is defined and globally in L_2 at least for

$$|\alpha| \geq m,\ \beta_0 + d \leq 2|\alpha| \leq \beta_\infty + d.$$

This is quite sufficient for the multiquadric case, because there β_∞ is arbitrarily large. For the thin–plate spline and the polyharmonic splines, we have $\beta_0 = \beta_\infty$ and see that we can still work with generalized derivatives of order $m = |\alpha| = d/2 + \beta_\infty/2 = d/2 + \beta_0/2 > d/2$ if this is an integer.

In all of these cases we can take the maximum possible $|\alpha|$ and get convergence of interpolants like $h^{|\alpha|-d/2}$ in the L_∞ norm when the data are from a function in the native space. This yields infinite order for the multiquadrics and convergence like $h^{\beta_0/2} = h^{\beta_\infty/2}$ for the thin–plate spline and the polyharmonic spline provided that $d + \beta_0 = d + \beta_\infty$ is even.

incomplete here...

8 Stability Theory

It would be very desirable to have recovery methods with small errors and good stability. However, these two goals cannot be met at the same time, since there is a connection between them that implies bad stability whenever the a-priori error bound is very small.

8.1 Uncertainty Relation

Let us look at this connection in the Lagrange interpolation setting and consider optimal recovery of a function $g \in \mathcal{N}$ in a Hilbert space $\mathcal{N}$ which is the native space of a reproducing kernel K on a domain $\Omega \subset I\!R^d$. This recovery should use data $g(x_j), 1 \leq j \leq M$ for a finite set $X = \{x_1, \ldots, x_M\} \subset \Omega \subset I\!R^d$.

We add a variable point x to X and define the kernel matrix

$$A_{x,X} := \begin{pmatrix} \Phi(x,x) & \Phi(x,x_1) & \ldots & \Phi(x,x_M) \\ \Phi(x_1,x) & \Phi(x_1,x_1) & \ldots & \Phi(x_1,x_M) \\ \vdots & \vdots & & \vdots \\ \Phi(x_M,x) & \Phi(x_M,x_1) & \ldots & \Phi(x_M,x_M) \end{pmatrix}$$

and the vector

$$(u_X^*)^T(x) := (1, -u_1^*(x), \ldots, -u_M^*(x))^T \in I\!R^{M+1}$$

with the Lagrange basis of (3.4) and get the special form

$$\begin{aligned} P_X^2(x) &= K(x,x) - 2\sum_{j=1}^N u_j^*(x)K(x_j,x) + \sum_{j,k=1}^N u_j^*(x)u_k^*(x)K(x_j,x_k) \\ &= (u_X^*)^T(x)A_{x,X}(u_X^*)(x) \\ &\geq \sigma(A_{x,X})\left(1 + \sum_{j=1}^M |u_j^*(x)|^2\right) \\ &\geq \sigma(A_{x,X}) \end{aligned} \tag{8.1}$$

of the power function (3.8), where $\sigma(A_{x,X})$ is the minimal eigenvalue of $A_{x,X}$. Note that both sides are continuous functions of x and X (or Λ standing for X) that vanish whenever x tends to points in X.

Theorem 8.2 *The error of kernel interpolation can only be small if the condition of the kernel matrix is large. In particular,*

$$P_X^2(x) \geq \sigma(A_{x,X})$$

holds for the power function P_X^2 in terms of the smallest eigenvalue $\sigma(A_{x,X})$ of the kernel matrix $A_{x,X}$.

We can call the above observation an **Uncertainty Principle** or a **Tradeoff principle**.

The interpretation of the above result is as follows. Assume we have a recovery process with a very good error bound (3.7) via the power function. Then $A_{x,X}$ must have a very small eigenvalue. The largest eigenvalue of $A_{x,X}$ can only be as large as a constant times N, thus it is not very relevant for the condition of $A_{x,X}$, which is the quotient of the largest by the smallest eigenvalue, if the condition is taken in the spectral norm. Thus the condition of $A_{x,X}$ is large whenever the recovery error is small. But $A_{x,X}$ is itself a kernel matrix, if we view x as the "'next" interpolation point. Or, when we change the meaning of x and X somewhat, we can rewrite the above result as

$$\min_{1 \leq j \leq N} P_{X \setminus x_j}^2(x_j) \geq \sigma(A_X),$$

bounding the smallest eigenvalue of a kernel matrix via the "leave–one–out" power function.

We now can give some hints to the results that follow in later sections. The Uncertainty Relation in the form (8.1) suggests to bound P^2 from above and σ from below, in order to have both upper bounds on the attainable error and on the numerical stability, measured by $1/\sigma$. We have seen in the previous chapter that upper bounds for P^2 take the form

$$P^2_{X_h}(x) \le F(h_{X,\Omega}) \text{ for all } x \in \Omega \tag{8.3}$$

where F is a monotonic function of the fill distance $h_{X,\Omega}$ defined in (5.2). On the other hand, the lower bounds for σ which we shall prove in this chapter, will be of the form

$$\sigma(A_X) \ge G(q_X) \text{ for all } X = \{x_1, \dots, x_M\} \subset \Omega \tag{8.4}$$

with the **separation distance**

$$q := q_X := \min_{1 \le i \ne j \le M} \|x_i - x_j\|_2. \tag{8.5}$$

For gridded data on $\epsilon\mathbb{Z}^d \cap \Omega$ we can roughly expect $h_{X,\Omega} = q_X\sqrt{d}$, and then the Uncertainty Relation necessarily implies

$$F(t\sqrt{d}) \ge G(t) \tag{8.6}$$

for all $t \ge 0$. This allows to check the quality of the bounds (8.3) and (8.4), since the lowest possible bounds F and the largest possible bounds G must necessarily satisfy (8.6) and are optimal, if they turn (8.6) into an equality. This opens the race for optimal bounds of the form (8.3) and (8.4), and this text will describe the current state-of-the-art. To cut the story short, we shall prove that F and G just differ by a factor in case of kernels of finite smoothness, i.e. (8.6) is extended to

$$F(t\sqrt{d}) \ge G(t) \ge C \cdot F(c \cdot t) \tag{8.7}$$

for all $t \ge 0$, proving that the square of the power function and the minimal eigenvalue of the kernel matrix are roughly proportional in all cases of finite smoothness.

8.2 Lower Bounds on Eigenvalues

This section uses Fourier transform techniques to prove results concerning the condition of the matrices that occur in the basic equations (3.2) and (4.10) for optimal recovery. This requires upper bounds for the largest, and lower bounds for the smallest eigenvalue. We start with the latter and restrict ourselves to the Lagrange case. The bounds should (if possible) should neither depend on the specific data locations $X = \{x_1, \dots, x_M\}$, nor on the number M of data points, but rather on certain real-valued quantities like the separation distance (8.5).

We generalize the technique of Narcowich and Ward [44] [45] [46] for calculating stability bounds, but we introduce Fourier transforms right from the start, which makes it much easier to treat large values of m, the order of conditional positive definiteness of Φ.

The starting point is that any conditionally positive definite function Φ of order m satisfying Assumption 7.40 allows the formula

$$\sum_{j=1}^{M}\sum_{k=1}^{M} \alpha_j\alpha_k\Phi(x_j - x_k) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{\Phi}(\omega) \left| \sum_{j=1}^{M} \alpha_j e^{ix_j\cdot\omega} \right|^2 d\omega \tag{8.8}$$

for all $I\!P_m^d$–nondegenerate sets $X = \{x_1, \ldots, x_M\}$ and all vectors $\alpha \in I\!R^M$ such that $\lambda_{X,M,\alpha}$ is a functional that annihilates $I\!P_m^d$. This is just another way of writing (7.44).

The left–hand side of (8.8) is the quantity $\alpha^T A_{X,\Phi}\alpha$ that we want to bound from below, and we can do this by any minorant $\widehat{\Psi}$ on $I\!R^d \setminus \{0\}$ of $\widehat{\Phi}$ that satisfies

$$\widehat{\Phi}(\omega) \geq \widehat{\Psi}(\omega) \qquad \text{on } I\!R^d \setminus \{0\} \tag{8.9}$$

and that itself leads to a similar quadratic form

$$\sum_{j=1}^{M}\sum_{k=1}^{M} \alpha_j \alpha_k \widehat{\Psi}(x_j - x_k) = (2\pi)^{-d/2} \int_{I\!R^d} \widehat{\Psi}(\omega) \left| \sum_{j=1}^{M} \alpha_j e^{ix_j \cdot \omega} \right|^2 d\omega \tag{8.10}$$

for another basis function $\widehat{\Psi}$ and a weaker constraint on $\alpha \in I\!R^M$ (or none at all). Furthermore, there should be an easy lower bound

$$\alpha^T A_{X,\Psi}\alpha \geq \sigma \|\alpha\|_2^2$$

for the left–hand side $\alpha^T A_{X,\Psi}\alpha$ of (8.10). Then clearly for all $\alpha \in I\!R^M$ that are admissible,

$$\alpha^T A_{X,\Phi}\alpha \geq \alpha^T A_{X,\Psi}\alpha \geq \sigma \|\alpha\|_2^2,$$

as required. The basic trick of Narcowich and Ward now is to make $A_{X,\Psi}$ diagonally dominant, while $\widehat{\Psi}$ is obtained by chopping off $\widehat{\Phi}$ appropriately near infinity.

Before we proceed any further, here is the main result:

Theorem 8.11 *Let Φ be a conditionally positive definite function on $I\!R^d$ that satisfies Assumption 7.40. Furthermore, let $X = \{x_1, \ldots, x_M\} \subset I\!R^d$ be any set of Lagrange data locations having separation distance (8.5). With the function*

$$\phi_0(r) := \inf_{\|\omega\|_\infty \leq 2r} \widehat{\Phi}(\omega), \tag{8.12}$$

the smallest eigenvalue σ of the quadratic form associated to the matrix

$$A_{X,\Phi} = \left(\Phi(x_j - x_k)\right)_{1 \leq j,k \leq M},$$

restricted as usual to the subspace of $I\!R^M$ that contains the coefficient vectors α of functionals $\lambda_{X,M,\alpha} \in \mathcal{P}_\Omega^\perp$ has the lower bound

$$\sigma \geq \frac{1}{2} \frac{\phi_0(K)}{\Gamma\left(d/2+1\right)} \left(\frac{K}{\sqrt{2}}\right)^d \tag{8.13}$$

for any $K > 0$ satisfying

$$K \geq \frac{4}{q} \left(2\pi\Gamma^2\left(d/2+1\right)\right)^{\dfrac{1}{d+1}} \tag{8.14}$$

or, a fortiori,

$$K \geq \frac{9.005\ d}{q}. \tag{8.15}$$

Proof: We start with any $K > 0$ and the characteristic function

$$\chi_K(x) = \left\{ \begin{array}{ll} 1 & \|x\|_2 \leq K \\ 0 & \text{else} \end{array} \right\}$$

of the L_2 ball $B_K(0)$ in $I\!R^d$ with radius K. Then we define

$$\widehat{\Psi}(\omega) := \widehat{\Psi}_K(\omega) := \frac{\phi_0(K)\Gamma\left(d/2+1\right)}{K^d \ \ \pi^{d/2}}(\chi_K * \chi_K)(\omega)$$

and immediately see that the support is

$$\text{supp}\left(\widehat{\Psi}_K\right) = \left\{x \in I\!R^d : \|x\|_2 \leq 2K\right\} =: B_{2K}(0).$$

We now use the formula (10.26) for the volume of the unit ball to get the L_∞ bound

$$\|\chi_K * \chi_K\|_\infty \leq vol(B_K(0)) = K^d \frac{\pi^{d/2}}{\Gamma\left(d/2+1\right)}$$

via the usual convolution integral. We adjusted the factors in the definition of $\widehat{\Psi}$ to guarantee (8.9) on all of $I\!R^d$.

This is part of what we wanted, but we still have to evaluate Ψ itself or at least to show diagonal dominance of $A_{X,\Psi}$. The radial basis function Ψ_K corresponding to $\widehat{\Psi}_K$ is obtained via the inverse Fourier transform as

$$\begin{aligned} \check{\chi}_K(x) &= \check{\chi}_1(\cdot/K)(x) \\ &= K^d \check{\chi}_1(Kx) \\ &= K^d (K\|x\|)^{-d/2} J_{d/2}(K \cdot \|x\|_2) \\ &= \left(\frac{K}{\|x\|}\right)^{d/2} J_{d/2}(K \cdot \|x\|_2) \end{aligned}$$

using scaling of Fourier transforms and (7.16). Then we apply the Fourier transform to the convolution to get

$$\begin{aligned} \Psi_K(x) &= \phi_0(K)\Gamma\left(d/2+1\right) K^{-d}\pi^{-d/2}(\chi_K * \chi_K)^\vee(x) \\ &= \phi_0(K)\Gamma\left(\frac{d}{2}+1\right) 2^{d/2}\|x\|^{-d} J_{d/2}^2(K \cdot \|x\|). \end{aligned}$$

Equation (10.39) yields

$$\Psi_K(0) = \frac{\phi_0(K)}{\Gamma\left(d/2+1\right)}\left(\frac{K}{\sqrt{2}}\right)^d$$

and we assert diagonal dominance of the quadratic form in (8.10) by a suitable choice of K. We have

$$\alpha^T A_{X,\Psi}\alpha \geq \|\alpha\|_2^2 \left(\Psi_K(0) - \max_{1\leq j\leq M} \sum_{\substack{k=1 \\ k\neq j}}^{M} \Psi_K(x_j - x_k) \right)$$

by Gerschgorin's theorem, and the final bound will be of the form

$$\sigma \geq \frac{1}{2}\Psi_K(0) = \frac{\phi_0(K)}{2\Gamma\left(d/2+1\right)}\left(\frac{K}{\sqrt{2}}\right)^d,$$

because we shall choose K such that

$$\max_{1\leq j\leq M} \sum_{\substack{k=1\\k\neq j}} \Psi_K(x_j - x_k) \leq \frac{1}{2}\Psi_K(0). \tag{8.16}$$

This is done by a tricky summation argument of Narcowich and Ward [47] using (10.38) which proves (8.16) for K satisfying (8.14). Since the technique is nice and instructive, we repeat it here in full detail.

To proceed towards diagonal dominance of the matrix, we should fix a point $x_j \in X = \{x_1, \ldots, x_M\}$ and exploit the observation that many of the distances $x_j - x_k$ to the remaining points should be large, if the separation distance $q > 0$ does not let two points to be too near to each other. But the number of far-away points will strongly depend on the space dimension d, and we need a precise argument to put the above reasoning on a solid basis.

To this end, define the sets

$$E_n := \{\ x_k \in X\ :\ nq \leq \|x_j - x_k\|_2 < (n+1)q\ \}$$

for all $n \in \mathbb{N}$ and observe that E_1 is empty due to the definition of the separation distance q, which implies

$$\|x_j - x_k\|_2 \geq 2q \text{ for all } 1 \leq j \neq k \leq M.$$

Now we can put a little ball $B_q(x_k)$ of radius q around each of the $x_k \in E_n$. Any two of these balls cannot overlap due to the definition of q. Since none of the x_k is farther away from x_j than $(n+1)q$, the balls are all contained in the ball $B_{(n+2)q}(x_j)$ of radius $(n+2)q$ around x_j. But all of the x_k are at least nq away from x_j, such that their surrounding balls cannot intersect the smaller ball $B_{(n-1)q}(x_j)$ around x_j of radius $(n-1)q$. Adding their volumes using (10.26) we get the bound

$$\begin{aligned} |E_n|\frac{q^d\pi^{d/2}}{\Gamma(1+d/2)} &\leq \frac{(q(n+2))^d\pi^{d/2}}{\Gamma(1+d/2)} - \frac{(q(n-1))^d\pi^{d/2}}{\Gamma(1+d/2)} \\ |E_n| &\leq (n+2)^d - (n-1)^d. \end{aligned}$$

for the number $|E_n|$ of elements of E_n. If both terms on the right-hand side are expanded with the binomial formula, the leading positive term is $3n^{d-1}$, and all the terms must combine into powers of n with nonnegative factors. Thus we arrive at

$$|E_n| \leq 3n^{d-1}.$$

For points $x_k \in E_n$ we can bound the values of Ψ via (10.38) as follows:

$$\begin{aligned}
\Psi_K(x_j - x_k) &= \phi_0(K)\Gamma\left(\frac{d}{2}+1\right)2^{d/2}\|x_j-x_k\|^{-d}J^2_{d/2}(K\cdot\|x_j-x_k\|)\\
&= \phi_0(K)\Gamma\left(\frac{d}{2}+1\right)2^{d/2}K^{-1}\|x_j-x_k\|^{-d-1}\\
&\quad\cdot(K\cdot\|x_j-x_k\|_2)J^2_{d/2}(K\cdot\|x_j-x_k\|)\\
&\le \phi_0(K)\Gamma\left(\frac{d}{2}+1\right)2^{d/2}K^{-1}((n-1)q)^{-d-1}\frac{2^{d+2}}{\pi}\\
&= \Psi_K(0)\left(\frac{4}{K(n-1)q}\right)^{d+1}\pi^{-1}\Gamma^2\left(\frac{d}{2}+1\right).
\end{aligned}$$

Now it is time to do the summation over all $k \neq j$, and this summation can be done by summing the points in the sets E_n. This yields

$$\begin{aligned}
\sum_{k\neq j}\Psi_K(x_j-x_k) &= \sum_{n=2}^{\infty}\sum_{x_k\in E_n}\Psi(x_j-x_k)\\
&\le \Psi_K(0)\left(\frac{4}{Kq}\right)^{d+1}\pi^{-1}\Gamma^2\left(\frac{d}{2}+1\right)\sum_{n=2}^{\infty}3n^{d-1}(n-1)^{-d-1}\\
&\le \Psi_K(0)\left(\frac{4}{Kq}\right)^{d+1}\pi^{-1}\Gamma^2\left(\frac{d}{2}+1\right)6\sum_{n=2}^{\infty}(n-1)^{-2}\\
&\le \Psi_K(0)\left(\frac{4}{Kq}\right)^{d+1}\pi^{-1}\Gamma^2\left(\frac{d}{2}+1\right)\pi^2\\
&= \Psi_K(0)\left(\frac{4}{Kq}\right)^{d+1}\pi\Gamma^2\left(\frac{d}{2}+1\right)\\
&\le \tfrac{1}{2}\Psi_K(0)
\end{aligned}$$

if we choose K according to (8.16).

It remains to show that (8.15) implies (8.14). We use a variation of Stirling's formula in the form

$$\Gamma(1+x) \le \sqrt{2\pi x}x^x e^{-x}e^{1/12x}, \qquad x>0$$

to get

$$\begin{aligned}
2\pi\Gamma^2\left(d/2+1\right) &\le 2\pi^2 d^{d+1}(2e)^{-d}e^{1/3d},\\
\left(2\pi\Gamma^2\left(d/2+1\right)\right)^{\frac{1}{d+1}} &\le \frac{d}{2e}\left(4e\pi^2\right)^{\frac{1}{d+1}}e^{\frac{1}{3d(d+1)}}\\
&\le d\frac{\pi}{\sqrt{e}}\cdot e^{1/6} \le d\cdot 2.2511
\end{aligned}$$

such that

$$K \ge \frac{9.005}{qd}$$

is satisfactory for all cases. □

We now want to look at the specific cases for applications. From (8.13) and (8.14) we see that

$$\sigma = \sigma(q) =\geq \mathcal{O}\left(q^{-d}\phi_0(cd/q)\right)$$

with some positive constant c. Thus we only need to look at the decay of the Fourier transforms to get the asymptotics of σ with respect to $q \to 0$, keeping the space dimension d fixed. Our known Fourier transforms then yield the results of Table 2.

$\phi(r)$	Lower Bound in $\mathcal{O}$ form for $q \to 0$
r^β	q^β
$r^\beta \log r$	q^β
$(r^2+\gamma^2)^{\beta/2}$	$q^{-d}\exp(-c/q),\ c>0$
$e^{-\beta r^2}$	$q^{-d}\exp(-c/q^2),\ c>0$
$r^\nu K_\nu(r)$	$q^{2\nu}$
$(1-r)_+^2(2+r)$	q
$(1-r)_+^4(1+4r)$	q^3

Table 2: Lower Bounds of Smallest Eigenvalue Based on Lagrange Data with Separation Distance q

8.3 Stability in Function Space

This text is from a recent preprint with Stefano deMarchi, and needs some brushing–up.

8.3.1 Lebesgue Constants

Given a positive definite kernel $\Phi : \Omega \times \Omega \to I\!R$, the recovery of functions from function values $f(x_j)$ on the set $X = \{x_1, ..., x_N\} \subset \Omega \subseteq I\!R^d$ of N different data sites can be done via interpolants of the form

$$s_{f,X} := \sum_{j=1}^{N} \alpha_j \Phi(\cdot, x_j)\ . \tag{8.16}$$

This interpolant, as in classical polynomial interpolation, can also be written in terms of *cardinal functions* $u_j \in V_X := \text{span}\{\Phi(\cdot, x) : x \in X\}$ such that $u_j(x_k) = \delta_{j,k}$. Then, the interpolant (8.16) takes the usual Lagrangian form

$$s_{f,X} = \sum_{j=1}^{N} f(x_j)u_j. \tag{8.16}$$

As in the (univariate) polynomial case, based on the representation (8.16) we consider the *Lebesgue function*

$$\lambda_N(x) := \sum_{j=1}^{N} |u_j(x)|\ .$$

Its maximum value, $\Lambda_N := \max_{x\in\Omega} \lambda_N(x)$ is referred to as the associated *Lebesgue constant* and gives the norm of the interpolating projector $\mathcal{P}_X : \mathcal{C}(\Omega) \to V_X \subseteq V_\Omega$, with $V_\Omega = \text{span}\{\Phi(\cdot, x) :$

$x \in \Omega\}$, both spaces equipped with the sup-norm. As well-known in the polynomial case, either in the univariate and in the bivariate case, there exist upper bounds for the Lebesgue function. Moreover, many authors faced the problem of finding near-optimal points for polynomial interpolation. All these near-optimal sets of N points have a Lebesgue function that behaves in 1D like $\log(N)$ while as $\log^2(N)$ in 2D (cf. [] and references therein).

We want to bound the Lebesgue constant and the Lebesgue function for interpolation projectors using (8.16). For a rather large class of kernel-based multivariate interpolants, we can prove that the Lagrange basis functions for N well-distributed data locations in a bounded Lipschitz domain with an interior cone condition are uniformly bounded, and thus the Lebesgue constant grows only linearly with N, irrespective of the space dimension and the kernel used.

For conditionally positive definite kernels with finite smoothness, sharper results are possible. The classical Lebesgue constants grow only like $\sqrt{N}$, and the generalized L_2 Lebesgue constants, defined as the norms of the interpolation projectors between $\mathbb{R}^N$ under a scaled ℓ_2 norm and $L_2(\Omega)$ are uniformly bounded, provided that the data locations are well-distributed. Specific estimates for general scattered data locations are also available, and some numerical examples in the next section show that the results are realistic.

We shall consider interpolation of d-variate functions on a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$ with an outer cone condition [**?**]. Interpolation is done on a set $X = \{x_1, \ldots, x_N\}$ of N scattered *data locations* or *centers*. Their geometric relation to the domain Ω is described by the *fill distance* or *mesh norm*

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

and the *separation distance*

$$q_X = \frac{1}{2} \min_{\substack{x_i, x_j \in X \\ x_i \neq x_j}} \|x_i - x_j\| .$$

These parameters are used for standard error and stability estimates for multivariate interpolants, and they will be also of inportance here. The inequality $q_X \leq h_{X,\Omega}$ will hold in most cases, but if points of X nearly coalesce, q_X can be much smaller than $h_{X,\Omega}$, causing ionstability of the standard solution process. Point sets X are called *quasi–uniform* with *uniformity constant* $\gamma > 1$, if the inequality

$$\frac{1}{\gamma} q_X \leq h_{X,\Omega} \leq \gamma q_X$$

holds. Later, we shall consider arbitrary sets of arbitrary cardinality, but with uniformity constants bounded above by a fixed number. Note that $h_{X,\Omega}$ and q_X play an important role in finding good points for radial basis function interpolation, as recently studied in [**?**, **?**, **?**].

To generate interpolants, we allow conditionally positve definite translation-invariant kernels Φ of the form

$$(x, y) \mapsto \Phi(x - y), \; x, y \in \mathbb{R}^d$$

which have generalized Fourier transforms on $\mathbb{R}^d$ [**?**].

For reasons to become apparent later, we consider two different classes of kernels. First, there are kernels of *limited smoothness* measured by a parameter τ with

$$0 < c(1 + \|\omega\|_2^2)^{-\tau} \leq \hat{\Phi}(\omega) \leq C(1 + \|\omega\|_2^2)^{-\tau} \tag{8.16}$$

at infinity. This includes polyharmonic splines, thin-plate splines, the Sobolev/Matern kernel, and Wendland's compactly supported kernels. Second, there are kernels with *unlimited smoothness* where the Fourier transform decays exponentially at infinity, e.g. the *Gaussian* and various *multiquadrics*.

8.3.2 Results for Limited Smoothness

Under the assumption (8.16) the space V_X will be a subspace of Sobolev space $W_2^\tau(\Omega)$. Our central result then is

Theorem 8.17 *The classical Lebesgue constant for interpolation with Φ on N data locations $X = \{x_1, \ldots, x_n\}$ in a bounded domain $\Omega \subseteq \mathbb{R}^d$ satisfying an outer cone condition has a bound of the form*

$$\lambda_N \leq C\sqrt{N}\left(\frac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2}.$$

For quasi-uniform sets with bounded uniformity γ, this simplifies to

$$\lambda_N \leq C\sqrt{N}.$$

Each single cardinal function is bounded by

$$\|u_j\|_{L_\infty(\Omega)} \leq C\left(\frac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2},$$

which in the quasi-uniform case simplifies to

$$\|u_j\|_{L_\infty(\Omega)} \leq C.$$

There also is an L_2 analog of this. We compare the $L_2(\Omega)$ norm of f with its discrete counterpart $h_{X,\Omega}^{d/2}\|f_{|X}\|_2$ and note that the latter converges to a multiple of the former, if f is smooth and if the discrete set X is quasi-uniform and asymptotically dense. The *generalized L_2 Lebesgue constant* can then be defined as the norm of the map

$$f_{|X} \mapsto s_{f,X}, \quad \mathbb{R}^N \to L_2(\Omega)$$

if the above norms are chosen.

Theorem 8.18 *Under the above assumptions,*

$$\|s_{f,X}\|_{\tilde{L}_2(\Omega)} \leq C\left(\frac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2} h_{X,\Omega}^{d/2}\|f\|_{2,X},$$

and for quasi-uniform data locations with bounded uniformity γ the generalized L_2 Lebesgue constant is uniformly bounded. The cardinal functions have a bound

$$\|u_j\|_{L_2(\Omega)} \leq C\left(\frac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2} h_{X,\Omega}^{d/2}$$

and for quasi-uniform data locations they behave like

$$\|u_j\|_{L_2(\Omega)} \leq Ch_{X,\Omega}^{d/2}.$$

8.3.3 L_∞ Bounds

Our most important tool for the proof of Theorem 8.17 is the *sampling inequality* (cf. [**?**, Th. 2.6])

$$\|u\|_{L_\infty(\Omega)} \le C\left(h_{X,\Omega}^{\tau-d/2}\|u\|_{W_2^\tau(\Omega)} + \|u\|_{\infty,X}\right), \quad \forall u \in W_2^\tau(\Omega), \tag{8.18}$$

where $X \subset \Omega$ is a discrete set of points in Ω with fill distance $h_{X,\Omega}$. This is independent of kernels.

We can apply the sampling inequality in two ways

$$\begin{aligned}
\|s_{f,X}\|_{L_\infty(\Omega)} &\le C\left(h_{X,\Omega}^{\tau-d/2}\|s_{f,X}\|_{W_2^\tau(\Omega)} + \|s_{f,X}\|_{\infty,X}\right)\\
&\le C\left(h_{X,\Omega}^{\tau-d/2}\|s_{f,X}\|_{W_2^\tau(\Omega)} + \|f\|_{\infty,X}\right),\\
\|u_j\|_{L_\infty(\Omega)} &\le C\left(h_{X,\Omega}^{\tau-d/2}\|u_j\|_{W_2^\tau(\Omega)} + \|u_j\|_{\infty,X}\right)\\
&\le C\left(h_{X,\Omega}^{\tau-d/2}\|u_j\|_{W_2^\tau(\Omega)} + 1\right)
\end{aligned}$$

since we know that the space V_X is contained in $W_2^\tau(\Omega)$. To get a bound on either the Lebesgue constant or the norm of a cardinal function, we have to find bounds of the form

$$\|s\|_{W_2^\tau(\Omega)} \le C(X,\Omega,\Phi)\|s\|_{\infty,X}$$

for arbitrary elements $s \in V_X$. Such bounds are available from [**?**], but we repeat the basic notation here. Let Φ satisfy (8.16). Then [**?**] has

$$\|s\|_{W_2^\tau(\Omega)}^2 \le Cq_X^{-2\tau+d}\|s\|_{2,X}^2 \le CNq_X^{-2\tau+d}\|s\|_{\infty,X}^2 \text{ for all } s \in V_X$$

with a different generic constant. If we apply this to u_j, we get

$$\|u_j\|_{L_\infty(\Omega)} \le C\left(\left(\frac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2} + 1\right),$$

while application to $s_{f,X}$ yields

$$\|s_{f,X}\|_{L_\infty(\Omega)} \le C\left(\left(\frac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2}\|f\|_{2,X} + \|f\|_{\infty,X}\right) \le C\left(\sqrt{N}\left(\frac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2} + 1\right)\|f\|_{\infty,X}.$$

Then the assertions of Theorem 8.17 follow. □.

8.3.4 L_2 Bounds

For the L_2 case covered by Theorem 8.18, we take the sampling inequality

$$\|f\|_{L_2(\Omega)} \le C\left(h_{X,\Omega}^{\tau}\|f\|_{W_2^\tau(\Omega)} + \|f\|_{\ell_2(X)}h_{X,\Omega}^{d/2}\right), \quad \forall f \in W_2^\tau(\Omega) \tag{8.18}$$

of [**?**, Thm. 3.5]. We can apply the sampling inequality as

$$\begin{aligned}
\|s_{f,X}\|_{L_2(\Omega)} &\le C\left(h_{X,\Omega}^{\tau}\|s_{f,X}\|_{W_2^\tau(\Omega)} + \|s_{f,X}\|_{\ell_2(X)}h_{X,\Omega}^{d/2}\right)\\
&\le C\left(h_{X,\Omega}^{\tau}\|s_{f,X}\|_{W_2^\tau(\Omega)} + \|f\|_{\ell_2(X)}h_{X,\Omega}^{d/2}\right),\\
&\le C\left(\tfrac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2}\|f\|_{\ell_2(X)}h_{X,\Omega}^{d/2},\\
\|u_j\|_{L_2(\Omega)} &\le C\left(h_{X,\Omega}^{\tau}\|u_j\|_{W_2^\tau(\Omega)} + \|u_j\|_{\ell_2(X)}h_{X,\Omega}^{d/2}\right)\\
&\le C\left(h_{X,\Omega}^{\tau-d/2}\|u_j\|_{W_2^\tau(\Omega)} + 1\right)h_{X,\Omega}^{d/2}\\
&\le C\left(\left(\tfrac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2} + 1\right)h_{X,\Omega}^{d/2}
\end{aligned}$$

This proves Theorem 8.18. □

9 Hilbert Space Theory

This is intended as a short tutorial on Hilbert spaces as required in this text. We only require fundamentals on linear spaces, bilinear forms, and norms. If a reader has problems with any of the stated facts below, it is time to go back to an introductory course on Calculus and Numerical Analysis.

9.1 Normed Linear Spaces

For completeness, we recall some basics from normed linear spaces:

1. A sequence $\{u_n\}_{n\in\mathbb{N}} \subset \mathcal{N}$ of a normed linear space $\mathcal{N}$ with norm $\|\cdot\|_{\mathcal{N}}$ is a **zero sequence** in $\mathcal{N}$, if the sequence $\{\|u_n\|_{\mathcal{N}}\}_{n\in\mathbb{N}}$ converges to zero in $\mathbb{R}$.

2. A sequence $\{u_n\}_{n\in\mathbb{N}} \subset \mathcal{N}$ is a **convergent sequence** in $\mathcal{N}$ with limit u, if the sequence $\{u_n - u\}_n$ is a zero sequence.

3. A subspace $\mathcal{M}$ of $\mathcal{N}$ is a **closed subspace**, if for every convergent sequence $\{u_n\}_{n\in\mathbb{N}} \subset \mathcal{M} \subset \mathcal{N}$ with limit u one can conclude that the limit u also belongs to $\mathcal{M}$.

4. The normed linear space $\mathcal{N}$ is **complete** or a **Banach space**, if every sequence which is a Cauchy sequence in the norm $\|\cdot\|_{\mathcal{V}}$ is necessarily convergent in $\mathcal{V}$.

5. A complete normed linear space is closed, since each convergent sequence is a Cauchy sequence.

6. A subset $\mathcal{M}$ of a normed linear space $\mathcal{N}$ is **dense**, if each element of $\mathcal{N}$ can be written as a limit of a convergent sequence from $\mathcal{M}$.

7. A linear mapping (or operator) $A : \mathcal{N} \to \mathcal{M}$ with values in a normed linear space $\mathcal{M}$ with norm $\|\cdot\|_{\mathcal{M}}$ is a **continuous mapping** or a **bounded mapping**, if there is a constant C such that
$$\|Ax\|_{\mathcal{M}} \leq C\|x\|_{\mathcal{N}}$$
holds for all $x \in \mathcal{N}$.

8. The mapping A then has an **operator norm**
$$\|A\|_{\mathcal{N},\mathcal{M}} := \sup_{x\in\mathcal{N}\setminus\{0\}} \frac{\|Ax\|_{\mathcal{M}}}{\|x\|_{\mathcal{N}}} \leq C$$
and the bound
$$\|Ax\|_{\mathcal{M}} \leq \|A\|_{\mathcal{N},\mathcal{M}}\|x\|_{\mathcal{N}}$$
is best possible.

9. The most important special case arises for $\mathcal{M} = \mathbb{R}$, i.e. for linear **functionals** $\lambda : \mathcal{N} \to \mathbb{R}$. If they are *continuous*, they have an operator norm
$$\|\lambda\|_{\mathcal{N}^*} := \|\lambda\|_{\mathcal{N},\mathbb{R}} := \sup_{x\in\mathcal{N}\setminus\{0\}} \frac{|\lambda(x)|}{\|x\|_{\mathcal{N}}} \leq C.$$

10. The space of continuous linear functionals on a normed linear space $\mathcal{N}$ is a normed linear space under the above **dual** norm, and it is called the **dual** space $\mathcal{N}^*$ to $\mathcal{N}$.

11. The kernel of a continuous linear map on a normed linear space is always a **closed** subspace.

9.2 Pre–Hilbert Spaces

Definition 9.1 *A set $\mathcal{H}$ and a mapping $(\cdot,\cdot)_{\mathcal{H}} : \mathcal{H}\times\mathcal{H}\to I\!R$ form a* **pre-Hilbert space** *or a* **Euclidean** *space over $I\!R$, if the following holds:*

1. *$\mathcal{H}$ is a vector space over $I\!R$.*

2. *$(\cdot,\cdot)_{\mathcal{H}}$ is a symmetric positive definite bilinear form.*

A symmetric positive bilinear form as $(\cdot,\cdot)_{\mathcal{H}} : \mathcal{H}\times\mathcal{H}\to I\!R$ is often called an **inner product** on $\mathcal{H}$. Then

$$\|x\|_{\mathcal{H}}^2 := (x,x)_{\mathcal{H}},\ x\in\mathcal{H} \tag{9.2}$$

defines a norm on $\mathcal{H}$, and we assume all readers to be familiar with this notion. Sometimes, we shall weaken the assumptions on $(\cdot,\cdot)_{\mathcal{H}}$ and only ask for symmetry and positive semidefiniteness. Even in this more general situation, we have the **Cauchy-Schwarz inequality**

$$|(u,v)_{\mathcal{H}}| \le |u|_{\mathcal{H}}|v|_{\mathcal{H}}$$

for all $u,v\in\mathcal{H}$, where we use the notation $|x|_{\mathcal{H}}^2 := (x,x)_{\mathcal{H}}$ to denote a **seminorm** instead of a norm as in (9.2). To prove the Cauchy-Schwarz inequality as a warm-up, just consider the quadratic function

$$\varphi(t) := |u+tv|_{\mathcal{H}}^2 = |u|_{\mathcal{H}}^2 + 2t(u,v)_{\mathcal{H}} + t^2|v|_{\mathcal{H}}^2.$$

It must be nonnegative, and thus it has none or a double real zero. This property is satisfied for a general function $\varphi(t) = at^2+2bt+c$, iff $b^2\le ac$ holds. But this is the square of the Cauchy-Schwarz inequality. An argument like the one above is very frequent, and we call it the "parabola argument".

Now we add some simple facts about pre-Hilbert spaces:

1. For two nonzero elements x,y of $\mathcal{H}$ one can define the cosine of the **angle** $\angle(x,y)$ as

$$\cos(\angle(x,y)) = \frac{(x,y)_{\mathcal{H}}}{\|x\|_{\mathcal{H}}\|y\|_{\mathcal{H}}}.$$

2. Two elements x,y of $\mathcal{H}$ are **orthogonal**, if $(x,y)_{\mathcal{H}}=0$. In that case, the theorem of Pythagoras is

$$\|x+y\|_{\mathcal{H}}^2 = \|x\|_{\mathcal{H}}^2 + \|y\|_{\mathcal{H}}^2$$

and trivially proven by evaluating the left–hand side as

$$\|x+y\|_{\mathcal{H}}^2 = (x+y,x+y)_{\mathcal{H}} = \|x\|_{\mathcal{H}}^2 + 2(x,y)_{\mathcal{H}} + \|y\|_{\mathcal{H}}^2.$$

3. Two subspaces $\mathcal{U},\ \mathcal{V}$ of a pre-Hilbert space are **orthogonal**, if all vectors $u\in\mathcal{U},\ v\in\mathcal{V}$ are orthogonal, i.e.: $(u,v)_{\mathcal{H}}=0$.

Roughly speaking, **Euclidean geometry** needs the definition of angles and orthogonality. This is why one can also use the notion of an **Euclidean** space here.

9.3 Sequence Spaces

For illustration, we can look at **sequence spaces**. Let I be a finite or countably infinite set, and take the space

$$S_0 := \text{ span } \{\{\xi_i\}_{i\in I} \ : \ \xi_i \neq 0 \text{ for only finitely many } i \in I\}. \tag{9.3}$$

Then take a sequence $\{\lambda_i\}_{i\in I}$ of positive numbers, and define the inner product

$$(\{\xi_i\}_{i\in I}, \{\eta_i\}_{i\in I})_{\lambda,I} := \sum_{i\in I} \lambda_i \xi_i \eta_i$$

on S_0. Then S_0 is a pre–Hilbert space with the above inner product, and we should call it $S_{0,\lambda,I}$ now to make the dependence on the topology on λ_i apparent. The dual of $S_{0,\lambda,I}$ is at least as large as the full sequence space

$$S_\infty := \text{ span } \{\{\mu_i\}_{i\in I}\}$$

because we can let each $\mu := \{\mu_i\}_{i\in I} \in S_\infty$ act on each $\xi := \{\xi_i\}_{i\in I} \in S_0$ via

$$\mu(\xi) := \sum_{i\in I} \xi_i \mu_i$$

because we only have finitely many nonzero ξ_i.

If we allow infinite sequences, we have to be careful with convergence and duality. But we can define the space

$$S_{\lambda,I} := \text{ span } \{\{\xi_i\}_{i\in I} \ : \ \sum_{i\in I} \lambda_i \xi_i^2 < \infty\} \tag{9.4}$$

which clearly also has the above inner product, and it contains $S_{0,\lambda,I}$. We assert that its dual contains $S_{1/\lambda,I}$, and it can surely not be as large as S_∞. We can let each $\mu := \{\mu_i\}_{i\in I} \in S_{1/\lambda,I}$ act on each $\xi := \{\xi_i\}_{i\in I} \in S_{\lambda,I}$ via

$$\mu(\xi) := \sum_{i\in I} \xi_i \mu_i$$

because we can use the Cauchy–Schwarz inequality

$$\begin{aligned}
\mu(\xi)^2 &= \left(\sum_{i\in I} \xi_i \mu_i \right)^2 \\
&= \left(\sum_{i\in I} \sqrt{\lambda_i} \xi_i \frac{1}{\sqrt{\lambda_i}} \mu_i \right)^2 \\
&\leq \left(\sum_{i\in I} \lambda_i \xi_i^2 \right) \left(\sum_{i\in I} \frac{1}{\lambda_i} \mu_i^2 \right) \\
&= \|\xi\|_{\lambda,I}^2 \|\mu\|_{1/\lambda,I}^2.
\end{aligned}$$

9.4 Best Approximations

Definition 9.5 *An element u^* of a subspace $\mathcal{M}$ of a normed linear space $\mathcal{N}$ is a* **best approximation** *to a given element $u \in \mathcal{N}$, if*

$$\|u - u^*\|_{\mathcal{N}} = \sup_{v\in\mathcal{M}} \|u - v\|_{\mathcal{N}} =: E_{\mathcal{M}}(u).$$

Theorem 9.6 *An element u^* of a subspace $\mathcal{M}$ of a pre-Hilbert space $\mathcal{H}$ is a best approximation to a given element $u \in \mathcal{H}$, iff the variational identity*

$$(u - u^*, v)_{\mathcal{H}} = 0 \text{ for all } v \in \mathcal{M} \tag{9.7}$$

holds. If it exists, the best approximation is unique. If $\mathcal{M}$ is finite-dimensional and spanned by linearly independent elements $u_1 \ldots, u_M$, then the coefficients α^* of the representation

$$u^* = \sum_{j=1}^{M} \alpha_j^* u_j$$

are solutions of the **normal equations**

$$\sum_{j=1}^{M} \alpha_j^* (u_j, u_k)_{\mathcal{H}} = (u, u_k)_{\mathcal{H}},\ 1 \le k \le M.$$

The symmetric and positive definite matrix with entries $(u_j, u_k)_{\mathcal{H}}$ in the above system is called a **Gram matrix**. *In this special case, the best approximation exists uniquely and can theoretically be calculated via the normal equations.*

Proof: Let u^* be a best approximation to u. To have another instance of the parabola argument, consider an arbitrary $v \in \mathcal{M}$ and form the quadratic function

$$u_v(\alpha) := \|u - u^* + \alpha v\|_{\mathcal{H}}^2 = \|u - u^*\|_{\mathcal{H}}^2 + 2\alpha(u - u^*, v)_{\mathcal{H}} + \alpha^2 \|v\|_{\mathcal{H}}^2$$

whose minimum must be attained at $\alpha = 0$. This implies $(u - u^*, v)_{\mathcal{H}} = 0$. Conversely, assume (9.7) and write any other element $v \in \mathcal{M}$ as $v = u^* + 1 \cdot (v - u^*)$. Then (9.7) implies that the quadratic function $u_{u^* - v}$ is minimal at $\alpha = 0$, proving $u_{u^*-v}(1) = \|u - v\|_{\mathcal{H}} \ge u_{u^*-v}(0) = \|u - u^*\|_{\mathcal{H}}$. If u^* and u^{**} are two best approximations from $\mathcal{M}$ to u, then we can subtract the two variational identities $(u - u^*, v)_{\mathcal{H}} - (u - u^{**}, v)_{\mathcal{H}} = (u^{**} - u^*, v)_{\mathcal{H}} = 0$ for all $v \in \mathcal{M}$ and insert $v = u^{**} - u^*$ to get $u^{**} = u^*$. The third assertion is a specialization of (9.7). □

Corollary 9.8 *The first statement of Theorem 9.6 holds also in the case of a positive semidefinite bilinear form. The Gram matrix in the finite-dimensional case now is only positive semidefinite.* □

Corollary 9.9 *Let $\lambda_1, \ldots, \lambda_M$ be linear functionals on a pre-Hilbert space $\mathcal{H}$ and let some $u \in \mathcal{H}$ be given. An element u^* of $\mathcal{H}$ solves the problem*

$$\|u^*\|_{\mathcal{H}} = \inf_{\substack{v \in \mathcal{H} \\ \lambda_j(v) = \lambda_j(u) \\ 1 \le j \le M}} \|v\|_{\mathcal{H}},$$

iff the variational identity

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

holds, or iff there are real numbers $\alpha_1, \ldots, \alpha_M$ such that

$$(u^*, v)_{\mathcal{H}} = \sum_{j=1}^{M} \alpha_j \lambda_j(v) \text{ for all } v \in \mathcal{H}.$$

Proof: Consider the subspace

$$\mathcal{M} = \{\ v \in \mathcal{H}\ :\ \lambda_j(v) = 0,\ 1 \le j \le M\}$$

and reformulate the problem by writing any $v \in \mathcal{H}$ with $\lambda_j(v) = \lambda_j(u),\ 1 \le j \le M$ as $v = u - w$ for $w \in \mathcal{M}$. Then we have a problem of best approximation to u from $\mathcal{M}$ and can simply use Theorem 9.6 to prove the first assertion. We then have to prove that the first variational identity implies the second. But this follows from a standard linear algebra argument that we include for completeness as the next lemma. □

Lemma 9.10 *If $A\ :\ X \to Y$ and $B\ :\ X \to Z$ are linear maps between linear spaces, and if B vanishes on the kernel* ker *A of A, then B factorizes over $A(X)$, i.e.: there is a map $C\ :\ A(X) \to Z$ such that $B = C \circ A$. If Z is normed and if Y is finite-dimensional, then C is continuous.*

Proof: There is an isomorphism $D\ :\ A(X) \to X/\ker A$, and one can define $\widetilde{B}\ :\ A/\ker A \to Z$ by $\widetilde{B}(x + \ker A) := B(x)$ because $B(\ker A) = \{0\}$. Then $C := \widetilde{B} \circ D$ does the job, since

$$C(A(x)) = \widetilde{B}(D(A(x))) = \widetilde{B}(x + \ker A) = B(x)$$

for all $x \in X$. If Y is finite-dimensional, the isomorphic spaces $A(X) \subseteq Y$ and $X/\ker A$ must also be finite-dimensional. Since all linear mappings defined on finite-dimensional linear spaces with values in normed linear spaces are continuous, we are finished. □

9.5 Hilbert Spaces

So far, Theorem 9.6 does not imply existence of best approximations from subspaces of infinite dimension. It just characterizes them. To get existence, we need that certain nice sequences actually have limits:

Definition 9.11 *A pre-Hilbert space $\mathcal{H}$ with inner product $(\cdot,\cdot)_{\mathcal{H}}$ is a* **Hilbert space** *over $\mathbb{R}$, if $\mathcal{H}$ is* **complete** *under the norm $\|\cdot\|_{\mathcal{H}}$, i.e.: as a normed linear space.*

We now prove the crucial **projection theorem** in Hilbert spaces:

Theorem 9.12 *If $\mathcal{H}$ is a Hilbert space with a closed subspace $\mathcal{M}$, then any element $u \in \mathcal{H}$ has a unique best approximation $u^*_{\mathcal{M}}$ from $\mathcal{M}$, and the elements $u^*_{\mathcal{M}}$ and $u - u^*_{\mathcal{M}}$ are orthogonal. The map $\Pi_{\mathcal{M}}\ :\ \mathcal{H} \to \mathcal{M}$ with $\Pi_{\mathcal{M}}(u) := u^*_{\mathcal{M}}$ is linear, has norm one if $\mathcal{M}$ is nonzero, and is a* **projector**, *i.e. it is* **idempotent**, *meaning $\Pi^2_{\mathcal{M}} = \Pi_{\mathcal{M}}$. If Id is the identity mapping, then $Id - \Pi_{\mathcal{M}}$ is another projector, mapping $\mathcal{H}$ onto the* **orthogonal complement**

$$\mathcal{M}^{\perp} := \{\ u \in \mathcal{H}\ :\ (u,v)_{\mathcal{H}} = 0 \text{ for all } v \in \mathcal{M}\ \}.$$

of $\mathcal{M}$. Finally, the decomposition

$$\mathcal{H} = \mathcal{M} + \mathcal{M}^{\perp}$$

is a direct and orthogonal sum of spaces.

Proof: The existence proof for approximations from finite-dimensional subspaces is a consequence of Theorem 9.6, and we postpone the general case for a moment. The orthogonality statement follows in general from Theorem 9.6, and it yields Pythagoras' theorem in the form

$$\|u\|_{\mathcal{H}}^2 = \|u - u_{\mathcal{M}}^*\|_{\mathcal{H}}^2 + \|u^*\|_{\mathcal{H}}^2.$$

This in turn proves that both projectors have a norm not exceeding one. It is easy to prove that $\alpha u_{\mathcal{M}}^* + \beta v_{\mathcal{M}}^*$ is a best approximation to $\alpha u + \beta v$ for all $\alpha, \beta \in I\!R$ and all $u, v \in \mathcal{H}$, using the variational identity in Theorem 9.6. To prove linearity of the projectors, we use uniqueness of the best approximation, as follows from Theorem 9.6. Finally, surjectivity of the projectors is easily proven from the best approximation property of their definition.

Thus we are left with the existence proof for the infinite-dimensional case. The nonnegative real number $E_{\mathcal{M}}(u)$ can be written as the limit of a decreasing sequence $\{\|u - v_n\|_{\mathcal{H}}\}_n$ for certain elements $v_n \in \mathcal{M}$, because it is defined as an infimum. Forming the subspaces

$$\mathcal{M}_n := \text{ span } \{v_1, \ldots, v_n\} \subseteq \mathcal{M}$$

and unique best approximations w_n to u from $\mathcal{M}_n$, we get

$$E_{\mathcal{M}}(u) \leq \|u - w_n\|_{\mathcal{H}} \leq \|u - v_n\|_{\mathcal{H}},$$

such that the sequence $\{\|u - w_n\|_{\mathcal{H}}\}_n$ converges to $E_{\mathcal{M}}(u)$, too. We now fix indices $m \geq n$ and use that $(u - w_m, w_m - w_n)_{\mathcal{H}} = 0$ follows from the best approximation property of w_m. Then we have

$$\begin{aligned} \|u - w_n\|_{\mathcal{H}}^2 - \|u - w_m\|_{\mathcal{H}}^2 &= \|u - w_m + w_m - w_n\|_{\mathcal{H}}^2 - \|u - w_m\|_{\mathcal{H}}^2 \\ &= \|u - w_m\|_{\mathcal{H}}^2 + 2(u - w_m, w_m - w_n)_{\mathcal{H}} \\ &\quad + \|w_m - w_n\|_{\mathcal{H}}^2 - \|u - w_m\|_{\mathcal{H}}^2 \\ &= \|w_m - w_n\|_{\mathcal{H}}^2, \end{aligned}$$

and since the sequence $\{\|u - w_n\|_{\mathcal{H}}^2\}_n$ is convergent and thus a Cauchy sequence, we get that $\{w_n\}_n \subset \mathcal{M}$ is a Cauchy sequence in $\mathcal{M} \subseteq \mathcal{H}$. Now the completeness of $\mathcal{H}$ assures the existence of a limit $w^* \in \mathcal{H}$ of this sequence, and since $\mathcal{M}$ was assumed to be closed, the element w^* must belong to $\mathcal{M}$. The above identity can be used to let m tend to infinity, and then we get

$$\|u - w_n\|_{\mathcal{H}}^2 - \|u - w^*\|_{\mathcal{H}}^2 = \|w^* - w_n\|_{\mathcal{H}}^2.$$

This proves

$$E_{\mathcal{M}}(u) \leq \|u - w^*\|_{\mathcal{H}} \leq \|u - w_n\|_{\mathcal{H}},$$

and since the right-hand side converges to $E_{\mathcal{M}}(u)$, the element w^* must be the best approximation to u. □

We add two little applications:

Lemma 9.13 *If an element f from a Hilbert space $\mathcal{H}$ is orthogonal to $\mathcal{H}$, it is zero.*

Proof: Take $\mathcal{M} = \mathcal{H}$ in Theorem 9.12. The space $\mathcal{M}^\perp$ contains f, but it is necessarily zero, so that f is zero. But a more simple and direct proof just uses that f is orthogonal to itself:

$$\|f\|_{\mathcal{H}}^2 = (f, f)_{\mathcal{H}} = 0.$$

□

Lemma 9.14 *If $\mathcal{M}$ is a dense subspace of a Hilbert space $\mathcal{H}$, then the closure of $\mathcal{M}$ is isometrically isomorphic to $\mathcal{H}$.*

Proof: The closure of $\mathcal{M}$ can be identified with a closed subspace $\mathcal{N}$ of $\mathcal{H}$, and we assert that $\mathcal{N} = \mathcal{H}$. To this end, decompose $\mathcal{H}$ into $\mathcal{H} = \mathcal{N} + \mathcal{N}^\perp$ and take an element u from $\mathcal{N}^\perp$. It must be orthogonal to all elements from $\mathcal{M}$, and by continuity of the functional $v \mapsto (u, v)_\mathcal{H}$ it must be orthogonal to all of $\mathcal{H}$. Thus it must be zero. □

9.6 Riesz Representation Theorem

We further need the **Riesz representation theorem** for continuous linear functionals. To this end, we recall that the **dual** $\mathcal{N}^*$ of a normed linear space $\mathcal{N}$ consists of all continuous linear functionals $\lambda : \mathcal{N} \to I\!R$ with **dual norm**

$$\|\lambda\|_{\mathcal{N}^*} := \sup_{f \in \mathcal{N}, f \neq 0} \frac{\lambda(f)}{\|f\|_\mathcal{N}}.$$

It is a normed linear space under this norm.

Theorem 9.15 *(**Riesz representation theorem**)*
Any continuous linear real-valued functional λ on a Hilbert space $\mathcal{H}$ can be written as

$$\lambda = (\cdot, g_\lambda)_\mathcal{H} \tag{9.16}$$

with a unique element $g_\lambda \in \mathcal{H}$ satisfying $\|\lambda\|_{\mathcal{H}^*} = \|g_\lambda\|_\mathcal{H}$.

Proof: If $\lambda = 0$, then $g_\lambda = 0$ does the job and is unique. If $\lambda \neq 0$, the kernel $\mathcal{L}$ of λ is not the full space $\mathcal{H}$. It is, however, a closed linear subspace, and thus there is some element $h_\lambda \in \mathcal{L}^\perp$ with $\|h_\lambda\|_\mathcal{H} = 1$. Now for each $u \in \mathcal{H}$ the element $\lambda(u)h_\lambda - \lambda(h_\lambda)u$ must necessarily be in $\mathcal{L}$ and thus orthogonal to h_λ. This means

$$\begin{aligned} 0 &= (h_\lambda, \lambda(u)h_\lambda - \lambda(h_\lambda)u)_\mathcal{H}, \\ \lambda(u)(h_\lambda, h_\lambda)_\mathcal{H} &= \lambda(h_\lambda)(u, h_\lambda)_\mathcal{H}, \\ \lambda(u) &= (u, \lambda(h_\lambda)h_\lambda)_\mathcal{H} \end{aligned}$$

and we can define $g_\lambda := \lambda(h_\lambda)h_\lambda$ to get (9.16).

The norm of λ is bounded by

$$\begin{aligned} \|\lambda\|_{\mathcal{H}^*} &:= \sup_{u \in \mathcal{H} \setminus \{0\}} \frac{|\lambda(u)|}{\|u\|_\mathcal{H}} \\ &\leq |\lambda(h_\lambda)| \end{aligned}$$

due to Cauchy-Schwarz, but using $u = h_\lambda$ in the definition of the norm yields equality. Since we set $g_\lambda := \lambda(h_\lambda)h_\lambda$, we get $\|\lambda\|_{\mathcal{H}^*} = \|g_\lambda\|_\mathcal{H}$. Uniqueness of g_λ satisfying (9.16) is easy to prove, because for any other $\tilde{g}_\lambda$ with (9.16) we have

$$(g_\lambda - \tilde{g}_\lambda, f)_\mathcal{H} = \lambda(f) - \lambda(f) = 0 \text{ for all } f \in \mathcal{H},$$

and thus $g_\lambda - \tilde{g}_\lambda = 0$ because it is orthogonal to the full space. □

Definition 9.17 *The map*

$$R \,:\, \mathcal{H}^* \to \mathcal{H} \text{ with } \lambda \mapsto g_\lambda \text{ for all } \lambda \in \mathcal{H}^*$$

on the dual $\mathcal{H}^*$ of a Hilbert space $\mathcal{H}$ is called the **Riesz** *map.*

Theorem 9.18 *The Riesz map is a linear isometric bijection between a Hilbert space $\mathcal{H}$ and its dual $\mathcal{H}^*$. In particular, the dual norm can be written as a Hilbert space norm based on an inner product $(.,.)_{\mathcal{H}^*}$ satisfying*

$$(R(\lambda), R(\mu))_{\mathcal{H}} = (\lambda, \mu)_{\mathcal{H}^*} \text{ for all } \lambda, \mu \in \mathcal{H}^*.$$

Thus any Hilbert space is isometrically isomorphic to its dual via the Riesz map.

Proof: We already know that the Riesz map is well–defined and satisfies $\|R(\lambda)\|_{\mathcal{H}} = \|\lambda\|_{\mathcal{H}^*}$. It also is linear due to

$$\begin{aligned}(R(a\lambda + b\mu), f)_{\mathcal{H}} &= (a\lambda + b\mu)(f) \\ &= a\lambda(f) + b\mu(f) \\ &= a(f, R(\lambda))_{\mathcal{H}} + b(f, R(\mu))_{\mathcal{H}} \\ &= (f, a \cdot R(\lambda) + b \cdot R(\mu))_{\mathcal{H}} \text{ for all } f \in \mathcal{H},\ a, b \in I\!R\end{aligned}$$

because this implies that $R(a\lambda + b\mu) - (a \cdot R(\lambda) + b \cdot R(\mu)$ is orthogonal to all of $\mathcal{H}$, thus zero.

We already have

$$\lambda(f) = (f, R(\lambda))_{\mathcal{H}} \text{ for all } f \in \mathcal{H},\ \lambda \in \mathcal{H}^*.$$

But now we use $f = R(\mu)$ and get

$$\lambda(R(\mu)) = (R(\mu), R(\lambda))_{\mathcal{H}} = \mu(R(\lambda)) \text{ for all } \lambda, \mu \in \mathcal{H}^*.$$

We define a "new" bilinear form

$$(\lambda, \mu)_* := \lambda(R(\mu)) = (R(\mu), R(\lambda))_{\mathcal{H}} = \mu(R(\lambda)) \text{ for all } \lambda, \mu \in \mathcal{H}^*$$

on $\mathcal{H}^*$ which clearly is positive definite, thus an inner product. It generates the same norm as $\|.\|_{\mathcal{H}^*}$ due to $\|\lambda\|_{\mathcal{H}^*} = \|R(\lambda)\|_{\mathcal{H}}$ for all $\lambda \in \mathcal{H}^*$, and we can rewrite it in the notation $(.,.)_* = (.,.)_{\mathcal{H}^*}$.

Altogether, we now have that R is an injective isometric linear map from $\mathcal{H}^*$ to $\mathcal{H}$ conserving the inner product. To prove that it is surjective, we can take any $f \in \mathcal{H}$ and generate a functional λ_f with

$$\lambda_f(g) := (f, g)_{\mathcal{H}} \text{ for all } g \in \mathcal{H}.$$

This functional clearly is continuous due to

$$|\lambda_f(g)| = |(f, g)_{\mathcal{H}}| \leq \|f\|_{\mathcal{H}} \|g\|_{\mathcal{H}} \text{ for all } g \in \mathcal{H}.$$

by Cauchy–Schwarz, and we compare f now with $R(\lambda_f)$ to get

$$\begin{aligned}(f - R(\lambda_f), g)_{\mathcal{H}} &= (f, g)_{\mathcal{H}} - (R(\lambda_f), g)_{\mathcal{H}} \\ &= (f, g)_{\mathcal{H}} - \lambda_f(g) \\ &= (f, g)_{\mathcal{H}} - (f, g)_{\mathcal{H}} = 0\end{aligned}$$

for all $g \in \mathcal{H}$, proving $f = R(\lambda_f)$. □

9.7 Reproducing Kernel Hilbert Spaces

As an important application of the Riesz theorem, we consider a Hilbert space $\mathcal{H}$ with an inner product $(.,.)_{\mathcal{H}}$ and assume that $\mathcal{H}$ consists of real–valued functions on a set Ω. Furthermore, we assume that all point–evaluation functionals

$$\delta_x \in \mathcal{H}^* \ : \ \mathcal{H} \to I\!R, \ f \mapsto f(x)$$

for all $x \in \Omega$ are continuous, i.e.

$$|\delta_x(f)| = |f(x)| \leq \|f\|_{\mathcal{H}} \|\delta_x\|_{\mathcal{H}^*} \text{ for all } f \in \mathcal{H}, \ x \in \Omega.$$

Then the Riesz map takes each δ_x into a function

$$R(\delta_x)(y) =: K(x,y) \text{ for all } x, y \in \Omega.$$

This defines a **kernel**

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

which is symmetric due to

$$\begin{array}{rcl} K(x,y) & := & R(\delta_x)(y) \\ & = & \delta_y(R(\delta_x)) \\ & = & (\delta_x, \delta_y)_{\mathcal{H}^*} \\ & = & (\delta_y, \delta_x)_{\mathcal{H}^*} \\ & = & K(y,x). \end{array}$$

Furthermore, we have

$$f(x) = \delta_x(f) = (f, R(\delta_x))_{\mathcal{H}} = (f, K(x,\cdot))_{\mathcal{H}} \text{ for all } x \in \Omega, \ f \in \mathcal{H},$$

and this is called a **reproduction equation**. In particular, when taking $f(\cdot) = K(y,\cdot)$, we get

$$K(y,x) = (K(y,\cdot), K(x,\cdot))_{\mathcal{H}} \text{ for all } x, y \in \Omega.$$

Then $\mathcal{H}$ is called a **reproducing kernel Hilbert space** with a reproducing kernel K. By the Riesz theorem, we have proven that all Hilbert spaces of functions with continuous point evaluations have a reproducing kernel.

9.8 Completion of Pre–Hilbert Spaces

Like the transition from rational numbers to real numbers by "completion", we can perform a transition from pre–Hilbert spaces to Hilbert spaces by a very similar technique via equivalence classes of Cauchy sequences.

Theorem 9.19 *Let $\mathcal{H}$ be a pre-Hilbert space with inner product $(\cdot,\cdot)_{\mathcal{H}}$. Then there is a Hilbert space $\mathcal{J}$ and an isometric embedding $J : \mathcal{H} \to \mathcal{J}$ such that the following is true:*

1. *$J(\mathcal{H})$ is dense in $\mathcal{J}$.*
2. *Any continuous mapping $A : \mathcal{H} \to \mathcal{N}$ with values in a Banach space $\mathcal{N}$ has a unique continuous extension $B : \mathcal{J} \to \mathcal{N}$ such that $B \circ J = A$.*
3. *The Hilbert space $\mathcal{J}$ is unique up to a Hilbert space isometry.*

Proof: We first form the space of all Cauchy sequences in $\mathcal{H}$, which clearly is a linear space over $\mathbb{R}$. Two such sequences are called equivalent, if their difference is a sequence in $\mathcal{H}$ converging to zero. The space $\mathcal{J}$ now is defined as the space of equivalence classes of Cauchy sequences in $\mathcal{H}$ modulo zero sequences. These classes clearly form a vector space under the usual operations on sequences. If we use an overstrike to stand for "class of", we write an element of $\mathcal{J}$ as $\overline{\{u_n\}_n}$ for a Cauchy sequence $\{u_n\}_n \in \mathcal{H}$. Now it is time to define an inner product

$$(\overline{\{u_n\}_n}, \overline{\{v_n\}_n})_{\mathcal{J}} := \lim_{n\to\infty} (u_n, v_n)_{\mathcal{H}}$$

on $\mathcal{J}$ and the embedding J via the constant Cauchy sequences

$$Ju := \{u\}_n := \{u_n = u\}_n$$

for each $u \subset \mathcal{H}$. Then

$$(Ju, Jv)_{\mathcal{J}} = (u, v)_{\mathcal{H}}$$

makes sure that J is an isometry and injective. But we still have to show that the inner product on $\mathcal{J}$ is well-defined and positive definite. If $\{u_n\}_n$ and $\{v_n\}_n$ are Cauchy sequences in $\mathcal{H}$, then

$$|\|u_n\|_{\mathcal{H}} - \|u_m\|_{\mathcal{H}}| \le \|u_n - u_m\|_{\mathcal{H}}$$

implies that the sequences $\{\|u_n\|_{\mathcal{H}}\}_n$ and $\{\|v_n\|_{\mathcal{H}}\}_n$ are Cauchy sequences in $\mathbb{R}$, and thus convergent and bounded by constants C_u and C_v. But then

$$\begin{aligned}(u_n, v_n)_{\mathcal{H}} - (u_m, v_m)_{\mathcal{H}} &= (u_n, v_n)_{\mathcal{H}} - (u_n, v_m)_{\mathcal{H}} - (u_m, v_m)_{\mathcal{H}} + (u_n, v_m)_{\mathcal{H}} \\ &= (u_n, v_n - v_m)_{\mathcal{H}} - (u_m - u_n, v_m)_{\mathcal{H}} \\ &\le C_u\|v_n - v_m\|_{\mathcal{H}} + C_v\|u_m - u_n\|_{\mathcal{H}}\end{aligned}$$

proves that $\{(u_n, v_n)_{\mathcal{H}}\}_n$ is a Cauchy sequence in $\mathbb{R}$ and thus convergent. Two representatives of a class $\overline{\{u_n\}_n}$ differ just by a zero sequence that does not affect the inner product's value. The proof of definiteness again uses that zero sequences represent zero in $\mathcal{J}$. This finishes the proof of well-definedness of the new inner product.

Thus $\mathcal{J}$ is another pre-Hilbert space that contains an isometric image of $\mathcal{H}$, and we first want to prove that $J(\mathcal{H})$ is dense in $\mathcal{J}$. Let us take an element $\overline{\{u_n\}_n} \in \mathcal{J}$ and use the fact that for each $\epsilon > 0$ there is some $K(\epsilon)$ such that for all $n, m \ge K(\epsilon)$ we have

$$\|u_n - u_m\|_{\mathcal{H}} \le \epsilon.$$

Now take $m \ge K(\epsilon)$ and the fixed Cauchy sequence $\overline{\{u_m\}_n} = J(u_m)$. Then

$$\|J(u_m) - \overline{\{u_n\}_n}\|_{\mathcal{J}} = \lim_{n\to\infty} \|u_m - u_n\|_{\mathcal{H}} \le \epsilon$$

proves the density assertion.

We now proceed to prove completeness of $\mathcal{J}$. To do this we have to form a Cauchy sequence $\{\overline{\{u_n^{(m)}\}_n}\}_m$ of equivalence classes $\overline{\{u_n^{(m)}\}_n}$ of Cauchy sequences $\{u_n^{(m)}\}_n \subset \mathcal{H}$. For each $m \in \mathbb{N}$ we can use the density property of $\mathcal{H}$ in $\mathcal{J}$ to find an element $v_m \in \mathcal{H}$ such that

$$\|\overline{\{u_n^{(m)}\}_n} - J(v_m)\|_{\mathcal{J}} \le 1/m.$$

Due to

$$\begin{aligned}\|v_n - v_m\|_{\mathcal{H}} &= \|J(v_n) - \overline{J(v_m)}\|_{\mathcal{J}} \\ &\leq \|J(v_n) - \overline{\{u_n^{(n)}\}_n}\|_{\mathcal{J}} + \\ &\quad + \|\overline{\{u_n^{(n)}\}_n} - \overline{\{u_n^{(m)}\}_n}\|_{\mathcal{J}} + \|\overline{\{u_n^{(m)}\}_n} - J(v_m)\|_{\mathcal{J}} \\ &\to 0\end{aligned}$$

for $n, m \to \infty$, the sequence $\{v_m\}_m$ is a Cauchy sequence in $\mathcal{H}$. We now form

$$\begin{aligned}\|\overline{\{u_n^{(k)}\}_n} - \{v_n\}_n\|_{\mathcal{J}} &\leq \|\overline{\{u_n^{(k)}\}_n} - J(v_k)\|_{\mathcal{J}} + \|J(v_k) - \{v_n\}_n\|_{\mathcal{J}} \\ &\leq 1/k + \lim_{n\to\infty} \|v_k - v_n\|_{\mathcal{H}} \\ &\to 0\end{aligned}$$

for $k \to \infty$, proving convergence towards $\{v_n\}_n$.

Now let $A : \mathcal{H} \to \mathcal{N}$ be a linear continuous mapping with values in a complete normed linear space $\mathcal{N}$. If $\overline{\{u_n\}_n}$ is an element of $\mathcal{J}$, we define the extension B on $\overline{\{u_n\}_n}$ by

$$B(\overline{\{u_n\}_n}) := \lim_{n\to\infty} A(u_n). \tag{9.20}$$

Since A is continuous, it is bounded and due to

$$\|A(u_m) - A(u_n)\|_{\mathcal{N}} \leq \|A\| \|u_m - u_n\|_{\mathcal{H}}$$

the sequence $\{Au_n\}_n$ is a Cauchy sequence in $\mathcal{N}$. But as $\mathcal{N}$ is a Banach space, the sequence is convergent and (9.20) is well-defined. Clearly $B \circ J = A$ holds by definition. Any two such extensions must agree on the dense subspace $A(\mathcal{H})$ of $\mathcal{J}$, and since they are continuous, they must agree on all of $\mathcal{J}$.

Finally, if there are two completions $\mathcal{J}$ and $\tilde{\mathcal{J}}$, we apply the first parts of the theorem to the embeddings

$$\begin{aligned} J &: \mathcal{H} \to \mathcal{J} \\ \tilde{J} &: \mathcal{H} \to \tilde{\mathcal{J}} \end{aligned}$$

This leads to two continuous maps

$$\begin{aligned} B &: \mathcal{J} \to \tilde{\mathcal{J}}, \quad B \circ J = \tilde{J} \\ \tilde{B} &: \tilde{\mathcal{J}} \to \mathcal{J}, \quad \tilde{B} \circ \tilde{J} = J \end{aligned} \tag{9.21}$$

and we conclude

$$\begin{aligned} \tilde{B} \circ B &= Id \quad \text{on } J(\mathcal{H}) \\ B \circ \tilde{B} &= Id \quad \text{on } \tilde{J}(\mathcal{H}) \end{aligned}$$

and this extends continuously to the completion, thus

$$\begin{aligned} \tilde{B} \circ B &= Id \quad \text{on } \mathcal{J} \\ B \circ \tilde{B} &= Id \quad \text{on } \tilde{\mathcal{J}}. \end{aligned}$$

But then we have isomorphisms between $\mathcal{J}$ and $\tilde{\mathcal{J}}$ which must be isometric due to (9.21). The isometry property follows first on the dense subspaces $J(\mathcal{H})$ and $\tilde{J}(\mathcal{H})$, but then also on the completions $\mathcal{J}$ and $\tilde{\mathcal{J}}$. □

9.9 Applications

As an application, consider a pre–Hilbert space $\mathcal{H}$ of functions on some set Ω having a reproducing kernel $K : \Omega \times \Omega \to I\!R$ in the sense

$$f(x) = (f, K(x,\cdot))_{\mathcal{H}} \text{ for all } f \in \mathcal{H},\ x \in \Omega.$$

We now go to the completion $\mathcal{J}$ with the embedding $J : \mathcal{H} \to \mathcal{J}$. The linear functionals

$$\delta_x : f \mapsto f(x) \text{ and } \lambda_x : f \mapsto (f, K(x,\cdot))_{\mathcal{H}}$$

coincide on $\mathcal{H}$ and are continuous there. Thus there is an extension

$$\tilde{\lambda}_x : \mathcal{J} \to I\!R$$

with

$$\tilde{\lambda}_x(J(f)) = \lambda_x(f) = (f, K(x,\cdot))_{\mathcal{H}} = f(x) \text{ for all } x \in \Omega.$$

On general elements $g \in \mathcal{J}$ we can define the functional

$$g \mapsto \tilde{\lambda}_x(g) - (g, J(K(x,\cdot)))_{\mathcal{J}}$$

which is in $\mathcal{J}^*$ and vanishes on the dense subset $J(\mathcal{H})$ due to

$$\begin{aligned} (J(f), J(K(x,\cdot)))_{\mathcal{J}} &= (f, K(x,\cdot))_{\mathcal{H}} \\ &= f(x) \\ &= \tilde{\lambda}_x(J(f)) \text{ for all } f \in \mathcal{H},\ x \in \Omega. \end{aligned}$$

Thus the functional is zero, proving the identity

$$\tilde{\lambda}_x(g) = (g, J(K(x,\cdot)))_{\mathcal{J}} \text{ for all } g \in \mathcal{J},\ x \in \Omega.$$

This equation can be interpreted as follows. Each abstract element $g \in \mathcal{J}$ is a function on Ω in the sense that

$$g(x) := \tilde{\lambda}_x(g) = (g, J(K(x,\cdot)))_{\mathcal{J}} \text{ for all } x \in \Omega.$$

Via

$$\tilde{K}(x,\cdot) := J(K(x,\cdot)) \text{ for all } x \in \Omega$$

we get a kernel $\tilde{K}$ such that the reproduction equation

$$g(x) = (g, \tilde{K}(x,\cdot))_{\mathcal{J}} \text{ for all } x \in \Omega,\ g \in \mathcal{J}$$

holds on $\mathcal{J}$. This proves that the original kernel, when embedded into the completion $\mathcal{J}$ of $\mathcal{H}$, still works as a reproducing kernel in the completion, and the completion is not just an abstract construction, but rather a Hilbert space of functions on Ω.

Let us now look at sequence spaces from subsection 9.3. They can be viewed as spaces of functions on I with a continuous point evaluation which we simply define for a sequence $\xi = \{\xi_i\}_{i \in I}$ as

$$\xi(i) = \xi_i,\ i \in I.$$

The pre–Hilbert space $S_{0,\lambda,I}$ under its inner product $(.,.)_{\lambda,I}$ has a reproducing kernel

$$K(i,j) := \frac{\delta_{ij}}{\lambda_i},\ i,j \in I$$

since

$$\xi(j) = \sum_{i\in I} \lambda_i \xi_i K(j,i) = (\xi, K(j,\cdot))_{\lambda,I} \text{ for all } \xi \in S_{0,\lambda,I},\ j \in I.$$

By completion of $S_{0,\lambda,I}$ under its inner product, we get some Hilbert space $\mathcal{S}$ with a continuous embedding $J\ :\ S_{0,\lambda,I} \to \mathcal{S}$, and it is a sequence space because we have a reproduction equation using the extended kernel. Since we have an isometric embedding of $S_{0,\lambda,I}$ into the space $S_{\lambda,I}$ of (9.4), we get that $\mathcal{S}$ must be isometrically embedded in $S_{\lambda,I}$. To prove that $S_{\lambda,I} = \mathcal{S}$, we take an arbitrary element $\xi \in S_{\lambda,I}$ which is orthogonal to the subspace $S_{0,\lambda,I}$. But then

$$\xi_i = (\xi, K(i,\cdot))_{\lambda,I} = 0 \text{ for all } i \in I$$

proves $\xi = 0$. Note that this avoids a direct proof that the space $S_{\lambda,I}$ is a Hilbert space, using completion arguments instead.

Thus we see that $S_{\lambda,I}$ of (9.4) is the Hilbert space completion of $S_{0,\lambda,I}$, and it is a reproducing kernel Hilbert space with the kernel K defined as above. The Riesz map

$$R\ :\ S_{1/\lambda,I} = S^*_{\lambda,I} \to S_{\lambda,I}$$

comes out as

$$R(\{\mu_i\}_{i\in I}) := \left\{\frac{\mu_i}{\lambda_i}\right\}_{i\in I} \text{ for all } \{\mu_i\}_{i\in I} \in S_{1/\lambda,I}$$

and the kernel is the Riesz representer of the point evaluation functional, as readers will quickly verify.

10 Necessary Results from Real Analysis

10.1 Lebesgue Integration

10.1.1 L_2 spaces

Lemma 10.1 *The shift operator $S_z\ :\ f(\cdot) \mapsto f(\cdot - z)$ is a continuous function of z near zero in the following sense: for each given $u \in L_2(\mathbb{R}^d)$ and each given $\epsilon > 0$ there is some $\delta > 0$ such that*

$$\|S_z(u) - u\|_{L_2(\mathbb{R}^d)} \leq \epsilon$$

for all $\|z\|_2 \leq \delta$.

Proof: to be supplied later....

We now want to prove that the space $\mathcal{S}$ of tempered test functions is dense in $L_2(\mathbb{R}^d)$. For this, we have to study functions like (10.12) in some more detail. They are in $\mathcal{S}$ for all positive values of ϵ, and Lemma 10.13 tells us that the operation

$$f \mapsto M_\epsilon(f) := \int_{\mathbb{R}^d} f(y)\varphi(\epsilon, \cdot - y)dy$$

maps each continuous L_1 function f to a "mollified" function $M_\epsilon(f)$ such that

$$\lim_{\epsilon \to 0} M_\epsilon(f)(x) = f(x)$$

uniformly on compact subsets of $\mathbb{R}^d$.

It is common to replace the Gaussian in (10.15) by an infinitely differentiable function with compact support, e.g.

$$\varphi(\epsilon, x) = \left\{ \begin{array}{ll} c(\epsilon) \exp(-1/(\epsilon^2 - \|x\|_2^2)) & \|x\|_2 < \epsilon \\ 0 & \|x\|_2 \geq \epsilon \end{array} \right\} \tag{10.2}$$

where the constant $c(\epsilon)$ is such that

$$\int_{\mathbb{R}^d} \varphi(\epsilon, x) dx = 1$$

holds for all $\epsilon > 0$. This **Friedrich's mollifier** can also be used in the definition of M_ϵ. It has the advantage that Lemma 10.13 holds for more general functions, i.e.: for functions which are in L_1 only locally.

We now want to study the action of M_ϵ on L_2 functions. Let $u \in L_2(\mathbb{R}^d)$ be given, and apply the Cauchy-Schwarz inequality to

$$M_\epsilon(f)(x) = \int_{\mathbb{R}^d} (f(y) \sqrt{\varphi(\epsilon, x-y)}) \sqrt{\varphi(\epsilon, x-y)} dy$$

to get

$$\begin{array}{rcl} |M_\epsilon(f)(x)|^2 & \leq & \int_{\mathbb{R}^d} |f(y)|^2 \varphi(\epsilon, x-y) dy \int_{\mathbb{R}^d} \varphi(\epsilon, x-y) dy \\ & = & \int_{\mathbb{R}^d} |f(y)|^2 \varphi(\epsilon, x-y) dy \end{array}$$

and

$$\int_{\mathbb{R}^d} |M_\epsilon(f)(x)|^2 dx \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(y)|^2 \varphi(\epsilon, z) dy dz = \int_{\mathbb{R}^d} |f(y)|^2 dy$$

such that M_ϵ has norm less than or equal to one in the L_2 norm. It is even more simple to prove the identity

$$(f, M_\epsilon g)_{L_2(R^d)} = (M_\epsilon f, g)_{L_2(R^d)}$$

for all $f, g \in L_2(\mathbb{R}^d)$ by looking at the integrals. Here, we used the Fubini theorem on $\mathbb{R}^d$ which requires some care, but there are no problems because everything can either be done with a Friedrich's mollifier, or be done on sufficiently large compact sets first, letting the sets tend to $\mathbb{R}^d$ later.

We now use a Friedrich's mollifier to study the L_2 error of the mollification. This is very similar to the arguments we already know. The error is representable pointwise as

$$f(x) - M_\epsilon(f)(x) = \int_{\mathbb{R}^d} (f(x) - f(y)) \varphi(\epsilon, x-y) dy$$

and we can use the Cauchy-Schwarz inequality to get

$$|f(x) - M_\epsilon(f)(x)|^2 \leq \int_{\|x-y\|_2 < \epsilon} |f(x) - f(y)|^2 \varphi(\epsilon, x-y) dy.$$

This can be integrated to get

$$\int_{I\!R^d} |f(x) - M_\epsilon(f)(x)|^2 dx \leq \int_{\|z\|_2<\epsilon} \varphi(\epsilon, z) \int_{I\!R^d} |f(y+z) - f(y)|^2 dy dz,$$

and we use the continuity of the shift operator as proven in Lemma 10.1 to make this as small as we want by picking a suitably small ϵ. This shows

$$\lim_{\epsilon \to 0} \|f - M_\epsilon(f)\|_{L_2(I\!R^d)} = 0$$

and proves

Lemma 10.3 *The space $\mathcal{S}$ of test functions is dense in $L_2(I\!R^d)$.* □

Lemma 10.4 *The space $C_0^\infty(I\!R^d)$ of compactly supported infinitely differentiable functions is dense in $L_2(I\!R^d)$.*

Proof: We can use a standard density lemma to go over from an $f \in \tilde{L}_2(I\!R^d)$ to a compactly supported function, and then we can use Friedrich's mollifier to generate an infinitely differentiable function. Both processes work with arbitrarily small L_2 errors. □

10.2 Fourier Transforms on $I\!R^d$

This section concerns an important tool for analysis of kernels on $I\!R^d$. There are two major possibilities to pick a space $\mathcal{S}$ of test functions on $I\!R^d$ to start with, and we take the **tempered test functions** forming **Schwartz** space $\mathcal{S}$ that are verbally defined as complex-valued functions on $I\!R^d$ whose partial derivatives exist for all orders and decay faster than any polynomial towards infinity.

Definition 10.5 *For a test function $u \in \mathcal{S}$, the* **Fourier transform** *is*

$$\widehat{u}(\omega) := (2\pi)^{-d/2} \int_{I\!R^d} u(x) e^{-ix\cdot\omega} dx, \tag{10.6}$$

where ω varies in $I\!R^d$ and $x \cdot \omega$ is shorthand for the scalar product $x^T\omega = \omega^T x$ to avoid the ^{T} symbol in the exponent. Since the definition even works for general $u \in L_1(I\!R^d)$, it is well-defined on $\mathcal{S}$ and clearly linear. Note that we use the **symmetric** form of the transform and do not introduce a factor 2π in the exponent of the exponential. This sometimes makes comparisons to other presentations somewhat difficult.

To get used to calculations of Fourier transforms, let us start with the **Gaussian** $u_\gamma(x) = \exp(-\gamma\|x\|_2^2)$ for $\gamma > 0$, which clearly is in the space of test functions, since all derivatives are polynomials multiplied with the Gaussian itself. As a byproduct we shall get that the Gaussian is positive definite on $I\!R^d$. Fortunately, the Gaussian can be written as a d-th power of the entire analytic function $\exp(-\gamma z^2)$, and we can thus work on $\mathbb{C}^d$ instead of $I\!R^d$. We simply use substitution in

$$\begin{aligned}
\widehat{u_\gamma}(i\omega) &= (2\pi)^{-d/2} \textstyle\int_{I\!R^d} e^{-\gamma\|x\|_2^2} e^{x\cdot\omega} dx \\
&= (2\pi)^{-d/2} e^{\|\omega\|_2^2/4\gamma} \textstyle\int_{I\!R^d} e^{-\|\sqrt{\gamma}x - \omega/2\sqrt{\gamma}\|_2^2} dx \\
&= (2\pi\gamma)^{-d/2} e^{\|\omega\|_2^2/4\gamma} \textstyle\int_{I\!R^d} e^{-\|y\|_2^2} dy
\end{aligned}$$

and are done up to the evaluation of the dimension-dependent constant

$$\int_{\mathbb{R}^d} e^{-\|y\|_2^2} dy =: c^d$$

which is a d-th power, because the integrand factorizes nicely. We calculate c^2 by using polar coordinates and get

$$\begin{aligned} c^2 &= \int_{\mathbb{R}^2} e^{-\|y\|_2^2} dy \\ &= \int_0^{2\pi} \int_0^\infty e^{-r^2} r \, dr \, d\varphi \\ &= 2\pi \int_0^\infty e^{-r^2} r \, dr \\ &= -\pi \int_0^\infty (-2r) e^{-r^2} \, dr \\ &= \pi. \end{aligned}$$

This proves the first assertion of

Theorem 10.7 *The Gaussian*

$$u_\gamma(x) = \exp(-\gamma \|x\|_2^2)$$

has Fourier transform

$$\widehat{u_\gamma}(\omega) = (2\gamma)^{-d/2} e^{-\|\omega\|_2^2/4\gamma} \tag{10.8}$$

and is unconditionally positive definite on $\mathbb{R}^d$.

To understand the second assertion, we add

Definition 10.9 *A real-valued function*

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

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

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

is positive definite.

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

Proof of the theorem: Let us first invert the Fourier transform by setting $\beta := 1/4\gamma$ in (10.8):

$$\begin{aligned} \exp(-\beta \|\omega\|_2^2) &= (4\pi\beta)^{-d/2} \int_{\mathbb{R}^d} e^{-\|x\|_2^2/4\beta} e^{-ix\cdot\omega} dx \\ &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} (2\beta)^{-d/2} e^{-\|x\|_2^2/4\beta} e^{+ix\cdot\omega} dx. \end{aligned}$$

Then take any set $X = \{x_1, \ldots, x_M\} \subset I\!R^d$ of M distinct points and any vector $\alpha \in I\!R^M$ to form

$$\begin{aligned}
\alpha^T A_{X,u_\gamma} \alpha &= \sum_{j,k=1}^{M} \alpha_j \alpha_k \exp(-\gamma \|x_j - x_k\|_2^2) \\
&= \sum_{j,k=1}^{M} \alpha_j \alpha_k (4\pi\gamma)^{-d/2} \int_{I\!R^d} e^{-\|x\|_2^2/4\gamma} e^{-ix\cdot(x_j - x_k)} dx \\
&= (4\pi\gamma)^{-d/2} \int_{I\!R^d} e^{-\|x\|_2^2/4\gamma} \sum_{j,k=1}^{M} \alpha_j \alpha_k e^{-ix\cdot(x_j - x_k)} dx \\
&= (4\pi\gamma)^{-d/2} \int_{I\!R^d} e^{-\|x\|_2^2/4\gamma} \left| \sum_{j=1}^{M} \alpha_j e^{-ix\cdot x_j} \right|^2 dx \geq 0.
\end{aligned}$$

This proves positive semidefiniteness of the Gaussian. To prove definiteness, we can assume

$$f(x) := \sum_{j=1}^{M} \alpha_j e^{-ix\cdot x_j} = 0$$

for all $x \in I\!R^d$ and have to prove that all coefficients α_j vanish. Taking derivatives at zero, we get

$$0 = D^\beta f(0) = \sum_{j=1}^{M} \alpha_j (-ix_j)^\beta,$$

and this is a homogeneous system for the coefficients α_j whose coefficient matrix is a generalized Vandermonde matrix, possibly transposed and with scalar multiples for rows or columns. This proves the assertion in one dimension, where the matrix corresponds to the classical Vandermonde matrix. The multivariate case reduces to the univariate case by picking a nonzero vector $y \in I\!R^d$ that is not orthogonal to any of the finitely many differences $x_j - x_k$ for $j \neq k$. Then the real values $y \cdot x_j$ are all distinct for $j = 1, \ldots, M$ and one can consider the univariate function

$$g(t) := f(ty) = \sum_{j=1}^{M} \alpha_j e^{-ity\cdot x_j} = 0$$

which does the job in one dimension. □

Note that the Gaussian is mapped to itself by the Fourier transform, if we pick $\gamma = 1/2$. We shall use the Gaussian's Fourier transform in the proof of the fundamental **Fourier Inversion Theorem**:

Theorem 10.10 *The Fourier transform is bijective on $\mathcal{S}$, and its inverse is the transform*

$$\check{u}(x) := (2\pi)^{-d/2} \int_{I\!R^d} u(\omega) e^{ix\cdot\omega} d\omega. \tag{10.11}$$

Proof: The multivariate derivative D^α of $\widehat{u}$ can be taken under the integral sign, because u is in $\mathcal{S}$. Then

$$(D^\alpha \widehat{u})(\omega) = (2\pi)^{-d/2} \int_{I\!R^d} u(x)(-ix)^\alpha e^{-ix\cdot\omega} dx,$$

and we multiply this by ω^β and use integration by parts

$$\begin{aligned}
\omega^\beta (D^\alpha \widehat{u})(\omega) &= (2\pi)^{-d/2} \textstyle\int_{I\!R^d} u(x)(-ix)^\alpha (i)^\beta (-i\omega)^\beta e^{-ix\cdot\omega} dx \\
&= (2\pi)^{-d/2} \textstyle\int_{I\!R^d} u(x)(-ix)^\alpha (i)^\beta \frac{d^\beta}{dx^\beta} e^{-ix\cdot\omega} dx \\
&= (2\pi)^{-d/2} (-1)^{|\alpha|+|\beta|} i^{\alpha+\beta} \textstyle\int_{I\!R^d} e^{-ix\cdot\omega} \frac{d^\beta}{dx^\beta} (u(x)x^\alpha) dx
\end{aligned}$$

to prove that $\widehat{u}$ lies in $\mathcal{S}$, because all derivatives decay faster than any polynomial towards infinity. The second assertion follows from the Fourier inversion formula

$$u(x) := (2\pi)^{-d/2} \int_{I\!R^d} \widehat{u}(\omega) e^{ix\cdot\omega} d\omega$$

that we now prove for all $u \in \mathcal{S}$. This does not work directly if we naively put the definition of $\widehat{u}$ into the right-hand-side, because the resulting multiple integral does not satisfy the assumptions of Fubini's theorem. We have to do a regularization of the integral, and since this is a standard trick, we write it out in some detail:

$$\begin{aligned}
(2\pi)^{-d/2} \textstyle\int_{I\!R^d} \widehat{u}(\omega) e^{ix\cdot\omega} d\omega &= (2\pi)^{-d} \textstyle\int_{I\!R^d} \int_{I\!R^d} u(y) e^{i(x-y)\cdot\omega} dy d\omega \\
&= \lim_{\epsilon \searrow 0} (2\pi)^{-d} \int_{I\!R^d} \int_{I\!R^d} u(y) e^{i(x-y)\cdot\omega - \epsilon\|\omega\|_2^2} dy d\omega \\
&= \lim_{\epsilon \searrow 0} (2\pi)^{-d} \int_{I\!R^d} \left(\int_{I\!R^d} e^{i(x-y)\cdot\omega - \epsilon\|\omega\|_2^2} d\omega \right) u(y) dy \\
&= \lim_{\epsilon \searrow 0} \int_{I\!R^d} \varphi(\epsilon, x-y) u(y) dy
\end{aligned}$$

with

$$\varphi(\epsilon, z) := (2\pi)^{-d} \int_{I\!R^d} e^{iz\cdot\omega - \epsilon\|\omega\|_2^2} d\omega. \tag{10.12}$$

The proof is completed by application of the following result that is useful in many contexts: □

Lemma 10.13 *The family of functions $\varphi(\epsilon, z)$ of (10.12) approximates the point evaluation functional in the sense*

$$u(x) = \lim_{\epsilon \searrow 0} \int_{I\!R^d} \varphi(\epsilon, x-y) u(y) dy \tag{10.14}$$

for all functions u that are in $L_1(I\!R^d)$ and continuous around x.

Proof: We first remark that φ is a disguised form of the inverse Fourier transform equation of the Gaussian. Thus we get

$$\varphi(\epsilon, x) = (4\pi\epsilon)^{-d/2} e^{-\|x\|_2^2/4\epsilon} \tag{10.15}$$

and

$$\int_{I\!R^d} \varphi(\epsilon, x) dx = (4\pi\epsilon)^{-d/2} \int_{I\!R^d} e^{-\|x\|_2^2/4\epsilon} dx = 1.$$

To prove (10.14), we start with some given $\delta > 0$ and first find some ball $B_\rho(x)$ of radius $\rho(\delta)$ around x such that $|u(x) - u(y)| \le \delta/2$ holds uniformly for all $y \in B_\rho(x)$. Then we split the integral in

$$\begin{aligned}
|u(x) - \textstyle\int_{I\!R^d} \varphi(\epsilon, x-y) u(y) dy| &= |\textstyle\int_{I\!R^d} \varphi(\epsilon, x-y)(u(x) - u(y)) dy| \\
&\le \textstyle\int_{\|y-x\|_2 \le \rho} \varphi(\epsilon, x-y) |u(x) - u(y)| dy \\
&\quad + \textstyle\int_{\|y-x\| > \rho} \varphi(\epsilon, x-y) |u(x) - u(y)| dy \\
&\le \delta/2 + (4\pi\epsilon)^{-d/2} e^{-\rho^2/4\epsilon} 2\|u\|_1 \\
&\le \delta
\end{aligned}$$

for all sufficiently small ϵ. □

Due to the Fourier inversion formula, we now know that the Fourier transform is bijective on $\mathcal{S}$.

We now relate the Fourier transform to the L_2 inner product, but we have to use the latter over $\mathbb{C}$ to account for the possibly complex values of the Fourier transform. We define the inner product as

$$(f,g)_{L_2(\mathbb{R}^d)} := \int_{\mathbb{R}^d} f(x)\overline{g(x)}dx \tag{10.16}$$

without factors that sometimes are used.

Fubini's theorem easily proves the identity

$$(v,\widehat{u})_{L_2(\mathbb{R}^d)} = (2\pi)^{-d/2}\int_{\mathbb{R}^d} v(x)\int_{\mathbb{R}^d}\overline{u(y)}e^{+ix\cdot y}dydx = (\check{v},u)_{L_2(\mathbb{R}^d)}$$

for all test functions $u,v\in\mathcal{S}$. Setting $v=\widehat{w}$ we get the **Plancherel equation**

$$(\widehat{w},\widehat{u})_{L_2(\mathbb{R}^d)} = (w,u)_{L_2(\mathbb{R}^d)} \tag{10.17}$$

for the Fourier transform on $\mathcal{S}$, proving that the Fourier transform is isometric on $\mathcal{S}$ as a subspace of $L_2(\mathbb{R}^d)$.

10.3 Fourier Transform in $L_2(\mathbb{R}^d)$

The test functions from $\mathcal{S}$ are dense in $L_2(\mathbb{R}^d)$ (see Lemma 10.3 for details), and thus we have

Theorem 10.18 *The Fourier transform has an L_2-isometric extension from the space $\mathcal{S}$ of tempered test functions to $L_2(\mathbb{R}^d)$. The same holds for the inverse Fourier transform, and both extensions are inverses of each other in $L_2(\mathbb{R}^d)$. Furthermore, Parceval's equation (10.17) holds in $L_2(\mathbb{R}^d)$.* □

Note that this result does not allow to use the Fourier transform formula (or its inverse) in the natural pointwise form. For any $f\in L_2(\mathbb{R}^d)$ one first has to provide a sequence of test functions $v_n\in\mathcal{S}$ that converges to f in the L_2 norm for $n\to\infty$, and then, by continuity, the image $\widehat{f}$ of the Fourier transform is uniquely defined almost everywhere by

$$\lim_{n\to\infty}\|\widehat{f}-\widehat{v_n}\|_{L_2(\mathbb{R}^d)} = 0.$$

This can be done via Friedrich's mollifiers as defined in (10.2), replacing the Gaussian in the representation (10.15) by a compactly supported infinitely differentiable function.

A more useful characterization of $\widehat{f}$ is the variational equation

$$(\widehat{f},v)_{L_2(\mathbb{R}^d)} = (f,\check{v})_{L_2(\mathbb{R}^d)}$$

for all test functions $v\in\mathcal{S}$, or, by continuity, all functions $v\in L_2(\mathbb{R}^d)$.

10.4 Poisson Summation Formula

This comes in several forms:

$$\begin{aligned}
(2\pi)^{-d/2}\sum_{k\in\mathbb{Z}^d}\hat{u}(k) &= \sum_{j\in\mathbb{Z}^d}u(2\pi j)\\
(2\pi)^{-d/2}\sum_{k\in\mathbb{Z}^d}\hat{u}(k)e^{ik^Tx} &= \sum_{j\in\mathbb{Z}^d}u(x+2\pi j)\\
(2\pi)^{-d/2}\sum_{k\in\mathbb{Z}^d}u(k)e^{-ik^T\omega} &= \sum_{j\in\mathbb{Z}^d}\hat{u}(\omega+2\pi j)\\
(2\pi)^{-d/2}\sum_{k\in\mathbb{Z}^d}u(hk)e^{-ihk^T\omega} &= h^{-d}\sum_{j\in\mathbb{Z}^d}\hat{u}\left(\omega+\frac{2\pi j}{h}\right)
\end{aligned}$$

but we shall have to assure in which sense and under which assumptions it holds. The first clearly is a consequence of the second, if the second holds pointwise. But we shall not discuss this here. The final two are variations of the second, as follows from standard transformations.

Thus we focus on the second one first and see it as an equation in $L_2(\mathbb{R}^d)$. Both sides are 2π-periodic, and the left-hand side can be viewed as the Fourier series representation of the right-hand side. Thus we assume that the right-hand side is a pointwise absolutely convergent series which is also convergent in $L_2[-\pi,\pi]^d$. To make the left-hand side meaningful, we assume that u is in $L_1(\mathbb{R}^d)$.

If we write the Fourier analysis of a d-variate 2π-periodic function $f(x)$ as

$$f(x) = \sum_{k\in\mathbb{Z}^d} c_k e^{ik^T x}, \quad c_k = (2\pi)^{-d} \int_{[-\pi,\pi]^d} f(x) e^{-ik^T x} dx,$$

we can apply this to the right-hand side f of the second form of the Poisson summation formula. We get the coefficient

$$\begin{aligned}
c_k &= (2\pi)^{-d} \int_{[-\pi,\pi]^d} f(x) e^{-ik^T x} dx \\
&= (2\pi)^{-d} \int_{[-\pi,\pi]^d} \sum_{j\in\mathbb{Z}^d} u(x+2\pi j) e^{-ik^T x} dx \\
&= (2\pi)^{-d} \int_{[-\pi,\pi]^d} \sum_{j\in\mathbb{Z}^d} u(x+2\pi j) e^{-ik^T (x+2\pi j)} dx \\
&= (2\pi)^{-d} \int_{\mathbb{R}^d} u(t) e^{-ik^T t} dt \\
&= (2\pi)^{-d/2} \hat{u}(k)
\end{aligned}$$

under our assumptions. Note that the above argument uses only L_2–continuous transformations. This proves the second equation.

The third form can be deduced exactly like the second one, if we also interchange the role of u and $\hat{u}$ in the assumptions. Formally, we can use the second for $\hat{u}$ instead of u and apply

$$\hat{\hat{u}}(k) = \hat{u}^\vee(-k) = u(-k).$$

The final form takes $v(x) := u(hx)$ and applies the third inequality with

$$\hat{v}(\omega) = h^{-d} \hat{u}\left(\frac{\omega}{h}\right)$$

following from

$$\begin{aligned}
\hat{v}(\omega) &= (2\pi)^{-d/2} \int v(x) e^{-ix^T\omega} dx \\
&= (2\pi)^{-d/2} \int u(hx) e^{-ihx^T\omega/h} dx \\
&= h^{-d} (2\pi)^{-d/2} \int u(y) e^{-iy^T\omega/h} dx \\
&= h^{-d} \hat{u}\left(\tfrac{\omega}{h}\right).
\end{aligned}$$

This yields

$$\begin{aligned}(2\pi)^{-d/2}\sum_{k\in\mathbb{Z}^d} v(k)e^{-ik^T\eta} &= \sum_{j\in\mathbb{Z}^d}\hat{v}(\eta+2\pi j)\\ (2\pi)^{-d/2}\sum_{k\in\mathbb{Z}^d} u(hk)e^{-ik^T\eta} &= h^{-d}\sum_{j\in\mathbb{Z}^d}\hat{u}\left(\frac{\eta+2\pi j}{h}\right)\\ (2\pi)^{-d/2}\sum_{k\in\mathbb{Z}^d} u(hk)e^{-ihk^T\omega} &= h^{-d}\sum_{j\in\mathbb{Z}^d}\hat{u}\left(\omega+\frac{2\pi j}{h}\right)\end{aligned}$$

for $\eta =: h\omega$. But note that the above form is badly scaled. It should read

$$h^{d/2}\sum_{k\in\mathbb{Z}^d} u(hk)e^{-ihk^T\omega} = \left(\frac{2\pi}{h}\right)^{d/2}\sum_{j\in\mathbb{Z}^d}\hat{u}\left(\omega+\frac{2\pi j}{h}\right)$$

in order to represent the fact that the left-hand side is a summation over gridpoints with spacing h, while the right-hand side is a summation over a grid with spacing $\frac{2\pi}{h}$.

10.5 Fourier Transforms of Functionals

With Plancherel's equation in mind, let us look at the linear functional

$$\lambda_u(v) := (u,v)_{L_2(\mathbb{R}^d)}$$

on $\mathcal{S}$. We see that

$$\lambda_{\widehat{u}}(v) = (\widehat{u},v)_{L_2(\mathbb{R}^d)} = (u,\check{v})_{L_2(\mathbb{R}^d)} = \lambda_u(\check{v})$$

holds. A proper definition of the Fourier transform for functionals λ_u should be in line with the functions u that represent them, and thus we should define

$$\widehat{\lambda_u} := \lambda_{\widehat{u}}$$

or in more generality

$$\widehat{\lambda}(v) := \lambda(\check{v})$$

for all $v \in \mathcal{S}$. Since the space $\mathcal{S}$ of test functions is quite small, its dual, the space of linear functionals on $\mathcal{S}$, is quite large. In particular, the functionals of the form λ_u are defined on all of $\mathcal{S}$, if u is a **tempered function**. The latter form the space $\mathcal{K}$ of all continuous functions on $\mathbb{R}^d$ that grow at most polynomially for arguments tending to infinity.

Definition 10.19 *The Fourier transform of a linear functional λ on $\mathcal{S}$ is the linear functional $\widehat{\lambda}$ on $\mathcal{S}$ defined by*

$$\widehat{\lambda}(v) := \lambda(\check{v}) \text{ or } \widehat{\lambda}(\widehat{v}) := \lambda(v)$$

for all $v \in \mathcal{S}$. If the latter can be represented in the form λ_w with a tempered function $w \in \mathcal{K}$, we say that w is the Fourier transform of λ and write $w = \widehat{\lambda}$. The **generalized Fourier transform** *of a tempered function $u \in \mathcal{K}$ is the Fourier transform $\widehat{\lambda_u}$ of the functional λ_u.*

Example 10.20 *The functional $\delta_x(v) := v(x)$ has the form*

$$\delta_x(v) = v(x) = (2\pi)^{-d/2}\int_{\mathbb{R}^d}\widehat{v}(\omega)e^{+ix\cdot\omega}d\omega,$$

and its Fourier transform is of the form λ_{u_x} with

$$u_x(\omega) = \widehat{\delta_x}(\omega) = e^{-ix\cdot\omega}.$$

Here, the normalization of the L_2 inner product (10.16) pays off. Note that the Fourier transform is not a test function, but rather a tempered function from $\mathcal{K}$ and in particular a bounded function. The functional $\delta := \delta_0$ has the Fourier transform $u_0 = 1$.

Example 10.21 *A very important class of functionals for our purposes consists of the space $\mathcal{P}_\Omega^\perp = L$ of functionals of the form*

$$\lambda_{a,X} := \sum_{x_j \in X} a_j f(x_j) \tag{10.22}$$

for finite sets $X \subset \Omega$ and $a \in I\!R^{|X|}$ that vanish on $I\!P_m^d$. Their action on a test function v is

$$\begin{aligned}
\lambda_{a,X}(v) &= \sum_{j=1}^M a_j v(x_j) \\
&= (2\pi)^{-d/2} \int_{I\!R^d} \widehat{v}(\omega) \sum_{j=1}^M a_j e^{ix_j\cdot\omega} d\omega \\
&= \widehat{\lambda}_{a,X}(\widehat{v})
\end{aligned}$$

such that the Fourier transform of the functional $\lambda_{a,X}$ is the functional generated by the bounded function

$$\widehat{\lambda}_{a,X}(\omega) = p_{a,X}(\omega) := \sum_{j=1}^M a_j e^{-ix_j\cdot\omega}.$$

If we expand the exponential into its power series, we see that

$$\begin{aligned}
\widehat{\lambda}_{a,X}(\omega) &= \sum_{k=0}^\infty \sum_{j=1}^M a_j (-ix_j\cdot\omega)^k/k! \\
&= \sum_{k=m}^\infty \sum_{j=1}^M a_j (-ix_j\cdot\omega)^k/k!
\end{aligned}$$

since the functional vanishes on $I\!P_m^d$. Thus $\widehat{\lambda}_{a,X}(\omega)$ has a zero of order at least m in the origin. If the functional $\lambda_{a,X}$ itself were representable by a function u, the function u should be L_2-orthogonal to all polynomials from $I\!P_m^d$. We shall use both of these facts later.

Example 10.23 *The monomials x^α are in the space $\mathcal{K}$, and thus they should at least have generalized Fourier transforms in the sense of functionals. This can easily be verified via*

$$\begin{aligned}
\left(-i\tfrac{d}{dx}\right)^\alpha v(x) &= \left(-i\tfrac{d}{dx}\right)^\alpha (2\pi)^{-d/2} \textstyle\int_{I\!R^d} \widehat{v}(\omega) e^{+ix\cdot\omega} d\omega \\
&= (2\pi)^{-d/2} \textstyle\int_{I\!R^d} \widehat{v}(\omega) (-i\cdot i\omega)^\alpha e^{+ix\cdot\omega} d\omega \\
&= (2\pi)^{-d/2} \textstyle\int_{I\!R^d} \widehat{v}(\omega) \omega^\alpha e^{+ix\cdot\omega} d\omega,
\end{aligned}$$

and the associated functional is

$$v \mapsto \left(-i\frac{d}{dx}\right)^\alpha v(x)$$

at $x = 0$.

10.6 Special Functions and Transforms

This is intended as a reference and tutorial for classical formulas involving special functions (e.g.: Gamma, Beta, and Bessel functions) and their transforms. Results on Fourier transforms in general are in section 10.2. This section, so far, is in raw and unsorted form, because all required formulae are just collected here.

10.6.1 Gamma Function

The **Gamma function** is defined by

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \tag{10.24}$$

and has the properties

$$\begin{aligned} \Gamma(z+1) &= z\Gamma(z), \quad z \notin -I\!N \\ \Gamma(k+1) &= k!, \quad k \in I\!N \\ \Gamma(1/2) &= \sqrt{\pi}. \end{aligned}$$

The equation

$$\int_0^1 u^{x-1}(1-u)^{y-1} du = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{10.25}$$

for any $x, y > 0$ will be useful.

10.6.2 Volumes and Surface Integrals

The volume of the d-dimensional ball

$$B_r(0) := \{\, x \in I\!R^d \;:\; \|x\|_2 \le r \,\}$$

of radius r is

$$\text{vol } B_r(0) = \frac{r^d \pi^{d/2}}{\Gamma(1+d/2)}. \tag{10.26}$$

The surface area σ_{d-1} of the $d-1$-dimensional sphere in $I\!R^d$ for $d \ge 1$ is

$$\sigma_{d-1} = \text{vol}\,(S^{d-1}) = 2\pi^{d/2}/\Gamma(d/2). \tag{10.27}$$

This follows for $d > 2$ from the representation

$$d\sigma = \prod_{j=1}^{d-1} (\sin\varphi_j)^{d-1-j} d\varphi_j$$

of the surface element $d\sigma$ in terms of the angles

$$\varphi_j \in [0,\pi],\ 1 \le j \le d-2,\ \varphi_{d-1} \in [0, 2\pi]$$

and univariate integration, while $d = 1, 2$ are standard.

10.6.3 Bessel Functions

For Bessel functions, the standard source of information is [68].

We consider the function $F(r\|\omega\|_2, d)$ defined by the integral

$$F(t,d) := \int_{\|y\|_2=1} e^{-ity\cdot z} dy \tag{10.28}$$

over the surface of the unit ball in $\mathbb{R}^d$ for $t \geq 0, d \geq 2$, and some $\|z\|_2 = 1,\ z \in \mathbb{R}^d$. This integral is invariant under orthogonal transformations Q of $\mathbb{R}^d$, as is easily obtainable from replacement of z by Qz. Thus the integral is independent of z, as already indicated by the notation, and we can assume $z = (-1, 0, \ldots, 0)$ for its evaluation. Let σ_{d-1} be the surface area of the $d-1$-sphere, i.e.: the boundary of the unit ball in $\mathbb{R}^d$. We now assume $d \geq 3$ and integrate over the surface of the $d-1$-sphere by summing up the integrals over surfaces of $(d-2)$-spheres, splitting $y = (y_1, u)$ and setting $z \cdot y = \cos\varphi$. This yields

$$\begin{aligned} F(t,d) &= \int_{\|y\|_2=1} e^{ity\cdot z} dy \\ &= \int_0^\pi e^{it\cos\varphi} \int_{\|u\|_2^2 = 1-y_1^2} du d\varphi \\ &= \sigma_{d-2} \int_0^\pi e^{it\cos\varphi} (\sin(\varphi))^{d-2} d\varphi \\ &= \sigma_{d-2} \int_{-1}^1 e^{its} (1-s^2)^{(d-3)/2} ds \end{aligned}$$

and contains an instance of the **Bessel function**

$$J_\nu(t) = \frac{(t/2)^\nu}{\Gamma(\frac{2\nu+1}{2})\Gamma(\frac{1}{2})} \int_{-1}^1 e^{its} (1-s^2)^{\frac{2\nu-1}{2}} ds \tag{10.29}$$

which is well-defined for $\Re(\nu) > -\frac{1}{2}$. We end up with $\nu = \frac{d-2}{2}$ and get

$$\begin{aligned} F(t,d) &= \sigma_{d-2} \frac{\Gamma(\frac{d-1}{2})\Gamma(\frac{1}{2})}{(t/2)^{(d-2)/2}} J_{(d-2)/2}(t) \\ &= 2\pi^{d/2} (t/2)^{-(d-2)/2} J_{(d-2)/2}(t). \end{aligned} \tag{10.30}$$

Direct integration shows that this is also valid for $d = 2$ or $\nu = 0$, using $\sigma_0 = 2$.

10.6.4 Power Series of Bessel Functions

The Bessel function of (10.29) has the power series representation

$$J_\nu(t) = \left(\frac{t}{2}\right)^\nu \sum_{j=0}^\infty \frac{\left(-\frac{t^2}{4}\right)^j}{j!\Gamma(\nu+j+1)} \tag{10.31}$$

that is valid for all $t \in \mathbb{C} \setminus \{0\}$ and all $\nu \in \mathbb{C}$. The integral representation (10.29) is first proven to be identical to the power series representation (10.31) on its domain of definition. Since the

power series is convergent everywhere, the general definition of J_ν can then be done by (10.31). We first expand the exponential in

$$\begin{aligned}\int_{-1}^{1} e^{its}(1-s^2)^{(2\nu-1)/2}ds &= \sum_{j=0}^{\infty}\frac{(it)^j}{j!}\int_{-1}^{1} s^j(1-s^2)^{(2\nu-1)/2}ds \\ &= \sum_{j=0}^{\infty}\frac{(it)^{2j}}{2j!}\int_{-1}^{1} s^{2j}(1-s^2)^{(2\nu-1)/2}ds\end{aligned}$$

and use symmetry to cancel the odd powers. The equation (10.25) will come in handy after the substitution $s^2 = u$. Then

$$\begin{aligned}\sum_{j=0}^{\infty}\frac{(it)^{2j}}{2j!}\int_{-1}^{1} s^{2j}(1-s^2)^{(2\nu-1)/2}ds &= \sum_{j=0}^{\infty}\frac{(it)^{2j}}{2j!}\int_{0}^{1} u^{j-1/2}(1-u)^{(2\nu-1)/2}du \\ &= \sum_{j=0}^{\infty}\frac{\Gamma(j+\frac{1}{2})\Gamma(\frac{2\nu+1}{2})}{\Gamma(j+\nu+1)}\frac{(it)^{2j}}{2j!} \\ &= \sum_{j=0}^{\infty}\frac{\Gamma(\frac{1}{2})\Gamma(\frac{2\nu+1}{2})}{j!\Gamma(j+\nu+1)}\left(-\frac{t^2}{4}\right)^j\end{aligned}$$

uses the same split of $\Gamma(j+\frac{1}{2})$ as before. This can be put int (10.29) to yield the power series representation.

Looking at (10.31), we can define a function H_ν by

$$\left(\tfrac{z}{2}\right)^{-\nu} J_\nu(z) =: H_\nu(z^2/4) = \sum_{k=0}^{\infty}\frac{(-z^2/4)^k}{k!\Gamma(k+\nu+1)} \tag{10.32}$$

for $\nu \in \mathbb{C}$. This function often occurs in the text.

In a very special situation the power series representation (10.31) implies

$$\begin{aligned}J_{-1/2}(t) &= \left(\frac{t}{2}\right)^{-1/2}\sum_{j=0}^{\infty}\frac{\left(-\frac{t^2}{4}\right)^j}{j!\Gamma(j+1/2)} \\ &= \left(\frac{t}{2}\right)^{-1/2}\sum_{j=0}^{\infty}\frac{(-1)^j t^{2j}}{2^{2j}j!((j-1)/2)((j-3)/2)\ldots(1/2)\sqrt{\pi}} \\ &= \left(\frac{t}{2}\right)^{-1/2}\sum_{j=0}^{\infty}\frac{(-1)^j t^{2j}}{(2j)!\sqrt{\pi}} \\ &= \left(\frac{t}{2}\right)^{-1/2}\frac{1}{\sqrt{\pi}}\cos(t) \\ &= \sqrt{\frac{2}{\pi}}\frac{\cos(t)}{\sqrt{t}},\end{aligned} \tag{10.33}$$

and the other Bessel functions with half-integer order are similarly obtainable as linear combinations of elementary functions.

10.6.5 Relations Between Bessel Functions

By differentiation of the H_ν function from (10.32) we get

$$\begin{aligned} -\frac{d}{dt}H_\nu(rt) &= -\tfrac{d}{dt}\sum_{k=0}^{\infty}\frac{(-rt)^k}{k!\Gamma(\nu+k+1)} \\ &= r\sum_{k=1}^{\infty}\frac{(-rt)^{k-1}}{(k-1)!\Gamma(\nu+k+1)} \\ &= r\sum_{k=0}^{\infty}\frac{(-rt)^k}{k!\Gamma(\nu+k+2)} \\ &= rH_{\nu+1}(rt). \end{aligned} \tag{10.34}$$

and

$$\begin{aligned} \frac{d}{dt}t^\nu H_\nu(rt) &= \tfrac{d}{dt}\sum_{k=0}^{\infty}\frac{(-rt)^k t^\nu}{k!\Gamma(\nu+k+1)} \\ &= r\sum_{k=0}^{\infty}\frac{(-r)^k(\nu+k)t^{\nu+k-1}}{k!\Gamma(\nu+k+1)} \\ &= \sum_{k=0}^{\infty}\frac{(-rt)^k t^{\nu-1}}{k!\Gamma(\nu-1+k+1)} \\ &= t^{\nu-1}H_{\nu-1}(rt). \end{aligned} \tag{10.35}$$

We further need a special identity for Bessel functions:

$$J_{\mu+\nu+1}(t) = \frac{t^{\nu+1}}{2^\nu\Gamma(\nu+1)}\int_0^1 J_\mu(ts)s^{\mu+1}(1-s^2)^\nu ds,\ t>0, \nu>-1, \mu>-\frac{1}{2}. \tag{10.36}$$

Since the integral is finite, we can simply insert the power series and get

$$\begin{aligned} \int_0^1 J_\mu(ts)s^{\mu+1}(1-s^2)^\nu ds &= \int_0^1\left(\left(\frac{ts}{2}\right)^\mu\sum_{j=0}^{\infty}\frac{\left(-\frac{(ts)^2}{4}\right)^j}{j!\Gamma(\mu+j+1)}\right)s^{\mu+1}(1-s^2)^\nu ds \\ &= \sum_{j=0}^{\infty}\frac{(-1)^j(\frac{t}{2})^{\mu+2j}}{j!\Gamma(\mu+j+1)}\int_0^1 s^{2\mu+2j+1}(1-s^2)^\nu ds \\ &= \sum_{j=0}^{\infty}\frac{(-1)^j(\frac{t}{2})^{\mu+2j}}{j!\Gamma(\mu+j+1)}\frac{1}{2}\int_0^1 r^{\mu+j}(1-r)^\nu dr \\ &= \sum_{j=0}^{\infty}\frac{(-1)^j(\frac{t}{2})^{\mu+2j}}{j!\Gamma(\mu+j+1)}\frac{1}{2}\frac{\Gamma(\mu+j+1)\Gamma(\nu+1)}{\Gamma(\mu+\nu+j+2)} \\ &= \left(\sum_{j=0}^{\infty}\frac{(-1)^j(\frac{t}{2})^{\mu+\nu+1+2j}}{j!\Gamma(\mu+\nu+j+2)}\right)\frac{2\Gamma(\nu+1)}{t^{\nu+1}} \\ &= \frac{2\Gamma(\nu+1)}{t^{\nu+1}}J_{\mu+\nu+1}(t). \end{aligned}$$

There is a special application in the text for $\nu=0$ and $\mu=(d-2)/2$, with

$$J_{d/2}(t) = t\int_0^1 J_{(d-2)/2}(ts)s^{d/2}ds. \tag{10.37}$$

10.6.6 Bounds on Bessel Functions

We continue with two properties of Bessel functions from [47]:

$$J_{d/2}^2(z) \;\leq\; \frac{2^{d+2}}{\pi z}\,, \quad z>0 \tag{10.38}$$

$$\lim_{z\to 0} z^{-d} J_{d/2}^2(z) \;=\; \frac{1}{2^d \Gamma^2\,(1+d/2)}. \tag{10.39}$$

The second of these follows easily from the power series expansion, since

$$\begin{aligned}
\lim_{z\to 0} \left(\frac{z}{2}\right)^{-\nu} J_\nu(z) &= \frac{1}{\Gamma(1+\nu)} \\
\lim_{z\to 0} z^{-\nu} J_\nu(z) &= \frac{2^{-\nu}}{\Gamma(1+\nu)} \\
\lim_{z\to 0} \left(z^{-\nu} J_\nu(z)\right)^2 &= \frac{2^{-2\nu}}{\Gamma(1+\nu)^2}.
\end{aligned}$$

Unfortunately, equation (10.38) is much more difficult and must (for now) be left to the cited literature. Similarly, there is a weaker, but more general bound

$$|J_\nu(x)| \leq 1 \tag{10.40}$$

for all $x \in I\!R$ and $\nu \geq 0$ ([1], 9.1.60, p. 362). Both of the above bounds should combine into the general inequality

$$|J_\nu(|x|)| \leq \mathcal{O}(|x|^{-1/2}),\; x\to\infty \tag{10.41}$$

in view of [1], 9.2.1, p. 364. These things will be added later.

10.6.7 Integrals Involving Bessel Functions

From [1] 11.4.16, p. 486 we take the moment equations

$$\int_0^\infty t^\mu J_\nu(t)dt = 2^\mu \frac{\Gamma((\nu+\mu+1)/2)}{\Gamma((\nu-\mu+1)/2)} \tag{10.42}$$

which are valid for $\Re(\nu+\mu) > -1,\; \Re(\mu) < 1/2$. We now use these to derive similar equations for the H_ν functions by

$$\begin{aligned}
\int_0^\infty s^\rho H_\nu(s)ds &= \int_0^\infty (z^2/4)^\rho H_\nu(z^2/4)(z/2)dz \\
&= \int_0^\infty (z^2/4)^\rho (z/2)^{-\nu} J_\nu(z)(z/2)dz \\
&= 2^{\nu-1-2\rho} \int_0^\infty z^{2\rho-\nu+1} J_\nu(z)dz \\
&= \frac{\Gamma(\rho+1)}{\Gamma(\nu-\rho)}
\end{aligned} \tag{10.43}$$

for $\rho > -1$ and $\nu > 2\rho + \frac{1}{2}$.

Another citation from [1] 11.4.41, p. 487 is the **Weber-Schafheitlin integral**

$$\int_0^\infty t^{\mu-\nu+1} J_\mu(at) J_\nu(bt) dt$$

$$= \left\{ \begin{array}{ll} 0 & 0 < b < a \\ \frac{2^{\mu-\nu+1} a^\mu (b^2-a^2)^{\nu-\mu-1}}{b^\nu \Gamma(\nu-\mu)} & 0 < a < b \end{array} \right\} \tag{10.44}$$

for $\Re\nu > \Re\mu > -1$ and $a \neq b > 0$.

10.6.8 Bessel Functions of Third Kind

The Bessel function K_ν of third kind (alias Mcdonald function) is defined as

$$K_\nu(z) := \int_0^\infty e^{z\cosh t} \cosh(\nu t) \tag{10.45}$$

for $z \neq 0,\ |\arg z| < \pi/2$ and all $\nu \in \mathbb{C}$. From this it follows that

$$K_\nu = K_{-\nu} \tag{10.46}$$

holds and that K_ν is positive for real parameters $\nu,\ z$. For the special case $\Re\nu > -1/2$ there is an integral representation

$$K_\nu(z) = \frac{\pi^{1/2}(z/2)^\nu}{\Gamma(\nu+1/2)} \int_1^\infty e^{-zt}(t^2-1)^{\nu-1/2} dt. \tag{10.47}$$

Its asymptotics near zero is

$$K_\nu(z) = \frac{(z/2)^{-\nu}}{\Gamma(\nu)} + o(1) \tag{10.48}$$

for $\nu > 0$ and real, while it behaves like

$$K_\nu(z) = \frac{\sqrt{\pi}}{\sqrt{2z}} e^{-z}(1 + \mathcal{O}(z^{-1})), \tag{10.49}$$

near infinity for $|\nu| \geq 1/2$. The asymptotics of K_0 near zero are like

$$K_0(r) = \frac{1}{e} - \log(r/2) + o(1) \text{ for } r \to 0.$$

Due to [1], 11.4.44, p.488 it is related to the J_ν Bessel functions via the identity

$$\int_0^\infty \frac{t^{\nu+1} J_\nu(at)}{(t^2+z^2)^{\mu+1}} dt = \frac{a^\mu z^{\nu-\mu}}{2^\mu \Gamma(\mu+1)} K_{\nu-\mu}(az) \tag{10.50}$$

for $a, z > 0,\ -1 < \nu < 2\mu + 3/2$. It satisfies the differential equations

$$\begin{aligned} K'_\nu(z) &= K_{\nu-1}(z) - \frac{\nu}{z} K_\nu(z) \\ \frac{d}{dz}(z^\nu K_\nu(z)) &= -z^\nu K_{\nu-1}(z). \end{aligned} \tag{10.51}$$

The second equation, combined with (10.46), proves that the functions $K_\nu(x)x^\nu$ for $x > 0,\ \nu \geq 0$ are nondecreasing for $x > 0$ with exponential decay at infinity. These functions occur as reproducing kernels of Sobolev spaces and are often called **Matérn kernels** or **Sobolev kernels**.

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1970.

[2] R. Ahrem, A. Beckert, and H. Wendland. A meshless spatial coupling scheme for large-scale fluid-structure-interaction problems. *Computer Modeling in Engineering and Sciences*, 12:121–136, 2006.

[3] N. Arad, N. Dyn, and D. Reisfeld. Image warping by radial basis functions: applications to facial expressions. *Graphical Models and Image Processing*, 56:161–172, 1994.

[4] R. Askey. Radial characteristic functions. Technical Report 1262, Univ. of Wisconsin, 1973. MRC Technical Sum. Report.

[5] T. Belytschko, Y. Krongauz, D.J. Organ, M. Fleming, and P. Krysl. Meshless methods: an overview and recent developments. *Computer Methods in Applied Mechanics and Engineering, special issue*, 139:3–47, 1996.

[6] M. Botsch and L. Kobbelt. Real-time shape editing using radial basis functions. In *Computer Graphics Forum*, volume 24, pages 611 – 621, 2005.

[7] S.C. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods, Second edition*. Springer, 2002.

[8] M.D. Buhmann. Convergence of univariate quasi-interpolation using multiquadrics. *IMA J. Numer. Anal.*, 8:365–383, 1988.

[9] M.D. Buhmann. Multivariate cardinal interpolation with radial–basis functions. *Constr. Approx.*, 6:225–255, 1990.

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

[11] J.C. Carr, 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 *Computer Graphics Proceedings*, pages 67–76. Addison Wesley, 2001.

[12] J.C. Carr, W.R. Fright, and R.K. Beatson. Surface interpolation with radial basis functions for medical imaging. *IEEE Transactions on Medical Imaging*, 16:96–107, 1997.

[13] A.H.-D. Cheng and J.J.S.P. Cabral. Direct solution of certain ill-posed boundary value problems by collocation method. In A. Kassab, C.A. Brebbia, E. Divo, and D. Poljak, editors, *Boundary Elements XXVII*, pages 35–44, 2005.

[14] A.H.-D. Cheng and J.J.S.P. Cabral. Direct solution of ill-posed boundary value problems by radial basis function collocation method. *Internat. J. Numer. Methods Engrg.*, 64(1):45–64, 2005.

[15] N. Cristianini and J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, Cambridge, 2000.

[16] Carl de Boor. *A practical guide to splines*. Springer Verlag, New York, 2001. revised version.

[17] L. Demaret, N. Dyn, M.S. Floater, and A. Iske. Adaptive thinning for terrain modelling and image compression. In *Advances in multiresolution for geometric modelling*, Math. Vis., pages 319–338. Springer, Berlin, 2005.

[18] T.A. Driscoll and B. Fornberg. Interpolation in the limit of increasingly flat radial basis functions. *Comput. Math. Appl.*, 43:413–422, 2002.

[19] G.E. Fasshauer. RBF collocation methods and pseudospectral methods. Preprint, 2006.

[20] M.S. Floater and A. Iske. Thinning algorithms for scattered data interpolation. *BIT*, 38:4:705–720, 1998.

[21] B. Fornberg, E. Larsson, and G. Wright. A new class of radial basis functions. *Comput. Math. Appl.*, 451:1209–1222, 2006.

[22] B. Fornberg and G. Wright. Stable computation of multiquadric interpolants for all values of the shape parameter. *Comput. Math. Appl.*, 48(5-6):853–867, 2004.

[23] George Gasper. Positive integrals of Bessel functions. *SIAM J. Math. Anal.*, 6(5):868–881, 1975.

[24] P. Giesl and H. Wendland. Approximating the basin of attraction of time-periodic ODEs by meshless collocation. Preprint Göttingen/München, 2006.

[25] P. Giesl and H. Wendland. Meshless collocation: Error estimates with application to dynamical systems. Preprint Göttingen/München, 2006.

[26] G. Golub and C. van Loan. *Matrix computations*. The Johns Hopkins University Press, 1996. Third edition.

[27] P.C. Hansen. Analysis of discrete ill-posed problems by means of the *l*-curve. *SIAM Review*, 34:561–580, 1992.

[28] P.C. Hansen. Regularization tools: a MATLAB package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms*, 6:1–35, 1994.

[29] P.C. Hansen. The L-curve and its use in the numerical treatment of inverse problems. In P. Johnston, editor, *Computational Inverse Problems in Electrocardiology, Advances in Computational Bioengineering Series*. WIT Press, Southampton, 2000.

[30] P.C. Hansen and D.P. O'Leary. The use of the *l*-curve in the regularization of discrete ill-posed problems. *SIAM Journal of Scientific Computing*, 14:1487–1503, 1993.

[31] Y.C. Hon and T. Wei. A meshless scheme for solving inverse problems of Laplace equation. In *Recent development in theories & numerics*, pages 291–300. World Sci. Publishing, River Edge, NJ, 2003.

[32] Y.C. Hon and T. Wei. The method of fundamental solution for solving multidimensional inverse heat conduction problems. *CMES Comput. Model. Eng. Sci.*, 7(2):119–132, 2005.

[33] B. Jin and Y. Zheng. Boundary knot method for some inverse problems associated with the Helmholtz equation. *Internat. J. Numer. Methods Engrg.*, 62(12):1636–1651, 2005.

[34] E. Larsson and B. Fornberg. A numerical study of some radial basis function based solution methods for elliptic PDEs. *Comput. Math. Appl.*, 46:891–902, 2003.

[35] E. Larsson and B. Fornberg. Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions. *Comput. Math. Appl.*, 49(1):103–130, 2005.

[36] Y.J. Lee, G.J. Yoon, and J. Yoon. Convergence property of increasingly flat radial basis function interpolation to polynomial interpolation. preprint, 2005.

[37] J. Li. A radial basis meshless method for solving inverse boundary value problems. *Comm. Numer. Methods Engrg.*, 20(1):51–61, 2004.

[38] J. Li. Application of radial basis meshless methods to direct and inverse biharmonic boundary value problems. *Comm. Numer. Methods Engrg.*, 21(4):169–182, 2005.

[39] W. R. Madych. An estimate for multivariate interpolation. II. *J. Approx. Theory*, 142(2):116–128, 2006.

[40] W.R. Madych and S.A. Nelson. Bounds on multivariate polynomials and exponential error estimates for multiquadric interpolation. *J. Approx. Theory*, 70:94–114, 1992.

[41] V. Maz'ya and G. Schmidt. On approximate approximations using Gaussian kernels. *IMA Journal of Numerical Analysis*, 16:13–29, 1996.

[42] Herbert Meschkowski. *Hilbertsche Räume mit Kernfunktion*, volume 113 of *Die Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin-Göttingen-Heidelberg, 1962.

[43] A. Meyenburg. Interpolation mit Translaten einer geraden 2π–periodischen stetigen Funktion. Master's thesis, Universität Göttingen, 1996. Staatsexamensarbeit NAM–E–390.

[44] F.J. Narcowich and J.D. Ward. Norm of inverses and condition numbers for matrices associated with scattered data. *Journal of Approximation Theory*, 64:69–94, 1991.

[45] F.J. Narcowich and J.D. Ward. Norms of inverses for matrices associated with scattered data. In P.J. Laurent, A. Le Méhauté, and L.L. Schumaker, editors, *Curves and Surfaces*, pages 341–348. Academic Press, Boston, 1991.

[46] F.J. Narcowich and J.D. Ward. Generalized Hermite interpolation via matrix–valued conditionally positive definite functions. Technical Report 280, Department of Mathematics, Texas A & M University, 1992.

[47] F.J. Narcowich and J.D. Ward. Norm estimates for the inverses of a general class of scattered–data radial–function interpolation matrices. *Journal of Approximation Theory*, 69:84–109, 1992.

[48] F.J. Narcowich, J.D. Ward, and H. Wendland. Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. *Constr. Approx.*, 24:175–186, 2006.

[49] Z. Nashed. Applications of wavelets and kernel methods in inverse problems. In *Integral methods in science and engineering*, pages 189–197. Birkhäuser Boston, Boston, MA, 2006.

[50] J.Y. Noh, D. Fidaleo, and U. Neumann. Animated deformations with radial basis functions. In *ACM Virtual Reality and Software Technology (VRST)*, pages 166–174, 2000.

[51] Y. Ohtake, A. Belyaev, and H.-P. Seidel. A multi-scale approach to 3D scattered data interpolation with compactly supported basis functions. In *SMI '03: Proceedings of the Shape Modeling International 2003*, page 292, Washington, DC, USA, 2003. IEEE Computer Society.

[52] R. Opfer. *Multiscale Kernels*. PhD thesis, Universität Göttingen, 2004. Published by Shaker Verlag.

[53] P. Reuter, I. Tobor, C. Schlick, and S. Dedieu. Point-based modelling and rendering using radial basis functions. In *Proceedings of ACM Graphite 2003*, pages 111–118, 2003.

[54] C. Rieger and B. Zwicknagl. Sampling inequalities for infinitely smooth functions, with applications to interpolation and machine learning. preprint Göttingen/Leipzig, 2006.

[55] S. Saitoh. Applications of reproducing kernels to best approximations, Tikhonov regularizations and inverse problems. In *Advances in analysis*, pages 439–445. World Sci. Publ., Hackensack, NJ, 2005.

[56] R. Schaback. Error estimates and condition numbers for radial basis function interpolation. *Adv. Comput. Math.*, 3:251–264, 1995.

[57] R. Schaback. On the efficiency of interpolation by radial basis functions. In A. LeMéhauté, C. Rabut, and L.L. Schumaker, editors, *Surface Fitting and Multiresolution Methods*, pages 309–318. Vanderbilt University Press, Nashville, TN, 1997.

[58] R. Schaback. Convergence analysis of methods for solving general equations. In A. Kassab, C.A. Brebbia, E. Divo, and D. Poljak, editors, *Boundary Elements XXVII*, pages 17–24. WITPress, Southampton, 2005.

[59] R. Schaback. Convergence of unsymmetric kernel-based meshless collocation methods. Preprint, 2006.

[60] R. Schaback. Limit problems for interpolation by analytic radial basis functions. Preprint, to appear in J. of Computational and Applied Mathematics, 2006.

[61] R. Schaback. Unsymmetric meshless methods for operator equations. Preprint, 2006.

[62] R. Schaback and H. Wendland. Approximation by positive definite kernels. for the IDOMat2001 proceedings (invited lecture), 2002.

[63] R. Schaback and H. Wendland. Kernel techniques: From machine learning to meshless methods. *Acta Numerica*, 15:543–639, 2006.

[64] R. Schaback and Z. Wu. Operators on radial basis functions. To appear in the proceedings of the 1995 Cancun conference.

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

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

[67] Y.J. Tan, X.H. Wang, and B.G. Niu. Solving the inverse problem of determining the boundary of a parabolic equation by radial basis method. *J. Fudan Univ. Nat. Sci.*, 43(3):285–291, 2004.

[68] G. N. Watson. *A treatise on the theory of Bessel functions*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1995. Reprint of the second (1944) edition.

[69] T. Wei, Y.C. Hon, and Y.B. Wang. Reconstruction of numerical derivatives from scattered noisy data. *Inverse Problems*, 21:657 – 672, 2005.

[70] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4:389–396, 1995.

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

[72] H. Wendland and C. Rieger. Approximate interpolation with applications to selecting smoothing parameters. *Numer. Math.*, 101:643–662, 2005.

[73] J.H. Wu and L. Kobbelt. Efficient spectral watermarking of large meshes with orthogonal basis functions. In *The Visual Computers (Pacific Graphics 2005 Proceedings)*, volume 21 (8-10), pages 848–857, 2005.

[74] Z. Wu. Multivariate compactly supported positive definite radial functions. *Advances in Computational Mathematics*, 4:283–292, 1995.

[75] Z.M. Wu. Hermite–Birkhoff interpolation of scattered data by radial basis functions. *Approximation Theory Appl.*, 8/2:1–10, 1992.

[76] J. Yoon. Spectral approximation orders of radial basis function interpolation on the Sobolev space. *SIAM J. Math. Anal.*, 33(4):946–958 (electronic), 2001.

Index