Kernel–Based Meshless Methods

Version:

February 1, 2011

Contents

Preface

This is a text intended for use with my lecture "Approximationsverfahren II" in winter 2010/2011. Though the basic background material is in the book [Wen05] of Holger Wendland, some additional stuff is necessary at certain places. The text is an update of of a 2005 lecture handout. It is under construction at various marked places, and it will evolve during the term.

Readers might consult the books or surveys [Aro50, Mes62, BCR84, Sas94, Sch97b, BS00, Buh03, Wen05, SW06, Fas07] (in chronological order) for additional material. Single papers will be cited where needed, but a few things presented here cannot be found elsewhere.

Göttingen, February 1, 2011

R. Schaback

1 Introduction

This text provides some basic material on kernels. It turns out that kernels arise very naturally in Applied Mathematics, in various places, and for diferent purposes. To give the reader an impression of the kernels that are in our focus, we first list the most important cases without referring to specific properties.

1.1 Radial Kernels

In a sense that can be specified, the **Gaussian**

$$K(x,y) := \exp(-\|x-y\|_2/2) \text{ for all } x,y \in \mathbb{R}^d \text{ or } \mathbb{C}^d \tag{1.1}$$

is the mother of many kernels. We shall use the notation $\mathbb{K}$ in the following, standing for either $\mathbb{R}$ or $\mathbb{C}$.

In general, our notion of kernels will use the following

Definition 1.2. *Let Ω be an arbitrary nonempty set. A function*

$$K \ : \ \Omega \times \Omega \to \mathbb{K}, \quad \textit{i.e. } \mathbb{R} \textit{ or } \mathbb{C}$$

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

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

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

Since the Gaussian (1.1) can be written as a function

$$\phi(r) = K(\|x-y\|_2), \ \phi \ : \ [0,\infty) \to \mathbb{K}$$

of the Euclidean distance $r = \|x-y\|_2$, it is traditionally called a **radial basis function** (**RBF**). There are other prominent kernels of this type, e.g. the **multiquadrics**

$$\phi(r) = (1+r^2)^{\beta/2}.$$

For negative β, they are often called **inverse multiquadrics**.

Other kernels are the **powers**

$$\phi(r) = r^\beta, \ \beta \notin 2\mathbb{Z}.$$

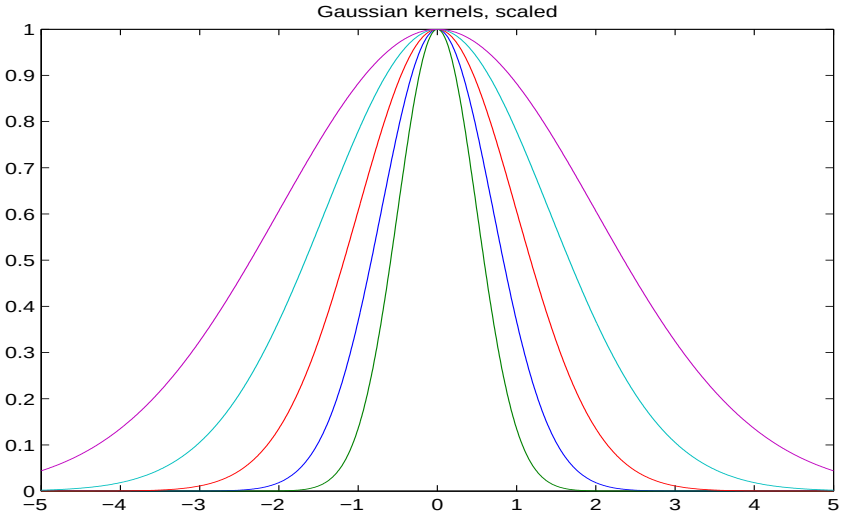

Figure 1: Gaussian kernels

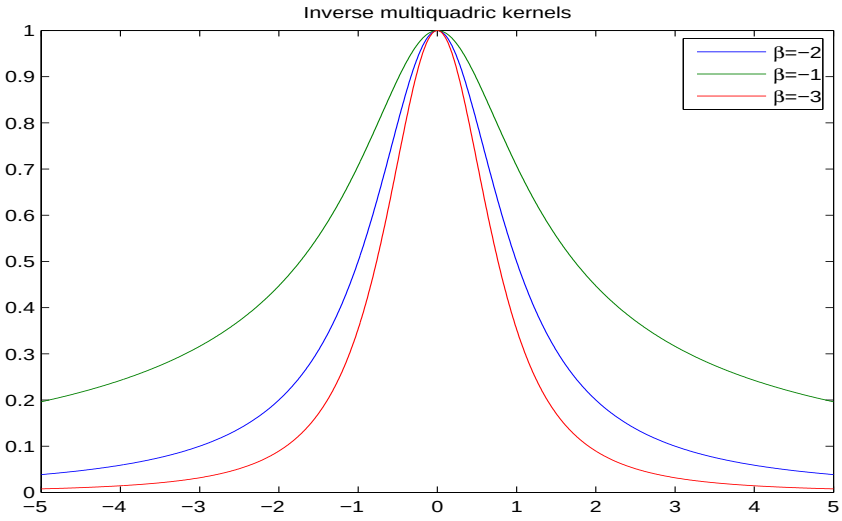

Figure 2: Inverse multiquadrics

The latter are part of an important family called **polyharmonic** for a reason to be explained later, and the other kernels of this family take the form

$$\phi(r) = r^{\beta} \log(r), \ \beta \in 2\mathbb{Z}$$

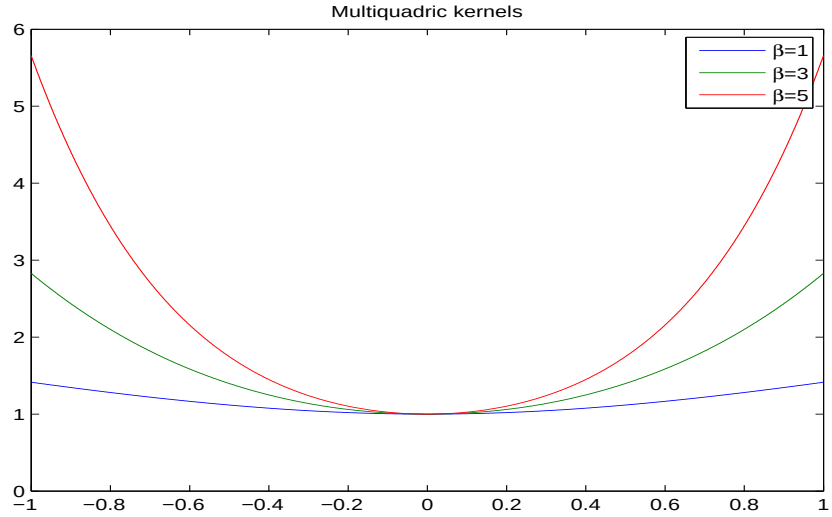

Figure 3: Multiquadrics

with the special case

$$\phi(r) = r^2 \log(r)$$

called the **thin–plate spline** because of its connection to the partial differential equation describing the bending of thin plates.

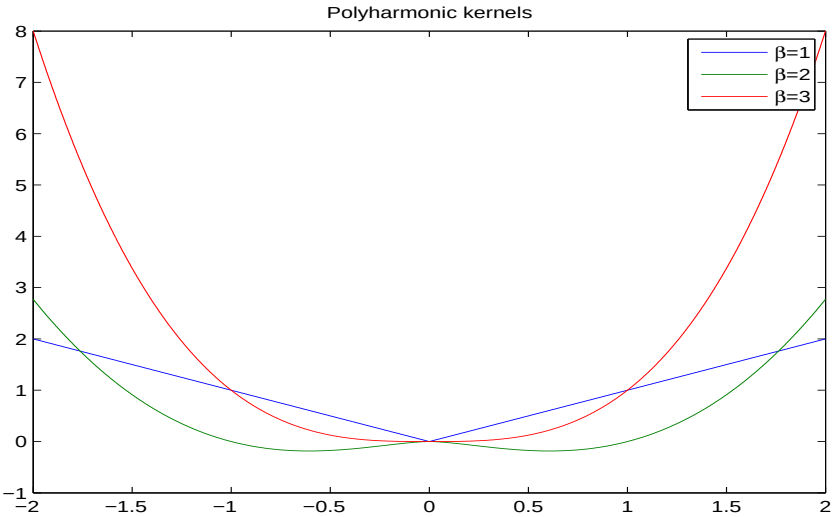

Figure 4: Polyharmonic kernels

The zoo of radial kernels also contains compactly supported kernels like Wendland's [Wen95] kernel

$$\phi_{3,1}(r) = (1-r)_+^4(1+4r)$$

with the **cutoff function**

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

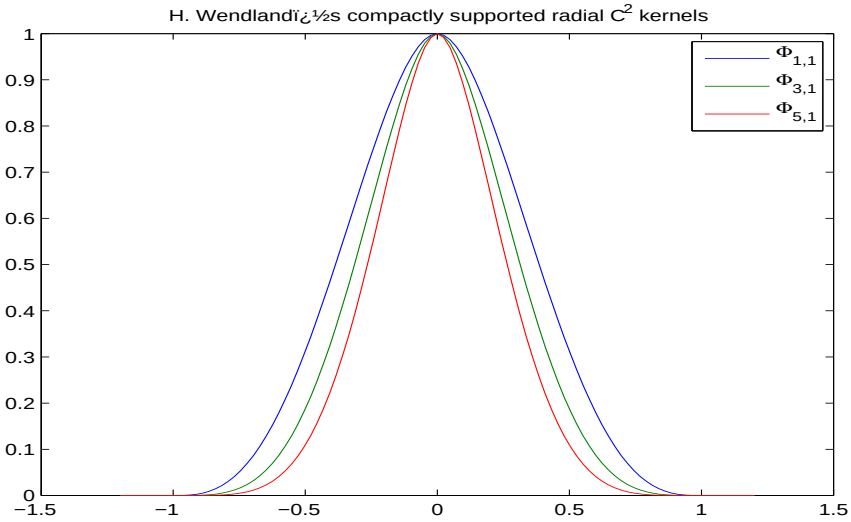

Figure 5: Some C^2 Wendland kernels

For reasons that come up later, a particularly important family of kernels

$$\phi(r) = r^\beta \; K_\beta(r)$$

is related to Matérn or Sobolev. It uses the Bessel functions K_ν of third kind, see the section 12.7 on Special Functions.

1.2 Stationary, Periodic, and Zonal Kernels

To get away from radial kernels, we can focus on **translation–invariant** or **stationary** kernels that are functions

$$K(x,y) = \Phi(x-y)$$

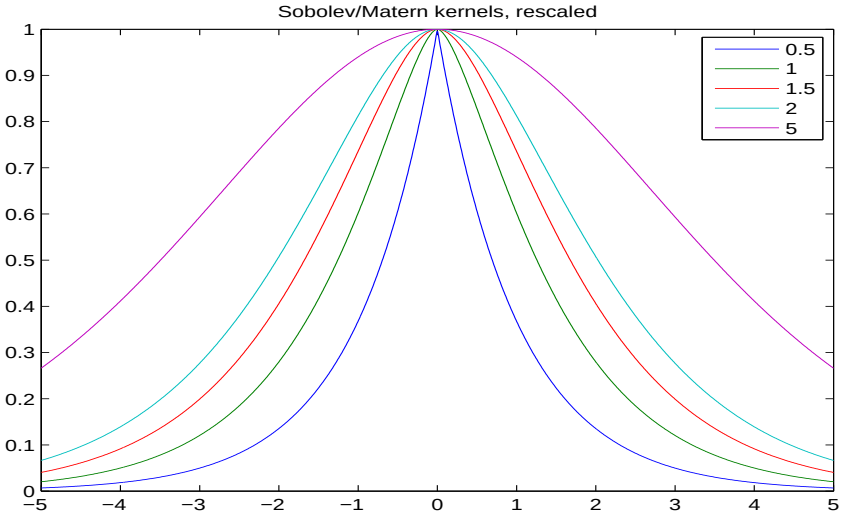

Figure 6: Sobolev/Matérn kernels

of differences, if the domain Ω allows an additive group operation. This, for instance, applies to periodic functions as well, and there we have examples like the **Dirichlet kernel**

$$D(\varphi) := \frac{1}{2} + \sum_{j=1}^{N} \cos(j\,\varphi) = \frac{1}{2} \frac{\sin\left(\left(n + \frac{1}{2}\right)\varphi\right)}{\sin\left(\frac{\varphi}{2}\right)}$$

which is applied to differences $\varphi = \alpha - \beta$ of angles or of 2π–periodic arguments. This kernel plays a dominat role in Fourier series theory, because it allows to write a Fourier partial sum as an integral.
Other non–radial kernels are functions of inner products, like

$$K(x, y) = \exp(x^T\, y) \text{ for all } x, y \in \mathbb{R}^d.$$

Such kernels are particularly important when working on the unit sphere, since then $x^T\, y$ is the cosine of the angle between the two vectors x and y, and thus the kernel can be represented as a function of an angle. Historically, these kernels are called **zonal**. There are many papers on kernel–based methods on the sphere, but no comprehensive book, so far.

At this point, we omit the general case of kernels on (semi–) groups [BCR84] or on Riemannian manifolds [Nar95]. But we remark that kernels can always be restricted to subsets of their domain without losing essential properties. This applies when defining kernels on embedded manifolds, e.g. the sphere.

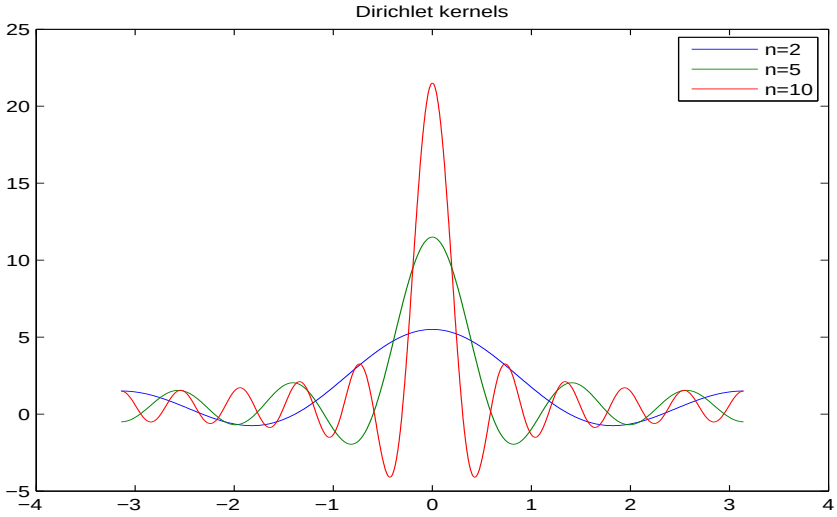

Figure 7: Dirichlet kernels

1.3 Kernels in Machine Learning

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 [CST00, SS02, STC04] are extremely general, but still they take a special form which can be tailored to meet the demands of applications.

This starts from defining the set Ω of objects one wants to learn about. Then 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**

$$K(x, y) := (\Phi(x), \Phi(y))_{\mathcal{F}} \text{ for all } x, y \in \Omega \tag{1.3}$$

which is automatically Hermitian. If the feature space is finite–dimensional or a sequence space, the resulting kernel is an expansion kernel. These will be dealt with in Section 1.6 and Chapter 4.

1.4 Spaces of Trial Functions

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

$$\mathcal{K}_0 := \text{ span}\,\{K(x, \cdot) \ : \ x \in \Omega\}. \tag{1.4}$$

of functions on Ω. In Learning Theory, the function $K(x, \cdot) = (\Phi(x), \Phi(\cdot))_{\mathcal{F}}$ relates each other input object to a fixed object x 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.

1.5 Convolution Kernels

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**

$$K *_\Omega f := \int_\Omega f(x) K(\cdot, x) d\mu(x) \tag{1.5}$$

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. See [BCR84] and the dissertation [Sch09b] for kernels on rotation groups.

Note that discretization of the integral in the convolution transform leads to functions in the space $\mathcal{K}_0$ from (1.4). 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.

1.6 Expansion Kernels

Integral operators (1.5) often have eigenfunction expansions of the form

$$K(x,y) = \sum_{i\in I}^{\infty} \lambda_i \overline{\varphi_i(x)}\varphi_i(y) \tag{1.6}$$

that go under the names of **Hilbert–Schmidt** or **Mercer** or **Karhunen–Loéve**, using a general index set I which usually is countable. We shall use the name **expansion kernels** for these, even if there is no integral operator behind them. Then they just are a series of the above form, with certain functions $\varphi_i \ : \ \Omega \to \mathbb{R}$, $i \in I$, certain positive **weights** λ_i, $i \in I$ and an **index set** I such that the summability conditions

$$K(x,x) := \sum_{i\in I} \lambda_i |\varphi_i(x)|^2 < \infty \tag{1.7}$$

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

$$\ell_{2,I,\lambda} := \{\{\xi_i\}_{i\in I} \ : \ \sum_{i\in I} \lambda_i |\xi_i|^2 < \infty\} \tag{1.8}$$

of sequences with indices in I.

Note further that the summability condition (1.7) 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

$$\begin{aligned} K(x,y) &:= \exp(-(x-y)^2) \\ &= \exp(-x^2)\exp(2xy)\exp(-y^2) \\ &= \exp(-x^2)\left(\sum_{n=0}^{\infty} \frac{2^n}{n!} x^n y^n\right)\exp(-y^2) \\ &= \sum_{n=0}^{\infty} \frac{2^n}{n!} \underbrace{x^n \exp(-x^2)}_{=:\varphi_n(x)} \underbrace{y^n \exp(-y^2)}_{=:\varphi_n(y)} \\ &= \sum_{n=0}^{\infty} \frac{2^n}{n!} \varphi_n(x)\varphi_n(y) \text{ for all } x,y \in \mathbb{R} \end{aligned} \tag{1.9}$$

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

1.7 Kernels from Transforms

A variation of the convolution and the expansion kernels are kernels obtained from transforms, e.g. Fourier series, Fourier transforms, or other instances of harmonic analysis. The basic principle is the representation

$$K(x,y) = \int_T \overline{g(\omega,x)} g(\omega,y) d\mu(\omega) \tag{1.10}$$

where integration or summation takes place with respect to a nonnegative measure μ on a transform domain T. In case of Fourier series,

$$K(\varphi,\psi) := \sum_{n\in\mathbb{Z}} \hat{K}(n) \exp(in(\varphi-\psi))$$

with nonnegative real numbers $\hat{K}(n)$. These are examples of expansion kernels. In case of Fourier transforms in d variables,

$$K(x,y) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{K}(\omega) \exp(i\omega^T(x-y)) d\omega$$

with a nonnegative transform function $\hat{K}$ on $\mathbb{R}^d$. We shall use this extensively in Section 9.3.

Theorem 1.11. *If a kernel K is defined via (1.10), it is Hermitian and positive semidefinite. If for all pairwise distinct points $x_k \in \Omega$ the functions $g(\cdot,x_k)$ are linearly independent on at least a set of positive measure, the kernel is positive definite.*

Proof: The standard quadratic form is

$$\begin{aligned} & \sum_{j,k=1}^N \overline{a_j} a_k K(x_j,x_k) \\ = & \sum_{j,k=1}^N \overline{a_j} a_k \int_T \overline{g(\omega,x_j)} g(\omega,x_k) d\mu(\omega) \\ = & \int_T \left| \sum_{k=1}^N a_k g(\omega,x_k) \right|^2 d\mu(\omega) \geq 0. \end{aligned}$$

If the form vanishes, the linear combination of the $g(\cdot,x_k)$ vanishes on all sets of positive measure. This proves the second assertion. □

1.8 Special Kernels

Another source of kernels are differential equations. Typical examples are Green's functions or fundamental solutions, or singular kernels like the single–layer or double–layer potential. We shall touch these cases here or there. But there also are kernels which are specially tailored for use with PDEs, e.g. **harmonic kernels** [Sch09a, HS10] or **matrix–valued kernels** [NW94, Low05a, Low05b, Fus08a] that allow to generate **divergence–free vector fields** [NWW07, Fus08b]

Finally, kernels often arise as covariance kernels in stochastic models. If for every t in some set Ω we have a random variable X_t with existing second moments, we can define the **covariance kernel**

$$K(s,t) := Cov(X_s, X_t),\ \Omega \times \Omega \to \mathbb{R} \tag{1.12}$$

and analyze its properties. It turns out that the statistical estimation techniques called **Kriging** in Geostatistics are algorithmically identical to interpolation with translates of kernels in Numerical Analysis, and a large part of this text will study these methods in detail. The connection of covariance kernels 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 (1.3) takes large positive values.

At this point, we leave out various other occurrences of kernels in the mathematical literature and in applications (see the survey article [SW06]). 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.

RS: the section on PDE-related kernels needs extension.

2 Kernels from Hilbert Spaces

From here on, we describe a common framework for most of the kernels that we saw in the previous section. We start with noting that all Hilbert spaces lead to " reproducing" kernels, and in the next chapter we shall see that most kernels lead to "native" Hilbert spaces in which they are reproducing.

2.1 Reproducing Kernel Hilbert Spaces

As Numerical Analysts, we want to work with real– or complex–valued functions on domains Ω. For each $x \in \Omega$ and each function f we consider, we

want that the evaluation

$$x \mapsto f(x) \in \mathbb{K}, \text{ i.e. } \mathbb{R} \text{ or } \mathbb{C}$$

is a reasonable operation. It depends on both f and x, and it should be stably computable.

But we shall place more emphasis on f than on x, because the set Ω is completely unstructured in various applications, e.g. in Machine Learning. If we place plenty of structure on the space $\mathcal{H}$ of functions we want to work with, we can assume $\mathcal{H}$ to be a Hilbert space with an inner product $(.,.)_{\mathcal{H}}$. See Section 11 for a basic account of Hilbert space theory.

For later use, we allow ourselves to consider arbitrary Hilbert spaces as well, without being spaces of functions on some set. But this is no generalization, because any Hilbert space $\mathcal{H}$ is a Hilbert space of functions on its own dual $\mathcal{H}^*$ via

$$f(\mu) := \mu(f) \text{ for all } f \in \mathcal{H},\ \mu \in \mathcal{H}^*. \tag{2.1}$$

In particular, the dual space now consists completely of point evaluation functionals, where "points" are functionals themselves. We shall come back to this.

Continuity of point evaluation functionals

$$\delta_x\ :\ f \mapsto f(x),\ x \in \Omega$$

then means that these functionals are in the topological dual $\mathcal{H}^*$ of $\mathcal{H}$ and satisfy

$$|\delta_x(f)| = |f(x)| \leq \|\delta_x\|_{\mathcal{H}^*}\|f\|_{\mathcal{H}} \text{ for all } x \in \Omega,\ f \in \mathcal{H}.$$

Furthermore, the Riesz isometry

$$R\ :\ \mathcal{H}^* \to \mathcal{H},\ \lambda(f) = (f, R(\lambda))_{\mathcal{H}} \text{ for all } f \in \mathcal{H},\ \lambda \in \mathcal{H}^*$$

maps δ_x into a function

$$K(x,\cdot) := R(\delta_x) \in \mathcal{H} \text{ for all } x \in \Omega$$

that is a kernel according to Definition 1.2. Then

$$\delta_y(f) = f(y) = (f, R(\delta_y))_{\mathcal{H}} = (f, K(y,\cdot))_{\mathcal{H}} \text{ for all } f \in \mathcal{H},\ y \in \Omega \tag{2.2}$$

is a **reproduction equation** for values of functions from the inner product. It is clear that any kernel $K(x,\cdot)$ satisfying the reproduction equation must

be the Riesz representer of the point evaluation functional δ_x. Thus the reproducing kernel is unique.

Specializing to $f = K(x, \cdot) \in \mathcal{H}$ we get

$$K(x, y) = (K(x, \cdot), K(y, \cdot))_{\mathcal{H}} = (\delta_y, \delta_x)_{\mathcal{H}^*} \text{ for all } x, y \in \Omega. \tag{2.3}$$

This is one of the various occurrences of **kernels** in Mathematics. We shall see a number of cases later. Note that in the right–hand side the points x and y are interchanged. This is due to the antilinearity of the Riesz map in the complex case, leading to the isometry property in the sense

$$(R(\lambda), R(\mu))_{\mathcal{H}} = (\mu, \lambda)_{\mathcal{H}^*} \text{ for all } \lambda, \mu \in \mathcal{H}^*.$$

See section 11.6 for details on Hilbert spaces and Riesz maps.

Theorem 2.4. *Each Hilbert space $\mathcal{H}$ of real– or complex–valued functions on some set Ω with continuous point evaluation functionals is a* **reproducing kernel Hilbert space** *(***RKHS***) with a unique* **kernel**

$$K \ : \ \Omega \times \Omega \to \mathbb{K}$$

satisfying the reproduction equation (2.2) and the representation (2.3). □

The above kernel K is **Hermitian** in the sense $K(x, y) = \overline{K(y, x)}$. In particular, $K(x, x)$ is always real. Note that $K(x, \cdot)$ is a function in $\mathcal{H}$, but not necessarily $K(\cdot, x)$, unless K is real–valued and thus **symmetric**, i.e. $K(x, y) = K(y, x)$. This is related to the fact that vector spaces over $\mathbb{C}$ need not have the property that they are closed under taking the complex conjugate or taking the real or imaginary parts.

The values of such kernels can always be bounded by

$$|K(x, y)|^2 \leq K(x, x)K(y, y) \text{ for all } x, y \in \Omega \tag{2.5}$$

due to (2.3), but we shall generalize this property later in Theorem 3.7 on page 42.

As a warm–up we state

Theorem 2.6. *All Hilbert spaces $\mathcal{H}$ of functions on some set Ω with a reproducing kernel K coincide with the closure of the linear combinations of functions $K(y, \cdot)$ for all $y \in \Omega$.*

Proof: Assume that some $f \in \mathcal{H}$ is orthogonal to all $K(y,\cdot)$. Then (2.2) proves that f is zero as a function on Ω. □

Theorem 2.7. *If a Hilbert (sub–) space of functions on Ω has a finite orthonormal basis $v_1, \ldots, v_N$ the reproducing kernel is*

$$K_N(x,\cdot) = \sum_{j=1}^{N} \overline{v_j(x)} v_j(\cdot) \text{ for all } x \in \Omega.$$

In case of a subspace, we have

$$K_N(x,x) = \sum_{j=1}^{N} |v_j(x)|^2 \leq K(x,x) \text{ for all } x \in \Omega.$$

Proof: Whatever the (always existing) kernel looks like, it must have a representation

$$\begin{aligned} K_N(x,\cdot) &= \sum_{j=1}^{N} (K_K(x,\cdot), v_j) v_j(\cdot) \\ &= \sum_{j=1}^{N} \overline{v_j(x)} v_j(\cdot) \end{aligned}$$

in the orthonormal basis. We could postpone the second assertion to Theorem 2.19, but here is another proof. Consider

$$\begin{aligned} K_N(x,x) &= (K_N(x,\cdot), K_N(x,\cdot))_{\mathcal{H}} \\ &= (K_N(x,\cdot), K(x,\cdot))_{\mathcal{H}} \end{aligned}$$

yielding

$$K_N(x,x) \leq \sqrt{K_N(x,x)} \sqrt{K(x,x)} \text{ for all } x \in \Omega.$$

□

The second assertion is somewhat surprising for an orthonormal basis, since it means that for increasing N the functions v_N must get small in spite of their normalization being independent of N. But in many cases the Hilbert space norm also includes derivatives, and since these are kept at bay by normalization, it is no miracle that the basis functions, exhibiting sharp spikes, tend to be small in their function values.

Corollary 2.8. *If a Hilbert (sub–) space with continuous point evaluation has a complete orthonormal basis, then Theorem 2.7 also holds for $N = \infty$.*

Proof: Just use a series expansion in the above proof. There are no convergence problems, because the Bessel inequality yields

$$\sum_{j=1}^{\infty} |(K(x,\cdot), v_j)_{\mathcal{H}}|^2 = \sum_{j=1}^{\infty} |v_j(x)|^2 = \|K(x,\cdot)\|_{\mathcal{H}}^2 = K(x,x) < \infty,$$

proving via the Cauchy–Schwarz inequality that the series

$$K(x,y) = \sum_{j=1}^{\infty} \overline{v_j(x)} v_j(y)$$

converges pointwise and absolutely. □

Note that **all** orthonormal bases give the same result. A change of basis will not change the kernel, only the representation will change.

2.2 The Dual Space

For later use, we need some information about the dual space of a reproducing kernel Hilbert space. The reason is that one often knows the space and the kernel, the latter as an explicit formula. But then one wants to know which linear functionals $\lambda : \mathcal{H} \to \mathbb{K}$ are in the dual of the Hilbert space. Here and elsewhere, we use superscript arguments to indicate the action of variables, i.e. λ^x means the action of λ with respect to the variable x.

Theorem 2.9. *The dual space $\mathcal{H}^*$ of a reproducing kernel Hilbert space of functions on some set Ω is the closure of the span of all point evaluation functionals δ_x for $x \in \Omega$. For each pair $\lambda,\ \mu$ of functionals from the dual $\mathcal{H}^*$ of $\mathcal{H}$, one can define $\lambda^x(\overline{\mu^y K(x,y)})$ uniquely via Cauchy sequences to yield*

$$\lambda^x(\overline{\mu^y K(x,y)}) = (\lambda,\mu)_{\mathcal{H}^*} \text{ for all } \lambda,\ \mu \in \mathcal{H}^* \tag{2.10}$$

as a generalization of (2.3). In particular, the Riesz representer of a functional $\lambda \in \mathcal{H}^*$ is $\overline{\lambda^x K(\cdot,x)}$, and this function lies in $\mathcal{H}$.

Proof: Clearly, the first assertion is just the dual form of Theorem 2.6.

The identity (2.10) holds for all linear combinations of point evaluation functionals, and by continuity it carries over to all limits of Cauchy sequences, i.e. to all functionals in the dual. For linear combinations of point evaluation functionals (and thus later for all functionals), the reproduction equation is

$$\lambda(f) = (f, \overline{\lambda^x K(\cdot,x)})_{\mathcal{H}}$$

proving $R(\lambda) = \overline{\lambda^x K(\cdot,x)}$. □

Roughly speaking, the dual space consists of a functionals which

- can be obtained via sequences of linear combinations of point evaluation functionals
- such that their application in the limit is possible for both arguments of K independently.

To make this more precise, we consider functionals that have a meaning outside the Hilbert space in question. Examples are functionals like

$$f \mapsto (\Delta f)(x),\ f \mapsto \int_\Omega f(t)dt.$$

We want to have a sufficient condition for these to lie in $\mathcal{H}^*$.

Theorem 2.11. *Assume that the kernel K of a reproducing kernel Hilbert space $\mathcal{H}$ of functions on some set Ω is explicitly known as a function on $\Omega\times\Omega$, and assume it allows the action of a general functional λ to both arguments, i.e. $\lambda^y\overline{\lambda^x K(y,x)} \in \mathbb{K}$ exists. Furthermore, assume that there is a sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ of linear combinations of point evaluation functionals on points of Ω such that for all $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that for all $n, m \geq N$ we have*

$$\left|\lambda_n^y\overline{\lambda_m^x K(y,x)} - \lambda^y\overline{\lambda^x K(y,x)}\right| \leq \epsilon. \tag{2.12}$$

Finally, assume

$$\lim_{n\to\infty} \lambda_n^y K(x,y) = \lambda^y K(x,y) \text{ for all } x \in \Omega. \tag{2.13}$$

Then λ lies in $\mathcal{H}^*$ and is the limit of the Cauchy sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ in $\mathcal{H}^*$.

Proof: We first want to show that $\{\lambda_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{H}^*$. This follows immediately from

$$\begin{aligned}
\|\lambda_n - \lambda_m\|_{\mathcal{H}^2}^2 &= \|\lambda_n\|_{\mathcal{H}^2}^2 + \|\lambda_m\|_{\mathcal{H}^2}^2 - (\lambda_n,\lambda_m)_{\mathcal{H}^*} - (\lambda_m,\lambda_n)_{\mathcal{H}^*} \\
&= \lambda_n^y\overline{\lambda_n^x K(y,x)} + \lambda_m^y\overline{\lambda_m^x K(y,x)} \\
&\quad -\lambda_n^y\overline{\lambda_m^x K(y,x)} - \lambda_m^y\overline{\lambda_n^x K(y,x)} \\
&\leq 4\epsilon
\end{aligned}$$

for all $n, m \geq N$. Now the sequence must have a limit $\tilde{\lambda} \in \mathcal{H}^*$, and we have to show that $\lambda = \tilde{\lambda}$ as functionals on $\mathcal{H}$. From (2.13) we get

$$\lim_{n\to\infty} \lambda_n^y K(x,y) = \tilde{\lambda}^y K(x,y) = \lambda^y K(x,y) \text{ for all } x \in \Omega.$$

By Theorem 2.6, this extends to all of $\mathcal{H}$. □

2.3 Implications for Ω

We first look at the situation where the Hilbert space $\mathcal{H}$ is invariant under a group $\mathcal{T}$ of transformations $\Omega \mapsto \Omega$ in the sense that

$$\begin{array}{rcl} f(T(\cdot)) & \in & \mathcal{H} \text{ for all } f \in \mathcal{H},\ T \in \mathcal{T} \\ (f,g)_{\mathcal{H}} & = & (f(T(\cdot)), g(T(\cdot)))_{\mathcal{H}} \text{ for all } f,g \in \mathcal{H},\ T \in \mathcal{T}. \end{array}$$

Theorem 2.14. *If $\mathcal{H}$ is invariant in the above sense under transformations, so is the reproducing kernel, i.e.*

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

Proof: Just consider

$$\begin{array}{rcl} f(T(x)) & = & (f, K(T(x),\cdot))_{\mathcal{H}} \\ & = & (f(T(\cdot)), K(T(x),T(\cdot)))_{\mathcal{H}} \end{array}$$

and introduce $g := f(T(\cdot))$ to get

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

to see that the point evaluation functional δ_x is also represented by $K(T(x),T(\cdot))$. □

Theorem 2.14 is behind many simplified kernels. Translation–invariance on $\Omega = \mathbb{R}^d$ is induced by invariance under shifts, while radial kernels arise from invariance under both shifts, rotations, and reflections, i.e. rigid–body motions. Zonal kernels on the sphere arise from rotational invariance.

The mapping $\delta : x \mapsto \delta_x$ takes Ω into $\mathcal{H}^*$. If it is not injective, we have a nontrivial equivalence relation $x \sim y$ on Ω defined by $\delta_x = \delta_y$ or $f(x) = f(y)$ for all $f \in \mathcal{H}$. In view of the Stone–Weierstraß theorem, and to avoid the above effect, there is

Definition 2.15. *A space $\mathcal{H}$ of functions on a set Ω* separates points *of Ω if for every pair $x \neq y$ of different points in Ω there is a function $f \in \mathcal{H}$ with $f(x) \neq f(y)$.*

If $\mathcal{H}$ separates points of Ω, we have injectivity of δ. Otherwise, we might eliminate this by going over to the factor set $\tilde{\Omega} := \Omega/\sim$ instead of Ω.

Under the hypotheses of Theorem 2.4 we can define

$$d(x,y) := \|\delta_x - \delta_y\|_{\mathcal{H}^*} \text{ for all } x,y \in \Omega \tag{2.16}$$

with the explicitly available representation

$$d(x,y)^2 = K(x,x) + K(y,y) - K(x,y) - K(y,x) \text{ for all } x,y \in \Omega. \quad (2.17)$$

This is a nonnegative symmetric function that satisfies the triangle inequality, but it can vanish for $x \neq y$ in case that $f(x) = f(y)$ for all $f \in \mathcal{H}$ or $x \sim y$. If this is assumed, the above function is a true metric on the otherwise unstructured set Ω, and the kernel could be redefined as a function $K^*(\delta_x, \delta_y) = (\delta_x, \delta_y)_{\mathcal{H}^*} = K(x,y)$ on $\Omega^* \times \Omega^*$ using the set

$$\Omega^* := \{\delta_x \ : \ x \in \Omega\}$$

that is the image of the embedding from Ω into $\mathcal{H}^*$ via $x \mapsto \delta_x$. The kernel in this form is nothing than the restriction of the inner product on $\mathcal{H}^* \times \mathcal{H}^*$. In the circumstances of (2.1), the kernel is the inner product itself, and the metric on $\Omega = \mathcal{H}^*$ is the usual metric induced by the norm.

Aiming at continuity of functions, we can look at

$$\begin{aligned} |f(x) - f(y)|^2 &= |(\delta_x - \delta_y)(f)|^2 \\ &\leq \|f\|_{\mathcal{H}}^2 \, \|\delta_x - \delta_y\|_{\mathcal{H}^*}^2 \\ &= \|f\|_{\mathcal{H}}^2 \, d(x,y)^2 \\ &= \|f\|_{\mathcal{H}}^2 \, (K(x,x) - K(x,y) - K(y,x) + K(y,y)) \\ &= \|f\|_{\mathcal{H}}^2 \, (K(x,x) - 2 \text{ Re } (K(x,y)) + K(y,y)) \end{aligned}$$

to see

Theorem 2.18. *Functions from $\mathcal{H}$ are always Lipschitz continuous with respect to d, and if K is continuous on $\Omega \times \Omega$, the functions in $\mathcal{H}$ are also continuous on Ω.* □

But note that talking about continuity on Ω needs a topology there, and we have not fixed a topology other than the one induced by d.

2.4 Kernels for Subspaces

Let $\mathcal{H}_0$ be a closed subspace of a Hilbert space $\mathcal{H}$ of functions on Ω. As such, it is a Hilbert space itself, and it has its own reproducing kernel K_0. With the projector $\Pi_0 \ : \ \mathcal{H} \to \mathcal{H}_0$ we have

Theorem 2.19. *The subspace kernel is*

$$K_0(x, \cdot) = \Pi_0(K_0(x, \cdot)) \text{ for all } x \in \Omega,$$

and the reproducing kernel for the orthogonal complement $\mathcal{H}_0^\perp$ is $K - K_0$.

Proof: We can decompose the identity on $\mathcal{H}$ into the orthogonal projectors

$$I = \Pi_0 + (I - \Pi_0) =: \Pi_0 + \Pi_0^\perp$$

and apply this to the reproduction equation (2.2). Then

$$\begin{aligned} f(y) &= (\Pi_0 f)(y) + (\Pi_0^\perp f)(y) \\ &= (f, K(y,\cdot))_{\mathcal{H}} \\ &= (\Pi_0 f + \Pi_0^\perp f, \Pi_0 K(y,\cdot) + \Pi_0^\perp K(y,\cdot))_{\mathcal{H}} \\ &= (\Pi_0 f, \Pi_0 K(y,\cdot))_{\mathcal{H}} + (\Pi_0^\perp f, \Pi_0^\perp K(y,\cdot))_{\mathcal{H}}. \end{aligned}$$

Specializing this to $f \in \mathcal{H}_0$ or to $f \in \mathcal{H}_0^\perp$ proves the assertions. □

In particular,

$$f(y) = (\Pi_0 f)(y) + (f - \Pi_0 f, \Pi_0^\perp K(y,\cdot))_{\mathcal{H}}$$

can in some instances be the Taylor formula, as we shall see.

The upshot here is that orthogonal space decompositions correspond to additive kernel decompositions using the appropriate projectors.

2.5 Subspaces from Point Sets

For what follows, we fix a nonempty subset $X \subseteq \Omega$ of Ω and consider the subspace

$$\mathcal{H}_X := \text{ clos span } \{K(x,\cdot) \,:\, x \in X\} \subseteq \mathcal{H} \tag{2.20}$$

of $\mathcal{H}$. It is closed by definition, and we have

Theorem 2.21.

$$\mathcal{H}_X^\perp = \{f \,:\, f \in \mathcal{H},\ f(X) = \{0\}\}.$$

Proof: If $f(X) = \{0\}$, then $f \in \mathcal{H}_X^\perp$ by the reproduction formula (2.2), and conversely. □

From standard results in Hilbert spaces, we know that there is a projector Π_X from $\mathcal{H}$ to $\mathcal{H}_X$. With the shorthand notation $f_X := \Pi_X(f)$ we get

Theorem 2.22. *Each function $f \in \mathcal{H}$ has an orthogonal decomposition*

$$f = f_X + f_X^\perp$$

with $f_X \in \mathcal{H}_X$ and $f_X^\perp \in \mathcal{H}_X^\perp$. This means that each function $f \in \mathcal{H}$ has an interpolant $f_X \in \mathcal{H}_X$ recovering the values of f on X. Furthermore,

$$\|f - f_X\|_{\mathcal{H}} = \inf_{g \in \mathcal{H}_X} \|f - g\|_{\mathcal{H}} \tag{2.23}$$

and

$$\|f_X\| = \inf_{\substack{f(x) = g(x) \\ \forall x \in \Omega \\ g \in \mathcal{H}}} \|g\|_{\mathcal{H}} = \inf_{v \in \mathcal{H}_X^\perp} \|f - v\|_{\mathcal{F}} \tag{2.24}$$

due to orthogonality of the decomposition. □

Note that Theorem 2.22 covers transfinite interpolation and provides two optimality principles known from spline theory. Because of their importance, we restate them as

Corollary 2.25. *The interpolant $f_X \in \mathcal{H}_X$ to a function f on X is at the same time the best approximation to f from all functions in $\mathcal{H}_X$.*

Proof: This is (2.23). □

Corollary 2.26. *The interpolant $f_X \in \mathcal{H}_X$ to a function f on X is minimizes the norm under all interpolants from the full space $\mathcal{H}$.*

Proof: This is (2.24). □

A third optimality property will follow in Section 2.8.

Defining $f_\emptyset = 0$, $f_\emptyset^\perp = f$ and $\mathcal{H}_\emptyset = \{0\}$ with $\mathcal{H}_\emptyset^\perp = \mathcal{H}$ for completeness, we can note a few simple observations:

Corollary 2.27. *For all sets $X \subseteq Y \subseteq \Omega$ and all $f \in \mathcal{H}$ we have*

$$\|f_X\|_{\mathcal{H}} \leq \|f_Y\|_{\mathcal{H}} \leq \|f\|_{\mathcal{H}}$$

and

$$\|f\|_{\mathcal{H}} \geq \|f - f_X\|_{\mathcal{H}} \geq \|f - f_Y\|_{\mathcal{H}}. \quad \square$$

2.6 Power Function

We now specialize to $f = K(x, \cdot)$ for a fixed $x \in \Omega$.

Definition 2.28. *The function*

$$P_X(x) := \|K(x, \cdot) - K(x, \cdot)_X\|_{\mathcal{H}},\ x \in \Omega$$

is called the **Power Function** *with respect to the set X and the kernel K.*

This is nonstandard in the transfinite case, so far. The error functional

$$\epsilon_{x,X} \ : \ f \mapsto f(x) - (\Pi_X(f))(x)$$

is well–defined and in $\mathcal{H}^*$. Thus another definition of the Power Function could be

$$P_X(x) := \|\epsilon_{x,X}\|_{\mathcal{H}^*} \text{ for all } x \in \Omega.$$

Theorem 2.29. *These definitions are equivalent. The Power Function has the properties*

$$\begin{array}{rcll} P_X(x) & = & 0 & \textit{for all } x \in X \\ P_\emptyset(x)^2 & = & K(x,x) & \textit{for all } x \in \Omega \\ P_\Omega(x) & = & 0 & \textit{for all } x \in \Omega \\ 0 = P_\Omega(x) \le P_Y(x) & \le & P_X(x) \le P_\emptyset(x) & \textit{for all } x \in \Omega,\ X \subseteq Y \subseteq \Omega, \\ P_X(x) & = & \inf\limits_{g \in \mathcal{H}_X} \|K(x,\cdot) - g\|_{\mathcal{H}} & \textit{for all } x \in \Omega \\ P_X(x) & = & \sup\limits_{\substack{f \in \mathcal{H} \\ \|f\|_{\mathcal{H}} \le 1 \\ f(X) = \{0\}}} f(x) & \textit{for all } x \in \Omega \end{array}$$

but the most important is the error bound

$$|f(x) - f_X(x)| = |f_X^\perp(x)| \le P_X(x)\|f_X^\perp\|_{\mathcal{H}} = P_X(x)\|f - f_X\|_{\mathcal{H}} \le P_X(x)\|f\|_{\mathcal{H}} \tag{2.30}$$

for all $x \in \Omega,\ f \in \mathcal{H}$.

Proof: For the equivalence, we have to prove that the Riesz representer of $\delta_x \circ \Pi_X$ is $K(x,\cdot)_X$. This follows from

$$\begin{array}{rcl} (f, R(\delta_x \circ \Pi_X))_{\mathcal{H}} & = & (\delta_x \circ \Pi_X)(f) \\ & = & f_X(x) \\ & = & (f_X, K(x,\cdot))_{\mathcal{H}} \\ & = & (f_X, K(x,\cdot)_X + K(x,\cdot)_X^\perp)_{\mathcal{H}} \\ & = & (f_X, K(x,\cdot)_X)_{\mathcal{H}} \\ & = & (f - f_X^\perp, K(x,\cdot)_X)_{\mathcal{H}} \\ & = & (f, K(x,\cdot)_X)_{\mathcal{H}} \end{array}$$

using the various orthogonalities.

The first five listed properties are easy consequences of Definition 2.28 and the previous results. The error bound follows as well from what we already know, but we can also use the error representation

$$\begin{array}{rcl} f(x) - f_X(x) & = & f_X^\perp(x) \\ & = & (f_X^\perp, K(x,\cdot))_{\mathcal{H}} \\ & = & (f_X^\perp, K(x,\cdot) - K(x,\cdot)_X)_{\mathcal{H}} \end{array}$$

because $f_X^\perp$ is orthogonal to $K(x,\cdot)_X$.

We are left with the sixth, the "dual" representation of the Power Function. From the first inequality of the error bound, we see that

$$P_X(x) \geq \sup_{\|f_X^\perp\|_{\mathcal{H}} \leq 1} f_X^\perp(x)$$

and equality must hold if we insert the representer of $\epsilon_{x,X}$. □

The fifth property of the previous theorem has another equivalent formulation. Consider the subspace

$$\mathcal{H}_X^* := \text{ clos span } \{\delta_x \ : \ x \in X\}$$

of the dual space $\mathcal{H}^*$. Then the property has the dual form of the fifth property of Theorem 2.29, i.e.

$$P_X(x) = \inf_{\lambda \in \mathcal{H}_X^*} \|\delta_x - \lambda\|_{\mathcal{H}^*} \tag{2.31}$$

for all $x \in \Omega$, and it indicates how well the point evaluation functional δ_x can be approximated by arbitrary linear combinations of the point functionals for points of X.

2.7 Interpolants on Finite Sets

We now consider finite sets $X = \{x_1, \ldots, x_N\} \subseteq \Omega$. For each $f \in \mathcal{H}$ we can write f_X as a linear combination

$$f_X = \sum_{j=1}^N \alpha_j K(x_j, \cdot) \tag{2.32}$$

with coefficients $\alpha_j \in \mathbb{R}$ or $\mathbb{C}$, but note that the coefficients might not be unique, since we do not assume that the $K(x_j,\cdot)$ are linearly independent. Since we know that f_X must interpolate f on X, we have

Theorem 2.33. *For each $f \in \mathcal{H}$, the linear system*

$$\sum_{j=1}^N \alpha_j K(x_j, x_k) = f_k, \ 1 \leq k \leq N \tag{2.34}$$

with the Hermitian **kernel matrix**

$$A = (K(x_j, x_k))_{1 \leq k,j \leq N} \tag{2.35}$$

is solvable. □

This is somewhat surprising, since the kernel matrix can be singular under the assumptions we made fo far.

Theorem 2.36. *In reproducing kernel Hilbert spaces, the kernel matrix for a finite set X is positive semidefinite. It is positive definite if the point evaluation functionals δ_x for $x \in X$ or, equivalently, the functions $K(x, \cdot)$ for $x \in X$ are linearly independent.*

Proof: This follows because any kernel matrix on a finite set $X = \{x_1, \ldots, x_N\}$ is a Gramian matrix for the functionals δ_{x_j} or the functions $K(x_j, \cdot)$ due to the representation equation (2.3). □

The upshot of Theorem 2.33 is that the right–hand side is always in the span of the columns of the matrix. Users must bear in mind that the system can be unsolvable for general right–hand sides.

Note that in the definition (2.35) of the kernel matrix, the row index k runs over the second argument of $K(x_j, x_k)$ to turn (2.32) into the interpolation system (2.34). For later use, we introduce

Definition 2.37. *A kernel on $\Omega \times \Omega$ is Hermitian and* **positive semidefinite**, *if all kernel matrices for all finite point sets of Ω are Hermitian and positive semidefinite.*

Then Theorem 2.36 is

Theorem 2.38. *All reproducing kernels of Hilbert spaces are Hermitian and positive semidefinite.* □

We mention this explicitly here, because we shall prove the converse in the next chapter:

Theorem 2.39. *Every Hermitian and positive definite kernel has a "native" Hilbert space in which it is reproducing.*

This means that there is a one–to–one relation between Hilbert spaces and Hermitian positive semidefinite kernels.

Now we specialize Theorem 2.33 to $f = K(x, \cdot)$. Then we know that

$$K(x, x_k) = \sum_{j=1}^{N} \overline{u_j(x)} K(x_j, x_k), \ 1 \le k \le N \tag{2.40}$$

has a solution $\overline{u_j(x)}$ as a function on Ω. Note that this also follows when calculating the optimal solution of (2.31). Furthermore, we now know that

$$K(x,\cdot)_X(z) = \sum_{j=1}^{N} \overline{u_j(x)} K(x_j, z) \tag{2.41}$$

holds for all $x, z \in \Omega$. Note that the functions u_j need not be contained in $\mathcal{H}_X$ at this point.

Theorem 2.42. *They are in $\mathcal{H}_X$ and a Lagrange basis, i.e.*

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

if the kernel matrix is nonsingular. In general, we still have

$$f_X = \sum_{j=1}^{N} u_j(\cdot) f(x_j). \tag{2.43}$$

Proof: The first assertion follows from (2.40), and it is now clear why we used the complex conjugates there. To prove the second assertion, we start from (2.32) and proceed via

$$\begin{aligned} f_X &= \sum_{k=1}^{N} \alpha_k K(x_k,\cdot) \\ &= \sum_{k=1}^{N} \alpha_k \sum_{j=1}^{N} u_j(\cdot) \overline{K(x_j,x_k)} \\ &= \sum_{j=1}^{N} u_j(\cdot) \sum_{k=1}^{N} \alpha_k K(x_k,x_j) \\ &= \sum_{j=1}^{N} u_j(\cdot) f(x_j). \square \end{aligned}$$

Going back to (2.41), we get

Theorem 2.44. *The Power Function has the explicit representation*

$$\begin{aligned} P_X^2(x) &= K(x,x) - \sum_{j=1}^{N} u_j(x) K(x,x_j) - \sum_{j=1}^{N} \overline{u_j(x)} K(x_j,x) \\ &\quad + \sum_{j=1}^{N} \sum_{k=1}^{N} \overline{u_j(x)} u_k(x) K(x_j,x_k) \\ &= K(x,x) - K(x,\cdot)_X(x). \end{aligned}$$

Proof: The Power Function is the norm of $K(x,\cdot) - K(x.\cdot)_X$. From (2.41) we then get the first assertion by direct calculation. Inserting (2.40), the third and fourth term cancel, and the second is $K(x,\cdot)_X(x)$ by (2.41). □

Going into the second part of Theorem 2.42 with $f = K(x,\cdot)$, we get

$$\begin{aligned} K(x,\cdot)_X(z) &= \sum_{j=1}^{N} u_j(z) K(x,x_j) \\ &= \sum_{j=1}^{N} \overline{u_j(x)} K(x_j,z) \end{aligned}$$

and see that this quantity is real in case $x = z$.

If the kernel matrix is singular, the point evaluation functionals at the points of x are not linearly independent. But then one can select a maximal linearly independent subset of those functionals and restrict oneself to the subset Y of X consisting of the evaluation points of the selected functionals. Function values of **all** functions of $\mathcal{F}$ on the discarded points are completely determined by the values on Y by an explicit linear dependence which is the same for all functions, and the same applies to the interpolant on Y. Thus it suffices to pose the interpolation problem on Y and ignore the other points. The interpolant on Y will automatically interpolate all functions from $\mathcal{H}$ on X as well. And then one can use the Lagrange basis for the points on Y. Note that this argument fails if the data are not from a function in the Hilbert space.

We shall make this point selection process more precise in Section 2.10 on page 27 and combine it with a numerical construction of an orthonormal basis.

2.8 Best Linear Estimation

From the above discussion, we know that f_X with the representation (2.43) is the interpolant to f on the set X of data locations. We also have all the background material that allows us to conclude that (2.43) at some point x is the best linear predictor for $f(x)$ in a way that we now describe. In particular, this is important if the kernel comes from a covariance (1.12).

Consider completely arbitrary estimation formulas

$$(x,f) \mapsto \sum_{j=1}^{N} v_j(x) f(x_j)$$

where no assumptions are made on the x–dependent scalar coefficients $v_j(x)$. These are linear in f, and for x fixed, their error functional is

$$f \mapsto f(x) - \sum_{j=1}^{N} v_j(x) f(x_j) = \left(\delta_x - \sum_{j=1}^{N} v_j(x) \delta_{x_j} \right) (f).$$

To let the estimation be optimal for all f, we should choose the $v_j(x)$ to minimize

$$V_{X,v}(x) := \left\| \delta_x - \sum_{j=1}^{N} v_j(x) \delta_{x_j} \right\|_{\mathcal{H}^*}.$$

But from (2.31) we know the solution: it is what we have already with our functions u_j, and the optimal error is described by the Power Function.

Theorem 2.45. *In the above sense, kernel–based interpolation yields the best linear predictor of unknown function values $f(x)$ from known function values $f(x_j)$ at points x_j, $1 \leq j \leq N$.* □

Let us specialize to the stochastic setting of (1.12) for a moment, with real–valued random variables X_t with mean zero and bounded variance. The above numerical estimation technique is then called **Kriging**, and $V_{X,v}^2$ is the variance of the prediction error, which is minimized if we proceed like in the previous section.

To supply the necessary details, we denote the error of the general linear predictor at x by

$$\epsilon_{x,X,v} := X_x - \sum_{j=1}^{N} v_j(x) X_{x_j}.$$

It has zero mean, and variance

$$\begin{aligned}
\mathbb{E}\left(\epsilon_{x,X,v}^2\right) &= Cov(X_x, X_x) - 2\sum_{j=1}^{N} v_j(x) Cov(X_x, X_{x_j}) \\
&\quad + \sum_{j=1}^{N} \sum_{k=1}^{N} v_j(x) v_k(x) Cov(X_{x_j}, X_{x_k}) \\
&= K(x,x) - 2\sum_{j=1}^{N} v_j(x) K(x, x_j) \\
&\quad + \sum_{j=1}^{N} \sum_{k=1}^{N} v_j(x) v_k(x) K(x_j, x_k) \\
&= V_{X,v}^2.
\end{aligned}$$

Note that there is no apparent Hilbert space here, but we shall see later that there is one behind the scene.

2.9 Power Function and Stability

It is a general observation that kernel matrices are often badly conditioned, and the condition is particularly poor in cases where the Power Function is small, i.e. when the interpolation error is small. This is kind of an **Uncertainty Principle**:

> It is impossible to make the Power Function and the condition of the kernel matrix small at the same time.

We shall analyze this effect now, following [Sch95].

The basic trick is to express the Power Function via the kernel matrix. In addition to the point set $X = \{x_1, \ldots, x_N\}$, we denote another point x by $x_0 := x$ and define $u_0(\cdot) := -1$. Then we define the extended kernel matrix $\tilde{A}$ with entries $K(x_j, x_k),\ 0 \le j, k \le N$ and the vector

$$u := (u_0(x), u_1(x), \ldots, u_N(x))^T = (-1, u_1(x), \ldots, u_N(x))^T$$

to consider the quadratic form

$$\begin{aligned}
\overline{u}^T \tilde{A}\, u &= \sum_{j=0}^{N}\sum_{k=0}^{N} \overline{u_j(x)} u_k(x) K(x_j, x_k) \\
&= K(x,x) - \sum_{j=1}^{N} u_j(x) K(x, x_j) - \sum_{j=1}^{N} \overline{u_j(x)} K(x_j, x) \\
&\quad + \sum_{j=1}^{N}\sum_{k=1}^{N} \overline{u_j(x)} u_k(x) K(x_j, x_k) \\
&= P_X^2(x)
\end{aligned}$$

using Theorem 2.44. The matrix $\tilde{A}$ is Hermitian and positive semidefinite. Thus it has $N+1$ nonnegative real eigenvalues

$$\lambda_0 \ge \lambda_1 \ge \ldots \ge \lambda_N \ge 0$$

and we get the inequality

$$P_X^2(x) \ge \lambda_{N+1} \left(1 + \sum_{j=1}^{N} |u_j(x)|^2 \right) \ge \lambda_{N+1}$$

from the usual bound

$$\lambda_{N+1} \|u\|_2^2 \le \overline{u}^T \tilde{A}\, u \le \lambda_0 \|u\|_2^2.$$

We can eliminate the special rôle of the point x:

Theorem 2.46. *The kernel matrix for N points $x_1, \ldots, x_N$ forming a set X has a smallest eigenvalue λ bounded above by*

$$\lambda \leq \min_{1 \leq j \leq N} P_{X \setminus \{x_j\}}(x_j). \quad \square$$

This implies that in settings where the Power Function still is small after one point is left out, the kernel matrix must be ill–conditioned.

But note that the kernel matrix is directly connected to the span of translates $K(x_j, \cdot)$, and therefore the ill–conditioning may be a consequence of a bad choice of the basis. This is indeed the case, since one can prove under additional assumptions that the functions u_j are uniformly bounded [DMS10]. We shall come back to stability questions later.

2.10 Newton Bases

The above discussion shows that one should avoid "nearly duplicate" points, or those whose omission does not let the Power Function or the Kriging variance increase too much. This can be cast into an adaptive algorithm [DMSW05, MS09] that we describe now. It constructs an orthonormal basis in a subspace of $\mathcal{H}_X$, which, for certain reasons, can be called a **Newton** basis.

We assume that we are given a fairly large and unordered set X of N points to work with. We start with $X_0 := \emptyset$ and have $P_\emptyset^2(x) = K(x, x)$ due to Theorem 2.29. We evaluate and store the N values $P_\emptyset^2(x) = K(x, x)$ for the x from X.

We introduce an integer $k \geq 0$ and assume that we have already chosen an ordered subset $X_k := \{x_1, \ldots, x_k\}$ of X with linearly independent point functionals $\delta_{x_1}, \ldots, \delta_{x_k}$. Furthermore, we assume that we have an orthonormal basis $v_1, \ldots, v_k$ of the space $\mathcal{H}_{X_k}$ with $v_{j+1} \in \mathcal{H}_{X_{j+1}} \cap \mathcal{H}_{X_j}^\perp$, $0 \leq j \leq k-1$. We assume that we have the values of these functions on X in storage, together with the values of $P_{X_k}(x)^2$ for all $x \in X$. So far, this uses $\mathcal{O}(N(k+1))$ storage.

To perform the next step, we can stop if $X \setminus X_k$ is empty. Then we check the values of P_{X_k} on X. If they are all zero or smaller than a chosen tolerance, we stop. Otherwise we pick some $x_{k+1} \in X$ with $P_{X_k}(x_{k+1}) > 0$. For a special "greedy" strategy we could also choose

$$x_{k+1} := \text{ arg max } \{P_{X_k}(x) \ : \ x \in X \setminus X_k\}. \tag{2.47}$$

If $P_{X_k}(x) = 0$ for all $x \in X$, we stop since there is nothing to gain by continuing.

Now we have $P_{X_k}(x_{k+1}) > 0$ and form $X_{k+1} = X_k \cup \{x_{k+1}\}$. If $\delta_{x_{k+1}}$ were linearly dependent on the functionals $\delta_{x_1}, \ldots, \delta_{x_k}$, we would have $P_{X_k}(x) = P_{X_{k+1}}(x)$ for all $x \in \Omega$ due to Theorem 2.29 and $\mathcal{H}_{X_k} = \mathcal{H}_{X_{k+1}}$, but then $P_{X_k}(x_{k+1}) = P_{X_{k+1}}(x_{k+1}) = 0$ is a contradiction.

Then we go for $v_{k+1} \in \mathcal{H}_{X_{k+1}} \cap \mathcal{H}_{X_k}^{\perp}$ with norm one. The standard way to do this is to ignore normalization first, an to make a function

$$w_{k+1} := K(x_{k+1}, \cdot) - \sum_{j=1}^{k} \alpha_j v_j$$

orthogonal to all $v_1, \ldots, v_k$. This means

$$\begin{aligned} (v_i, K(x_{k+1}, \cdot))_{\mathcal{H}} &= v_i(x_{k+1}) \\ . &= \sum_{j=1}^{k} \alpha_j (v_i, v_j)_{\mathcal{H}} \\ &= \alpha_i, \ 1 \le i \le k \end{aligned} \tag{2.48}$$

as simple evaluations of the functions v_i. We already have these values, but we need the additional n values $K(x_{k+1}, x)$ for $x \in X$ to calculate w_{k+1} on X by $\mathcal{O}(Nk)$ operations. Note that $v_{k+1} \in \mathcal{H}_{X_k}^{\perp}$ implies that $v_k(x_j) = 0, \ 1 \le j \le k$, but we do not let these conditions enter into the calculation.

The norm of w_{k+1} is also easy to calculate via the orthonormal decomposition

$$K(x_{k+1}, \cdot) - w_{k+1} = \sum_{j=1}^{k} \alpha_j v_j$$

leading to

$$\begin{aligned} \sum_{j=1}^{k} |\alpha_j|^2 &= \|K(x_{k+1}, \cdot) - w_{k+1}\|_{\mathcal{H}}^2 \\ &= K(x_{k+1}, x_{k+1}) - (K(x_{k+1}, \cdot), w_{k+1})_{\mathcal{H}} \\ &\quad -(w_{k+1}, K(x_{k+1}, \cdot))_{\mathcal{H}} + \|w_{k+1}\|_{\mathcal{H}}^2 \\ &= K(x_{k+1}, x_{k+1}) - 2 \text{ Re } (w_{k+1}(x_{k+1})) + \|w_{k+1}\|_{\mathcal{H}}^2. \end{aligned}$$

We can now define

$$v_{k+1}(x) := \frac{w_{k+1}(x)}{\|w_{k+1}\|}$$

and calculate its values on X. Finally, we need the Power Function $P^2_{X_{k+1}}$ on X. To this end, we use Theorem 2.44 in the form

$$P^2_{X_{k+1}}(x) = K(x,x) - K(x,\cdot)_{X_{k+1}}(x)$$

and take advantage of our orthonormal basis:

$$\begin{aligned} K(x,\cdot)_{X_{k+1}}(z) &= \sum_{j=1}^{k+1} (K(x,\cdot),v_j)_{\mathcal{H}} v_j(z) \\ &= \sum_{j=1}^{k+1} \overline{v_j(x)} v_j(z), \\ K(x,\cdot)_{X_{k+1}}(x) &= \sum_{j=1}^{k+1} |v_j(x)|^2 \end{aligned}$$

to arrive at the surprisingly simple recursion

$$\begin{aligned} P^2_{X_{k+1}}(x) &= K(x,x) - K(x,\cdot)_{X_{k+1}}(x) \\ &= K(x,x) - \sum_{j=1}^{k+1} |v_j(x)|^2 \\ &= P^2_{X_k}(x) - |v_{k+1}(x)|^2. \end{aligned} \tag{2.49}$$

We see that we are constructing increasing ordered sets of points where the associated point evaluation functionals are linearly independent, and thus all kernel matrices here are positive definite. Furthermore, we have a sequence of orthonormal functions $v_1, v_2, \ldots$ with the property

$$v_k(x_j) = 0, \ 1 \le j < k \tag{2.50}$$

like the basis

$$v_k(x) = \prod_{j=1}^{k-1} (x - x_j)$$

for the univariate interpolating polynomial in Newton form. This is why we call the v_j a **Newton basis**. We summarize:

Theorem 2.51. *The above adaptive algorithm selects for $0 \le k \le N$ an ordered subsequence of points $x_1, \ldots, x_k$ of an N–point set X such that the point evaluation functionals $\delta_{x_1}, \ldots, \delta_{x_k}$ are linearly independent. In addition, an orthonormal basis $v_1, \ldots, v_k$ of $\mathcal{H}_{X_k}$ is constructed with the Newton property (2.50). The overall storage is $\mathcal{O}(Nk)$, while computational operations are $\mathcal{O}(Nk^2)$. The original $N \times N$ kernel matrix is never formed or stored. The algorithm produces and monitors monotonically decreasing power functions with (2.49). It should be stopped when these are small on X.* □

So far, the algorithm only produces the values of the basis on X. In practice, one can often let N and X be as large as needed for plotting and function evaluation, stopping the method at reasonably small values of k. This implies that additional evaluations are not necessary at all. But if evaluation at some x is necessary, it can be done at $\mathcal{O}(k^2)$ cost as follows. We start with

$$v_1(x) = \frac{K(x_1, x)}{\sqrt{K(x_1, x_1)}}$$

and work our way up to $v_k(x)$ using

$$v_j(x) = \frac{w_j(x)}{\|w_j\|_{\mathcal{H}}}$$

and

$$w_j(x) = K(x_j, x) - \sum_{i=1}^{j-1} v_i(x_j) v_i(x)$$

using (2.48).

Theorem 2.52. *The Newton basis functions have the additional property*

$$\sum_{j=1}^{k} |v_j(x)|^2 \;\; \leq \;\; K(x, x) \text{ for all } x \in \Omega.$$

and for the "greedy" variation also

$$|v_j(x)| \;\; \leq \;\; |v_j(x_{j+1})| = P_{X_j}(x_{j+1}) \text{ for all } x \in X.$$

Proof: The first property follows from Theorem 2.7 because we constructed an orthonormal basis of $\mathcal{H}_{X_k}$. For the second, the "greedy" selection of x_{j+1} implies

$$|v_j(x)|^2 \leq P_{X_j}(x)^2 \leq P_{X_j}(x_{j+1})^2 = |v_j(x_{j+1})|^2$$

using that $P_{X_{j+1}}(x_{j+1}) = 0$ due to Theorem 2.29. □

The second property guarantees that the Newton basis has no higher maxima than the controlled one at x_{j+1}.

By orthonormality, we can write the interpolants $f_{X_k} =: f_k$ on X_k in the form

$$f_k = \sum_{j=1}^{k} (f, v_j) v_j, \tag{2.53}$$

and if we do so, we need the coefficients

$$\lambda_j(f) := (f, v_j),\ 1 \le j \le k.$$

This means that the v_j are the Riesz representers of the functionals λ_j. Consequently,

Theorem 2.54. *The functionals λ_j for the Newton basis are orthonormal, and their Riesz representers are the Newton basis functions. In particular,*

$$\sum_{j=1}^{k} |\lambda_j(f)|^2 \le \|f_k\|_{\mathcal{H}}^2 \le \|f\|_{\mathcal{H}}^2.$$

Looking at the numerical evaluation of (2.53) at some point x, given that we have both the $v_j(x)$ and the $\lambda_j(f)$, we get

$$|f_k(x)|^2 \le \left(\sum_{j=1}^{k} |\lambda_j(f)|^2\right) \left(\sum_{j=1}^{k} |v_j(x)|^2\right) \le \|f_k\|_{\mathcal{H}}^2 K(x,x).$$

The outer part is not surprising, but the message here is that both inner factors stay bounded. This is in sharp contrast to (2.32), where in most applications the coefficients α_j are large in absolute value, leading to severe cancellation when forming the sum.

2.11 Kernel Recursions and Expansions

We need not always assume the special choice (2.47) for the next point. To what we do now, it suffices to guarantee $P_{X_k}(x_{k+1}) > 0$ throughout. In view of the recursion (2.49) for the Power Function, we define kernels

$$K_k(x,y) := \sum_{j=1}^{k} \overline{v_j(x)} v_j(y) \text{ for all } x, y \in \Omega \tag{2.55}$$

and

$$K_k^{\perp}(x,y) := K(x,y) - K_k(x,y) \text{ for all } x, y \in \Omega$$

with the recursion

$$K_{k+1}(x,y) = K_k(x,y) + \overline{v_{k+1}(x)} v_{k+1}(y). \tag{2.56}$$

Theorem 2.57. *The kernel K_k is reproducing on $\mathcal{H}_{X_k}$, while $K_k^{\perp}$ is reproducing on $\mathcal{H}_{X_k}^{\perp}$. As reproducing kernels in Hilbert spaces, they are Hermitian and positive semidefinite by Theorem 2.38. Furthermore,*

$$P_{X_k}^2(x) = K_k^{\perp}(x,x) \text{ for all } x \in \Omega.$$

Proof: Each function in $\mathcal{H}_{X_k}$ is of the form (2.53), and this means

$$\begin{aligned} f_k(x) &= \sum_{j=1}^{k} (f_k, v_j) v_j(x) \\ &= (f_k, \sum_{j=1}^{k} \overline{v_j(x)} v_j)_{\mathcal{H}} \\ &= (f_k, K_k(x, \cdot))_{\mathcal{H}} \\ &= (f, K_k(x, \cdot))_{\mathcal{H}} \end{aligned}$$

where the last equality holds for all functions f that f_k interpolates on X_k.

Each function in $\mathcal{H}_{X_k}^{\perp}$ is of the form $g = f - f_k$, and then

$$\begin{aligned} g(x) &= f(x) - f_k(x) \\ &= (f, K(x, \cdot))_{\mathcal{H}} - (f, K_k(x, \cdot))_{\mathcal{H}} \\ &= (f, K_k^{\perp}(x, \cdot))_{\mathcal{H}} \\ &= (g + f_k, K_k^{\perp}(x, \cdot))_{\mathcal{H}} \\ &= (g, K_k^{\perp}(x, \cdot))_{\mathcal{H}}. \end{aligned}$$

Finally, (2.49) implies the rest. □

The adaptive matrix–free algorithm of the previous section is nothing else than a pivoted Cholesky decomposition:

Theorem 2.58. *If stopped after k steps, the algorithm for the Newton basis has produced a* **Cholesky decomposition** $A_k = \overline{L}\, L^T$ *of the kernel matrix A_k for $\mathcal{H}_{X_k}$. The matrix L has the entries $v_j(x_i)$, $1 \le i, j \le k$.*

Proof: From (2.55) we get that the kernel matrix entries are

$$\begin{aligned} K(x_i, x_j) &= K_k(x_i, x_j) + 0 \\ &= \sum_{j=1}^{k} \overline{v_j(x_i)} v_j(x_j) \text{ for all } x, y \in \Omega. \quad \square \end{aligned}$$

If we go from X_k one step further, we have the starting step of the algorithm again, but now acting on K_k. This means

$$v_{k+1}(x) = \frac{K_k(x_{k+1}, x)}{\sqrt{K_k(x_{k+1}, x_{k+1})}}$$

and we get the recursions

$$\begin{aligned} K_{k+1}(x, y) &= K_k(x, y) + \frac{K_k(x, x_{k+1}) K_k(x_{k+1}, y)}{K_k(x_{k+1}, x_{k+1})} \\ K_{k+1}^{\perp}(x, y) &= K_k^{\perp}(x, y) - \frac{K_k(x, x_{k+1}) K_k(x_{k+1}, y)}{K_k(x_{k+1}, x_{k+1})} \end{aligned}$$

from (2.56) which does not contain the Newton basis anymore.

But the main point of this discussion is that one can pass to the limit $k \to \infty$ if there is an infinite set X and if the calculation does not break down prematurely, then leading to a finite–dimensional subspace $\mathcal{H}_X$ of $\mathcal{H}$. In fact, the kernels are pointwise absolutely summable via (2.5) and

$$K_k(x,x) = K(x,x) - P_{X_k}^2(x) \leq K(x,x)$$

for all k.

Definition 2.59. *A subset $X \subseteq \Omega$ is* **unisolvent** *for a space of functions $\mathcal{F}$ on a set Ω, if a function $f \in \mathcal{F}$ vanishing on X must be zero on all of Ω.*

Note that we did not use Hilbert space structure here. But in our standard Hilbert space context, we can apply sections 2.5 and 2.6 to get

Theorem 2.60. *If X is an unisolvent set for a Hilbert space $\mathcal{H}$ of functions on Ω with continuous point evaluation, then $\mathcal{H}_X = \mathcal{H}$ and $\mathcal{H}_X^\perp = \{0\}$. Furthermore, $P_X = 0$.* □

At this point, one can ask whether "dense" point sets X are unisolvent, but we have only the distance (2.16) of Section 2.3 at our disposal.

Theorem 2.61. *If X is dense in the distance (2.16), i.e. if for all points $y \in \Omega$ and all $\epsilon > 0$ there is a point $x \in X$ with $d(x,y) < \epsilon$, then X is unisolvent for $\mathcal{H}$.*

Proof: From (2.31) we know that

$$P_X(y) \leq \inf_{x \in X} \|\delta_y - \delta_x\|_{\mathcal{H}^*} \text{ for all } y \in \Omega,$$

and this implies $P_X(y) = 0$ under the assumptions of the theorem. □

2.12 General Interpolants

We now depart from point evaluation functionals. We fix a subset Λ of the dual $\mathcal{H}^*$ that generalizes the set X we had before, and want to consider interpolation using the data $\lambda(f)$ for all $\lambda \in \Lambda$. This replaces point evaluation functionals by general functionals, and goes back to [Wu92].

In standard special cases, these functionals can contain derivatives, e.g.

$$\lambda(f) = \frac{\partial f}{\partial t_j}\,|_z$$

for the j–th partial derivative of a d–variate function f at a point z, or

$$\lambda(f) = \int_T f(t)v(t)dt$$

for a local integral over a subdomain T against a weight or "test" function v. Interpolation of **general** functionals $\lambda_1, \ldots, \lambda_N$ usually is a mess, because one wants to use interpolants from a span of functions $u_1, \ldots, u_N$ and has no chance to make sure that the matrix with entries $\lambda_k(u_j)$ is nonsingular. Even for univariate polynomials, the fully general **Hermite–Birkhoff interpolation problem** has no apparent and simple solution. In this situation, one resorts to **Hermite interpolation** requiring all necessary lower derivatives, too. In multivariate applications, things are even more complicated, but for kernel–based interpolation there is a solution we describe now. It is the starting point for various meshless methods for solving partial differential equations.

We have already derived it without knowing. In fact, we can deal with this seemingly more general situation by temporarily dropping the kernel K completely, using the kernel

$$K^*(\lambda, \mu) := (\lambda, \mu)_{\mathcal{H}^*} \text{ for all } \lambda, \mu \in \mathcal{H}^*$$

instead of K with

$$K(x, y) = (\delta_y, \delta_x)_{\mathcal{H}^*} = K^*(\delta_y, \delta_x)_{\mathcal{H}^*}.$$

This means that we simply redo the previous paragraphs using $\Omega = \mathcal{H}^*$ and replacing points x and x_j by functionals λ and λ_j, while K^* replaces K with swapped arguments. At the same time, this allows us to work in Hilbert spaces where users cannot rely on point evaluation and have to resort to weak methods. This applies to Hilbert spaces like $W_2^m(\Omega)$ for domains $\Omega \subset \mathbb{R}^d$ with $m \leq d/2$. The most important of such cases arises for $d = 2$ and $m = 1$. But we can also deal with fairly general Hilbert spaces $\mathcal{H}$ that are not necessarily a space of functions on some specific set $\Omega \neq \mathcal{H}^*$. Readers should note that in the sense of (2.1) on page 11 the functionals occur in a twofold way, namely as arguments of functions in $\mathcal{H}$ and as functionals on elements of $\mathcal{H}$.

Given a subset Λ of the dual $\mathcal{H}^*$, we define

$$\mathcal{H}_\Lambda := \text{ clos span } \{R(\lambda) \ : \ \lambda \in \Lambda\}.$$

To see the connection to (2.20), we consider

$$\begin{aligned}(R(\lambda))(\mu) &= \mu(R(\lambda)) \\ &= (R(\lambda), R(\mu))_{\mathcal{H}} \\ &= (\mu, \lambda)_{\mathcal{H}^*} \\ &= K^*(\mu, \lambda) \text{ for all } \lambda \in \mathcal{H}^*,\ \mu \in \mathcal{H}^*\end{aligned}$$

such that $R(\lambda)(\cdot) = K^*(\cdot, \lambda)$.

If the original kernel K is still there, readers can be trapped by assuming that $R(\delta_x)(y) = K(x, y)$ generalizes to $R(\lambda)(y) = \lambda(K(\cdot, y))$, but the action of functionals to the first argument of the kernel is undefined if we are in the truly complex case. Instead, the property $K(x, y) = R(\delta_x)(y)$ generalizes to

$$\begin{aligned}R(\lambda)(x) &= \overline{(R(\lambda), K(x, \cdot))_{\mathcal{H}}} \\ &= \overline{(K(x, \cdot), R(\lambda))_{\mathcal{H}}} \\ &= \overline{\lambda^y K(x, y)}\end{aligned}$$

where λ^y denotes action of λ with respect to y. If we let another functional μ act with respect to x on this, we get

$$\mu(R(\lambda)) = \mu^x(\overline{\lambda^y K(x, y)}) = (\mu, \lambda)_{\mathcal{H}^*}.$$

Thus entries of generalized kernel matrices are

$$(\lambda_j, \lambda_k)_{\mathcal{H}^*} = \lambda_j^x(\overline{\lambda_k^y K(x, y)}) \tag{2.62}$$

if the kernel K is still present.

Dropping K again, we generalize Theorem 2.21 to

Theorem 2.63.

$$\mathcal{H}_\Lambda^\perp = \{f \in \mathcal{H} \ : \ \lambda(f) = 0 \ \textit{for all} \ \lambda \in \Lambda\}.$$

A new proof is not necessary, but we can translate the original proof. Each $f \in \mathcal{H}_\Lambda^\perp$ is characterized by $(f, R(\lambda))_{\mathcal{H}} = 0$ for all $\lambda \in \Lambda$, and this means $\lambda(f) = 0$ for all $\lambda \in \Lambda$. □

Again, we define a projector Π_Λ onto $\mathcal{H}_\Lambda$ and denote $f_\Lambda := \Pi_\Lambda(f)$. Then Theorem 2.22 generalizes to

Theorem 2.64. *Each element $f \in \mathcal{H}$ has an orthogonal decomposition*

$$f = f_\Lambda + f_\Lambda^\perp$$

with $f_\lambda \in \mathcal{H}_\Lambda$ and $f_\Lambda^\perp \in \mathcal{H}_\Lambda^\perp$. Then f_Λ interpolates f in the sense

$$\lambda(f) = \lambda(f_\Lambda) \text{ for all } \lambda \in \Lambda.$$

Furthermore,

$$\|f - f_\Lambda\|_\mathcal{H} = \inf_{g \in \mathcal{H}_\Lambda} \|f - g\|_\mathcal{H}$$

and

$$\|f_\Lambda\|_\mathcal{H} = \inf_{\substack{g \in \mathcal{H} \\ \lambda(f) = \lambda(g) \\ \text{for all } \lambda \in \Lambda}} \|g\|_\mathcal{H}. \quad \square$$

This performs transfinite interpolation by general sets of functionals. Monotonicity like in Corollary 2.27 also prevails, and we have the optimality principles from the corollaries following Theorem 2.22.

To evaluate the error, we cannot use point evaluation functionals. Instead, we take an "evaluation" functional $\mu \in \mathcal{H}^*$ replacing a point x and consider the error functional

$$\mu(f - f_\Lambda) = (\mu - \mu \circ \Pi_\Lambda)(f)$$

The generalization of the Power Function then is

$$P_\Lambda(\mu) := \|\mu - \mu \circ \Pi_\Lambda\|_{\mathcal{H}^*} \text{ for all } \mu \in \mathcal{H}^*.$$

and we leave it to the reader to generalize Theorem 2.29, where the fifth property should be replaced by its dual form (2.31). The numerical construction of interpolants for finite sets $\Lambda = \{\lambda_1, \ldots, \lambda_n\}$ generalizes similarly. Instead of (2.32) we have

$$f_\Lambda = \sum_{j=1}^n \alpha_j R(\lambda_j),$$

and we impose the interpolation conditions to get

$$\lambda_k(f_\Lambda) = \lambda_k(f) = \sum_{j=1}^n \alpha_j \lambda_k(R(\lambda_j)),\ 1 \le k \le n.$$

The kernel matrix is replaced by the Gramian with elements

$$\lambda_k(R(\lambda_j)) = (R(\lambda_j), R(\lambda_k))_{\mathcal{H}} = (\lambda_k, \lambda_j)_{\mathcal{H}^*}.$$

Note that in presence of the original kernel K one has to calculate this using (2.62).

To generalize the system (2.40) we pick $f = R(\mu)$ to obtain

$$\begin{aligned} \lambda_k(R(\mu)_\Lambda) &= \lambda_k(R(\mu)) = (\lambda_k, \mu)_{\mathcal{H}^*} \\ &= \sum_{j=1}^{n} \overline{u_j(\mu)} (\lambda_k, \lambda_j)_{\mathcal{H}^*}, \ 1 \le k \le n \end{aligned}$$

where we cannot say that the u_j are elements of $\mathcal{H}$ unless the functionals in Λ are linearly independent. But in the latter case, we have that they are linear combinations of

$$\overline{\lambda_k(R(\mu))} = (\mu, \lambda_k) = R(\lambda_k)(\mu)$$

as functions of μ, i.e. the u_j are in $\mathcal{H}_\Lambda$. This is in line with Theorem 2.42. In general, the solution of the interpolation problem can be written as

$$f_\Lambda = \sum_{j=1}^{n} \lambda_j(f) u_j$$

in the sense that

$$\mu(f_\Lambda) = \sum_{j=1}^{n} \lambda_j(f) u_j(\mu)$$

for all $\mu \in \mathcal{H}^*$. Also, the connection between the Power Function and stability generalizes to

Theorem 2.65. *The kernel matrix for N functionals $\lambda_1, \ldots, \lambda_N$ forming a set Λ has a smallest eigenvalue λ bounded above by*

$$\lambda \le \min_{1 \le j \le N} P_{\Lambda \setminus \{\lambda_j\}}(\lambda_j). \quad \square$$

Finally, we note that also the construction of the Newton basis generalizes verbatim.

For what follows, we can always stick to point evaluation, going back to $\Omega = \mathcal{H}^*$ and K^* if we want to deal with general functionals.

2.13 Factor Spaces

RS: Still somewhat incomplete, Dec. 2010

In certain important situations connected to the notion of **conditional** positive definiteness (see Section 5), there is no reproducing kernel Hilbert space of functions on Ω at first sight. Instead, there is a space $\mathcal{H}$ of $\mathbb{K}$–valued functions on Ω carrying a semi–inner product with a closed nullspace $\mathcal{P} \subset \mathcal{H}$, i.e.

$$(x, y)_{\mathcal{H}} = 0 \text{ for all } y \in \mathcal{H} \text{ holds iff } x \in \mathcal{P},$$

such that the factor space $\mathcal{H}/\mathcal{P}$ is a Hilbert space under the inner product

$$([f], [g]) := (f, g)_{\mathcal{H}} \text{ for all } f, g \in \mathcal{H},$$

where we adopt the notation $[f]$ for the class $f + \mathcal{P}$ represented by some f. Note that we now have a Hilbert space again, but the elements are equivalence classes modulo some subspace $\mathcal{P}$. With a linear projector Π onto $\mathcal{P}$ we assume that the functionals

$$\mu_x \ : \ f \mapsto f(x) - (\Pi(f))(x) \text{ for all } x \in \Omega, \ f \in \mathcal{H} \tag{2.66}$$

are continuous in the seminorm or on $\mathcal{H}/\mathcal{P}$, i.e.

$$|\mu_x(f)| \leq C_x \|[f]\|_{\mathcal{H}} = C_x \|[f]\| \text{ for all } f \in \mathcal{H}.$$

Then μ_x has a Riesz representer $[K(x, \cdot)]$ in $\mathcal{H}/\mathcal{P}$ with some $K(x, \cdot) \in \mathcal{H}$ which is for each x nonunique up to functions in $\mathcal{P}$. Then

$$\mu_x(f + \mathcal{P}) = f(x) - (\Pi(f))(x) = ([f], [K(x, \cdot)]) = (f, K(x, \cdot))_{\mathcal{H}}$$

holds for all $x \in \Omega, \ f \in \mathcal{H}$. This yields a Taylor–type representation formula

$$f(x) = (\Pi(f))(x) + (f, K(x, \cdot))_{\mathcal{H}} \tag{2.67}$$

for all $f \in \mathcal{H}, \ x \in \Omega$, replacing the reproduction equations we had so far. Since for each $x \in \Omega$ we are free to change $K(x, \cdot)$ by some function in $\mathcal{P}$, we can assume that $\Pi(K(x, \cdot)) = 0$ for all $x \in \Omega$, e.g. by going over to $K(x, y) - \Pi(K(x, \cdot))(y)$. Then the reproduction formula leads to

$$K(y, x) = (K(y, \cdot), K(x, \cdot))_{\mathcal{H}} \text{ for all } x, y \in \Omega. \tag{2.68}$$

This yields a positive semidefinite Hermitian kernel, but note that it depends on the chosen projector Π.

But we do not want to work with equivalence classes. If we want to recover functions from their values at points of a set $X = \{x_1, \ldots, x_N\} \subset \Omega$, we cannot use the data directly, because point evaluation functionals δ_x are not well defined. We have to use functionals μ_{x_j} instead. Thus we should start from a class $[f]$ and consider the problem

$$\arg\min \left\{ \|[s]\| \ : \ s \in \mathcal{H},\ \mu_{x_j}(s) = \mu_{x_j}(f),\ 1 \le j \le N \right\}.$$

This clearly has a minimizer in the factor space, and we know it is a linear combination of the Riesz representers of the μ_{x_j} which are the classes $[K(x_j, \cdot)]$. Thus the function

$$s_0(x) = \sum_{j=1}^{N} \overline{\alpha_j} K(x_j, x)$$

represents the solution class. We also have

$$\mu_{x_k}(s_0) = \mu_{x_k}(f) = 0 + \sum_{j=1}^{N} \overline{\alpha_j} \mu_{x_k}(K(x_j, \cdot)) = \sum_{j=1}^{N} \overline{\alpha_j} K(x_j, x_k) + 0,\ 1 \le k \le N.$$

This system is clearly solvable, but we do not have interpolation of f on all data. We have only cared for data defined by $[f]$, but we also have the same interpolation properties so far, if we change s_0 by some function $p \in \mathcal{P}$. Considering $s = s_0 + p$ with some $p \in \mathcal{P}$, we get

$$\begin{aligned} f(x_k) &= \mu_{x_k}(f) + \Pi(f)(x_k) \\ &= \mu_{x_k}(s_0) + \Pi(f)(x_k) \\ &= \mu_{x_k}(s) + \Pi(f)(x_k) \\ &= s(x_k) - \Pi(s)(x_k) + \Pi(f)(x_k),\ 1 \le k \le N. \end{aligned}$$

and see that we should change s_0 into $s + p$ in such a way that $\Pi(s) = \Pi(s_0 + p) = \Pi(p) = \Pi(f)$. Thus the function $s_0 + \Pi(f)$ solves the full interpolation problem. We summarize:

Theorem 2.69. *Let $\mathcal{H}$ be a space of functions on Ω which carries a semi–inner product with a closed nullspace $\mathcal{P}$ such that the factor space $\mathcal{H}/\mathcal{P}$ is a Hilbert space. Assume further that the functionals (2.66) are continuous in the seminorm, and denote a fixed projector onto $\mathcal{P}$ by Π. Then one can define a Hermitian kernel K on $\Omega \times \Omega$ such that the reproduction equation (2.67) and the standard identity (2.68) hold together with $\Pi(K(x, \cdot)) = 0$ for all $x \in \Omega$. Furthermore, interpolation of data of functions of $\mathcal{H}$ is always possible and has certain optimality properties in the factor space.* □

Note that we have used a good deal of freedom to define a kernel that suited our needs. When going backwards in Section 5, starting from a given conditionally positive definite kernel and procceding towards a Hilbert space, we will not be free to change the kernel.

3 Hilbert Spaces from Kernels

We now go back the the abstract definition 1.2 on page 1 of kernels on general sets Ω. We shall construct a Hilbert space in which the kernel is reproducing. This will then allow us to apply everything we did in the previous chapter.

3.1 Positive Definiteness

If we have no hypotheses to start with, we cannot expect to be able to develop a reasonable theory for kernels. The basic assumption we shall need is positive semidefiniteness, as already defined in Definition 2.37. We give another motivation for it here.

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

$$s(x) := \sum_{j=1}^{N} \overline{a_j} K(x_j, x), \ x \in \Omega, \ a_k \in \mathbb{K} \tag{3.1}$$

of "translates" of the kernel, acting as **trial functions** like we did in Section 1.4 already. This is a very convenient technique to generate functions on an otherwise unstructured set Ω. It will be clear later why we take the complex conjugate of the coefficients in (3.1).

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

$$A := (K(x_j, x_k))_{1 \le j,k \le N} \tag{3.2}$$

and pose the interpolation problem

$$\begin{array}{rcll} y_k & = & s(x_k), & 1 \le k \le N \\ y_k & = & \displaystyle\sum_{j=1}^{N} \overline{a_j} K(x_j, x_k), & 1 \le k \le N. \end{array} \tag{3.3}$$

for s from (3.1). In matrix notation, this is an $N \times N$ linear system

$$A\, a = y.$$

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 3.4. *(see also Definition 2.37)*
A Hermitian kernel K on Ω with values in $\mathbb{K}$ 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** *(3.2) is positive (semi–) definite.*

This means that the Hermitian quadratic form

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

has nonnegative real values. In the positive definite case, it additionally is zero only if the vector a is zero.

Theorem 3.5. *Expansion kernels of the form (1.6) are positive semidefinite. Also, kernels arising from feature maps via (1.3) 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, one can write the expansion via a suitable feature map with values in a weighted sequence space. To give an explicit proof which is typical for much more general cases, the quadratic form corresponding to the kernel matrix can be written as

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

for all vectors $a \in \mathbb{K}^N$. □

Note that this applies to the univariate **Gaussian** via (1.9).

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 (1.6) is Mercer's

Theorem 3.6. *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 (1.6), 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.

3.2 General Rules

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

Theorem 3.7. *Let K be a positive semidefinite kernel on Ω. Then*

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

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 3.4. The second property restates that our kernels will always be Hermitian. Since determinants of positive semidefinite Hermitian quadratic forms must be nonnegative, the third property follows if we take a set $X = \{x, y\}$. The final property follows from the third, using the standard inequality

$$2ab \leq a^2 + b^2$$

for nonnegative real numbers a, b. 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

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

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

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

□

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

Corollary 3.8. *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 real coefficients $c_1, \ldots, c_m$. The kernel K will be hermitian, and positive semidefinite due to

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

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 (1.3). Thus positive semidefiniteness of such kernels is no miracle.

3.3 Inner Product

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

$$H := \text{ span } \{K(x, \cdot) \ : \ x \in \Omega\} \tag{3.9}$$

of all finite linear combinations of generalized translates of the kernel. Similarly, we define the linear space

$$L := \text{ span } \{\delta_x \ : \ x \in \Omega, \ \delta_x \ : \ H \to \mathbb{K}\} \tag{3.10}$$

of all finite linear combinations of point evaluation functionals acting on functions in H. Note that we restrict the action of the functionals to functions in H.

Now all elements from L or H take the form

$$\lambda_{a,X} := \sum_{j=1}^{N} a_j \delta_{x_j}, \ f_{a,X}(x) := \overline{\lambda_{a,X}^y K(x,y)} = \sum_{j=1}^{N} \overline{a_j} K(x_j, x) \tag{3.11}$$

with $a \in \mathbb{K}^N$ while $X = \{x_1, \ldots, x_N\} \subset \Omega$, but different N and all point sets X are allowed. We introduced complex conjugates in the second form, because we want to end up with $f_{a,X} = R(\lambda_{a,X})$ for the antilinear Riesz map R. Note that $f_{a,X}(\cdot) = 0$ or $\lambda_{a,X}(\cdot) = 0$ do not imply $a = 0$, forcing us to be careful.

On L we can define a sesquilinear form

$$\begin{aligned} (\lambda_{a,X}, \lambda_{b,Y})_L &:= \sum_{j=1}^{M} \sum_{k=1}^{N} a_j \overline{b_k} K(y_k, x_j) \\ &= \lambda_{a,X}^x \overline{\lambda_{b,Y}^y K(x,y)} \\ &= \lambda_{a,X}(f_{b,Y}). \end{aligned} \tag{3.12}$$

It is well–defined, because the second form in (3.12) is obtained by an action of the functionals, thus it is independent of their representation. Furthermore, we have a positive semidefinite form due to the positive semidefiniteness of all kernel matrices.

Then we have

$$\begin{aligned} |\lambda_{a,X}(f_{b,Y})| &= |(\lambda_{a,X}, \lambda_{b,Y})_L| \\ &\leq \|\lambda_{a,X}\|_L \|\lambda_{b,Y}\|_L \end{aligned} \tag{3.13}$$

where we may still have just a seminorm, not a norm.

Strangely enough, the sesquilinear form is even positive definite:

Theorem 3.14. *If K is a positive semidefinite Hermitian kernel on Ω, the sesquilinear form $(.,.)_L$ of (3.12) is positive definite on the space L of (3.10) as a space of functionals defined on functions on Ω. Thus L is a pre–Hilbert or Euclidean space of functions on Ω.*

Proof: Assume that

$$(\lambda_{a,X}, \lambda_{a,X})_L = \sum_{j,k=1}^{N} a_j \overline{a_k} K(x_j, x_k) = \lambda_{a,X}^x \overline{\lambda_{a,X}^y K(x,y)} = \lambda_{a,X}(f_{a,X}) = 0$$

for $a \in \mathbb{K}^N$ and $X = \{x_1, \ldots, x_N\} \subset \Omega$. Then by (3.13) we have $\lambda_{a,X} = 0$ as a functional on H. Here it pays off to have the functionals in L restricted to functions in H. Note that we do not need or get $a = 0$. □

Theorem 3.15. *The mapping*

$$R \,:\, \lambda_{a,X} \mapsto f_{a,X} := \overline{\lambda_{a,X}^y K(\cdot, y)}$$

is antilinear and bijective from L onto H. Thus

$$(f_{a,X}, f_{b,Y})_H := (\lambda_{b,Y}, \lambda_{a,X})_L = (R(\lambda_{a,X}), R(\lambda_{b,Y}))_H$$

is an inner product on H, and R acts as the Riesz map.

Proof: If some $f_{b,Y} = R(\lambda_{b,Y}) \in H$ vanishes, (3.12) implies that $\lambda_{b,Y}$ is orthogonal to all of L, thus zero due to Theorem 3.14. The Riesz property is already in (3.12) since

$$\begin{aligned} (\lambda_{a,X}, \lambda_{b,Y})_L &= (R(\lambda_{b,Y}), R(\lambda_{a,X}))_H \\ &= (f_{b,Y}, R(\lambda_{a,X}))_H \\ &= \lambda_{a,X}(f_{b,Y}). \end{aligned}$$

□

When specializing (3.12) to $\lambda_{1,x}$ for a point $x \in \Omega$, we get the reproduction equation

$$\begin{aligned}(\lambda_{1,x}, \lambda_{b,Y})_L &= f_{b,Y}(x) \\ &= (R(\lambda_{b,Y}), R(\lambda_{1,x}))_H \\ &= (f_{b,Y}, R(\lambda_{1,x}))_H \\ &= (f_{b,Y}, K(x,\cdot))_H\end{aligned} \tag{3.16}$$

in H. Finally, (2.3) follows if we set $\lambda_{b,Y} = \lambda_{1,y}$ above.

3.4 Native Space

We now know that H is an inner–product or semi–Hilbert space of functions on Ω under the inner product $(.,.)_H$, provided that K is a positive semidefinite Hermitian kernel on Ω. Furthermore, we also have L as its dual, and we have the Riesz map R.

Then we can invoke a classical argument from Hilbert space theory to go over the closure $\mathcal{H}$ of H under $(.,.)_H$. This is an abstract space defined by equivalence classes of Cauchy sequences in H, but it is a complete space (thus a Hilbert space), and each continuous map from H to a Banach space Y extends uniquely to the closure.

Theorem 3.17. *Each symmetric positive semidefinite kernel K on a set Ω is the reproducing kernel of a Hilbert space called the* **native space** $\mathcal{H} := \mathcal{N}_K$ *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*

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

generalizing (3.16).

Proof: The existence of the native space follows from standard Hilbert space arguments we do not repeat here, see section 11.8. Since (3.16) is an equation with both sides being continuously dependent on $f \in H$, 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 (2.3) and the reproduction equation in T to conclude

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

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

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

is a contradiction. □

To mark the dependence of the native Hilbert space on the kernel K we started with, we prefer the notation $\mathcal{N}_K$ over the notation $\mathcal{H}$ we used in Chapter 2.

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.

4 Expansion Kernels

The previous two chapters showed that we can start from Hilbert spaces to arrive at positive semidefinite Hermitian kernels, but we also can start from the kernels and construct corresponding Hilbert spaces. This chapter illustrates this correspondence for the special case of expansion kernels.

4.1 Kernels from Orthonormal Bases

Let us start from Hilbert spaces first, and work our way towards kernels. We consider the fairly general "separable" case where a Hilbert space $\mathcal{H}$ has a complete orthonormal basis $\{\varphi_n\}_{n\in\mathbb{N}}$. The model situation is the space ℓ_2 of quadratically summable sequences. A practically important case are trigonometric polynomials in the space of square–integrable 2π–periodic functions, or any space of functions spanned by orthogonal polynomials. A third example is the space $\mathcal{B}_h$ of univariate bandlimited functions on $\mathbb{R}$ with spectrum in $[-\pi/h, \pi/h]$ with the orthonormal basis of shifted and scaled sinc functions

$$\mathrm{sinc}(x) = \frac{\sin(\pi\, x)}{\pi\, x},\ x \in \mathbb{R}$$

as used in the **Shannon–Whittaker–Kotelnikov theorem** representing functions $f \in \mathcal{B}_h$ as

$$f(x) = \sum_{k\in\mathbb{Z}} f(kh)\mathrm{sinc}\left(\frac{x - kh}{h}\right).$$

Another expansion kernel is given by the well–known formula

$$\exp\left(-\frac{x^2t^2 - 2txy + y^2t^2}{2(1-t^2)}\right) = \sqrt{1-t^2}\sum_{n=0}^{\infty} H_n(x)H_n(y)\frac{t^n}{n!}$$

of Mehler (cf. [Wat33]) with $x, y \in \mathbb{R}$. It consists of a weighted expansion into Hermite polynomials H_n on $\mathbb{R}$, which are orthogonal with weight $\exp(-x^2/2)$.

But first we stay general and note that we have not restricted ourselves to spaces of functions, so far.

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

$$f = \sum_{n\in\mathbb{N}} (f, \varphi_n)_{\mathcal{H}} \varphi_n \tag{4.1}$$

with the **Parseval equation**

$$\|f\|_{\mathcal{H}}^2 = \sum_{n\in\mathbb{N}} (f, \varphi_n)_{\mathcal{H}}^2 < \infty.$$

This is fine in the Hilbert space context, but even if all φ_n can be interpreted as functions on a domain Ω, it is not clear if an expression like

$$f(x) = \sum_{n\in\mathbb{N}} (f, \varphi_n)_{\mathcal{H}} \varphi_n(x)$$

makes sense. In fact, 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 $\mathcal{H}$. As we have seen in Section 2.12, these Hilbert spaces can always be interpreted as spaces of functions on their own dual, but this viewpoint is too theoretical at this point.

As well–known from Fourier series, the situation is better if the coefficients of the expansion satisfy a decay condition. We mimic this in general by formally introducing positive and uniformly bounded weights λ_n into the above sum, leading to

$$f(x) = \sum_{n\in\mathbb{N}} \frac{(f, \varphi_n)_{\mathcal{H}}}{\sqrt{\lambda_n}} \sqrt{\lambda_n}\varphi_n(x)$$

and the bound

$$|f(x)|^2 \le \left(\sum_{i\in\mathbb{N}} \frac{|(f, \varphi_n)_{\mathcal{H}}|^2}{\lambda_n}\right) \left(\sum_{i\in\mathbb{N}} \lambda_n |\varphi_n(x)|^2\right)$$

if we assume that both factors on the right are bounded. We see that we are coming back towards (1.7) on page 8, but we have no kernel yet. We simply assume the **summability condition**

$$\sum_{i\in\mathbb{N}} \lambda_n |\varphi_n(x)|^2 < \infty \text{ for all } x \in \Omega \tag{4.2}$$

and define the function space by

$$\mathcal{H}_1 := \{f \ : \ (4.1) \text{ with } \sum_{i\in\mathbb{N}} \frac{|(f,\varphi_n)_{\mathcal{H}}|^2}{\lambda_n} < \infty\}. \tag{4.3}$$

The inner product of two functions in $\mathcal{H}_1$ then is defined as

$$(f,g)_{\mathcal{H}_1} := \sum_{n\in\mathbb{N}} \frac{(f,\varphi_n)_{\mathcal{H}}\overline{(g,\varphi_n)_{\mathcal{H}}}}{\lambda_n}. \tag{4.4}$$

Uniform boundedness of the weights implies that $\mathcal{H}_1 \subset \mathcal{H}$ with bounded embedding, and since we have a space that can be isometrically mapped to ℓ_2 via coefficients, we have a Hilbert space. Thus $\mathcal{H}_1$ is a closed subspace of $\mathcal{H}$, and we have made sure that point evaluation functionals are continuous on $\mathcal{H}_1$.

From the definition (4.4) of the inner product on $\mathcal{H}_1$ we see that the φ_n are orthogonal, but not orthonormal in $\mathcal{H}_1$. But the functions $\sqrt{\lambda_n}\varphi_n$ are.

Now Theorem 2.4 on page 12 implies that $\mathcal{H}_1$ has a reproducing kernel K such that $K(x,\cdot)$ is the Riesz representer of δ_x. We assert that

$$K(x,y) := \sum_{n\in\mathbb{N}} \lambda_n \overline{\varphi_n(x)}\varphi_n(y) \tag{4.5}$$

does the job. The expansion coefficients of $K(x,\cdot)$ in $\mathcal{H}$ are

$$(K(x,\cdot),\varphi_n)_{\mathcal{H}} = \lambda_n \overline{\varphi_n(x)},$$

and they satisfy

$$\sum_{\in\mathbb{N}} \frac{\lambda_n^2 |\varphi_n(x)|^2}{\lambda_n} \le \sum_{i\in\mathbb{N}} \lambda_n |\varphi_n(x)|^2 < \infty,$$

proving $K(x,\cdot) \in \mathcal{H}_1$. Furthermore,

$$\begin{aligned}
(f, K(x,\cdot))_{\mathcal{H}_1} &= \sum_{n\in\mathbb{N}} \frac{(f,\varphi_n)_{\mathcal{H}}\overline{(K(x,\cdot),\varphi_n)_{\mathcal{H}}}}{\lambda_n} \\
&= \sum_{n\in\mathbb{N}} (f,\varphi_n)_{\mathcal{H}}\varphi_n(x) \\
&= f(x) \text{ for all } x \in \Omega
\end{aligned}$$

proves the reproduction equation, with absolute summability of the series.

Theorem 4.6. *If a Hilbert space $\mathcal{H}$ of functions on Ω has a countable orthonormal basis $\{\varphi_n\}_{n\in\mathbb{N}}$, each summability property of the form (1.7) leads to a reproducing expansion kernel (4.5) for a suitable subspace $\mathcal{H}_1$ with (4.3) of functions with continuous point evaluation. In the space $\mathcal{H}_1$, the functions $\sqrt{\lambda_n}\varphi_n$ are orthonormal.* □

Note that the reproduction equation uses the inner product $(.,.)_{\mathcal{H}_1}$. The mapping

$$x \mapsto (f, K(x,\cdot))_{\mathcal{H}}$$

does not yield $f(x)$, but rather something like a convolution $(f * K)(x)$. We evaluate it as

$$\begin{aligned}(f, K(x,\cdot))_{\mathcal{H}} &= \sum_{n\in\mathbb{N}} (f,\varphi_n)_{\mathcal{H}}\overline{(K(x,\cdot),\varphi_n)_{\mathcal{H}}} \\ &= \sum_{n\in\mathbb{N}} \lambda_n (f,\varphi_n)_{\mathcal{H}}\varphi_n(x)\end{aligned}$$

and this function lies in

$$\mathcal{H}_2 := \{f \ : \ (4.1) \text{ with } \sum_{i\in\mathbb{N}} \frac{|(f,\varphi_n)_{\mathcal{H}}|^2}{\lambda_n^2} < \infty\}.$$

Thus we get a scale of spaces

$$\mathcal{H}_m := \{f \ : \ (4.1) \text{ with } \sum_{i\in\mathbb{N}} \frac{|(f,\varphi_n)_{\mathcal{H}}|^2}{\lambda_n^m} < \infty\}.$$

with

$$\mathcal{H} =: \mathcal{H}_0 \supset \mathcal{H}_1 \supset \mathcal{H}_2 \supset \ldots$$

which are connected by the convolution map $f \mapsto f * K$. They carry inner products

$$(f,g)_{\mathcal{H}_m} := \sum_{n\in\mathbb{N}} \frac{(f,\varphi_n)_{\mathcal{H}}\overline{(g,\varphi_n)_{\mathcal{H}}}}{\lambda_n^m},$$

and have reproducing kernels

$$K_m(x,y) := \sum_{n\in\mathbb{N}} \lambda_n^m \overline{\varphi_n(x)}\varphi_n(y).$$

Note that the map $f \mapsto f * K = (f, K(x,\cdot))_{\mathcal{H}}$ coincides with the integral operator defined by a kernel, if it exists, and if the Mercer theorem holds. This is easy to see, since both operators act on expansions by coefficient–wise multiplication with the weights, i.e. they are **multipliers**. An integral or a connection to an L_2 space is not necessary.

Just for curiosity, let us ignore continuity of point evaluation for a moment. For linear functionals $\lambda, \mu \in \mathcal{H}^*$ we have square summability of the sequences $\lambda(\varphi_n),\ \mu(\varphi_n)$ and

$$(\lambda, \mu)_{\mathcal{H}^*} = \sum_{n \in \mathbb{N}} \lambda(\varphi_n)\overline{\mu(\varphi_n)}.$$

The kernel

$$K_0(x, y) := \sum_{n \in \mathbb{N}} \overline{\varphi_n(x)}\varphi_n(y)$$

makes no pointwise sense, but it satisfies

$$\begin{aligned} \lambda^x(\overline{\mu^y K_0(x, y)}) &= \sum_{n \in \mathbb{N}} \lambda^x(\varphi_n(x))\overline{\mu^y(\varphi_n(y))} \\ &= (\lambda, \mu)_{\mathcal{H}^*} \end{aligned}$$

i.e. it makes perfect sense in view of (2.62) if functionals are applied to the arguments. Again, we used a superscript x at λ^x to denote that λ acts with respect to the variable x.

We finish this section with an approximation result. Let K be an expansion kernel associated to a complete set of orthonormal functions φ_n as above, and let f_N be the orthogonal projection of some element $f \in \mathcal{H}$ to the span of $\varphi_1, \ldots, \varphi_N$ in the original inner product. Because projections have norm one, one cannot assert a convergence rate $f_N \to f$ in the Hilbert space norm which is better than the usual summability property

$$\|f - f_N\|^2_{\mathcal{H}} = \sum_{j=N+1}^{\infty} |(f, \varphi_n)|^2.$$

Things are better if we go to weighted kernels and consider pointwise convergence:

Theorem 4.7. *Let K be an expansion kernel as above, and let f_N be the orthogonal projection of some element $f \in \mathcal{H}$ to the span of $\varphi_1, \ldots, \varphi_N$ in the original inner product. Then*

$$|f(x) - f_N(x)|^2 \leq \|f - f_N\|^2_{\mathcal{H}}(K(x, x) - K_N(x, x)) \tag{4.8}$$

where K_N is the **truncated kernel**

$$K_N(x, y) := \sum_{n=1}^{N} \lambda_n \overline{\varphi_n(x)}\varphi_n(y).$$

Proof:

$$\begin{aligned} |f(x) - f_N(x)|^2 &= \left| \sum_{n=N+1}^{\infty} (f, \varphi_n)\varphi_n(x) \right|^2 \\ &\leq \left(\sum_{n=N+1}^{\infty} \frac{|(f,\varphi_n)|^2}{\lambda_n} \right) \left(\sum_{n=N+1}^{\infty} \lambda_n |\varphi_n(x)|^2 \right) \\ &= \|f - f_N\|_{\mathcal{H}}^2 (K(x,x) - K_N(x,x)). \quad \square \end{aligned}$$

Consequently, if K_N converges quickly to K for $N \to \infty$, this quick convergence goes over to the pointwise convergence of the projections.

Example 4.9

Consider the Gaussian kernel expansion (1.9) on $\Omega = [-1, 1] \in \mathbb{R}$. Then

$$\begin{aligned} K(x,x) - K_N(x,x) &= \exp(-x^2 - y^2) \sum_{n=N+1}^{\infty} \frac{2^n}{n!} (xy)^n \\ &\leq \frac{2^{N+1}}{(N+1)!} = \lambda_{N+1} \end{aligned} \tag{4.10}$$

via the residual of Taylor's formula. Thus

$$|f(x) - f_N(x)|^2 \leq \|f - f_N\|_{\mathcal{H}}^2 \frac{2^{N+1}}{(N+1)!} \leq \|f\|_{\mathcal{H}}^2 \frac{2^{N+1}}{(N+1)!}$$

is a stunningly fast convergence rate, but only for functions in the (small) native space $\mathcal{H}$.

Note that the error bound in Theorem 4.7 is sharp, because equality is attained for $f(x) = K(y, x)$, leading to the inequality (2.5) for $K - K_N$, which is sharp for $x = y$. Thus the convergence rates implied by Theorem 4.7 are what one should go for using interpolation.

4.2 Shannon Kernel

To understand the kernel theory behind the Shannon–Whittaker–Kotelnikov theorem, we can use an expansion into an orthonormal basis in $L_2(\mathbb{R})$. It can be proven (directly or via Fourier transforms) that the functions

$$s_k(x) = \frac{1}{\sqrt{h}} \operatorname{sinc}\left(\frac{x - kh}{h} \right) \text{ for all } k \in \mathbb{Z}$$

are orthonormal in $L_2(\mathbb{R})$, i.e. under the inner product

$$(f,g)_{L_2(\mathbb{R})} := \int_{\mathbb{R}} f(t)g(t)dt.$$

Thus we can write down the expansion kernel

$$K(x,y) := \frac{1}{h}\sum_{k\in\mathbb{Z}} \text{sinc}\left(\frac{x-kh}{h}\right)\text{sinc}\left(\frac{y-kh}{h}\right)$$

to get the reproduction equation

$$f(x) = (f, K(x,\cdot))_{L_2(\mathbb{R})} \text{ for all } x \in \mathbb{R}$$

but this does not hold for all functions in $L_2(\mathbb{R})$, but rather for functions f with

$$\sum_{k\in\mathbb{Z}} |(f,s_k)_{L_2(\mathbb{R})}|^2 = h\sum_{k\in\mathbb{Z}} |f(kh)|^2 < \infty$$

which form a Hilbert space that needs Fourier transforms to be analyzed. It turns out to be the space of functions bandlimited to $[-\pi/h, \pi/h]$, i.e. whose Fourier transforms exist in $L_2(\mathbb{R})$ but vanish outside that interval. Such functions are necessarily infinitely often differentiable, and pointwise evaluation of the function and all derivatives is continuous.

For these functions, the inner product takes the second form

$$(f,g)_{L_2(\mathbb{R})} = \int_{\mathbb{R}} f(t)g(t)dt = h\sum_{k\in\mathbb{Z}} f(kh)g(kh).$$

Since for each fixed x the function

$$S_x(y) = \frac{1}{h}\text{sinc}((x-y)/h)$$

is in that space, we can write the inner product with f in two ways:

$$\begin{aligned}(f,S_x)_{L_2(\mathbb{R})} &= \sum_{k\in\mathbb{Z}} f(kh)\text{sinc}((x-kh)/h) \\ &= f(x)\end{aligned}$$

to see that the series kernel $K(x,y)$ is identical to $S_x(y)$, i.e.

$$K(x,y) = \frac{1}{h}\sum_{k\in\mathbb{Z}} \text{sinc}\left(\frac{x-kh}{h}\right)\text{sinc}\left(\frac{y-kh}{h}\right) = \frac{1}{h}\text{sinc}\left(\frac{x-y}{h}\right).$$

This nice summation formula can also be proven by using the fact that the space is translation–invariant, so that the kernel must also be translation–invariant by Theorem 2.14 from page 16.

By the above discussion, we also see that the set $h\mathbb{Z} \subset \Omega = \mathbb{R}$ is unisolvent for $\mathcal{H}$ in the sense of Definition 2.59 on page 33.

A generalization to higher dimensions leads to the radial Bessel kernels

$$\phi_\nu(r) = r^{-\nu} J_\nu(r),$$

see Theorem 9.17 on page 181. The case $\nu = 1/2$ is the sinc kernel, and some others are in Figure 29 on page 182. These kernels were studied in [FLW06].

4.3 Trigonometric Kernels

Let us look at trigonometric series

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos(nx) + b_n \sin(nx) \right) \tag{4.11}$$

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

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

We can write these via the index set

$$J := (0,0) \cup (\mathbb{N}, 0) \cup (0, \mathbb{N})$$

and

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

as

$$f = \sum_{j \in J} (f, \varphi_j)_{\mathcal{H}} \varphi_j$$

in the sense of convergence in $\mathcal{H}$.

Note that all functions φ_j are uniformly bounded, such that the summability condition (1.7) is satisfied whenever the weights are summable. This works when the weights are n^{-2}, and thus we start with

$$\begin{aligned} K_1(x,y) &:= \frac{1}{\sqrt{2}} + \sum_{n=1}^{\infty} n^{-2} \left(\cos(nx)\cos(ny) + \sin(nx)\sin(ny)\right) \\ &= \frac{1}{\sqrt{2}} + \sum_{n=1}^{\infty} n^{-2} \cos(n(x-y)). \end{aligned}$$

In view of the previous section, where we had a sequence of kernels with weights λ_n^m, we define

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

to get the expansion kernels

$$\begin{aligned} K_{2m}(x,y) &:= \frac{1}{\sqrt{2}} + \sum_{n=1}^{\infty} n^{-2m} \left(\cos(nx)\cos(ny) + \sin(nx)\sin(ny)\right) \\ &= \frac{1}{\sqrt{2}} + \sum_{n=1}^{\infty} n^{-2m} \cos(n(x-y)) \end{aligned} \tag{4.12}$$

which must be positive semidefinite on $\Omega = [0, 2\pi)$. Plotting the kernel K_2 (see top left of Figure 8) 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.

To verify this by elementary means, 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.

In more generality, the functions

$$\sum_{n=1}^{\infty} \frac{1}{n^{2k}} \cos(n\, t) \tag{4.13}$$

represent polynomials of degree $2k$ on $[0, 2\pi]$. To see this, consider Hurwitz-Fourier expansions

$$B_m(x) = -\frac{m!}{(2\pi i)^m} \sum_{n=-\infty, n\neq 0}^{+\infty} n^{-m} e^{2\pi i n x}$$

of the Bernoulli polynomials[1] B_m of degree m on $[0, 1]$. If we set $t = 2\pi x$ and $m = 2k$, we get

$$\begin{aligned}
B_{2k}(\tfrac{t}{2\pi}) &= (-1)^{k+1} \frac{(2k)!}{(2\pi)^{2k}} \sum_{n=-\infty, n\neq 0}^{+\infty} n^{-2k} (\cos(nt) + i \sin(nt)) \\
&= 2(-1)^{k+1} \frac{(2k)!}{(2\pi)^{2k}} \sum_{n=1}^{+\infty} n^{-2k} \cos(nt)
\end{aligned}$$

that proves our claim. 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 (4.11) and K_{2m} takes the form

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

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}

[1] `http://mathworld.wolfram.com/BernoulliPolynomial.html`

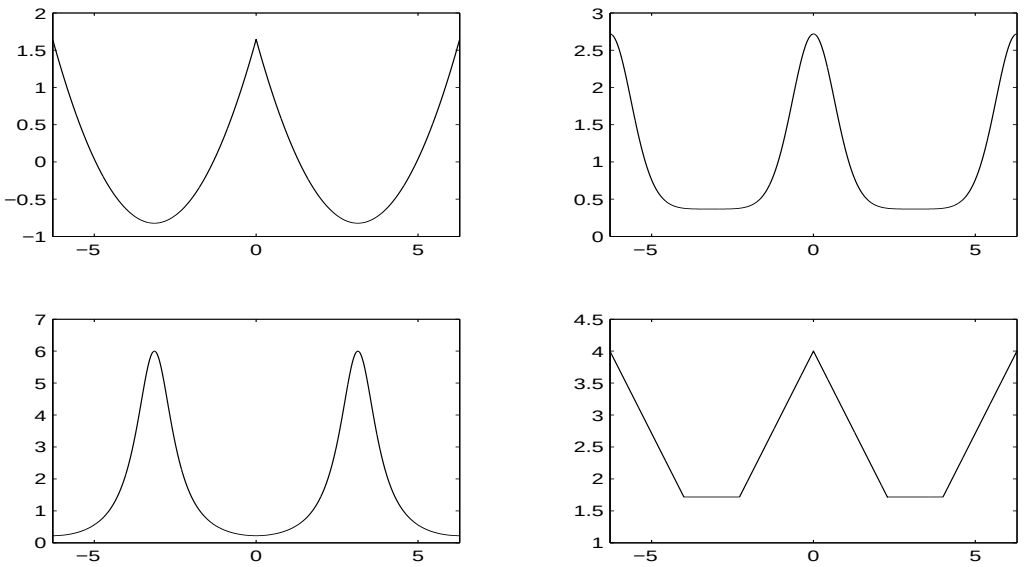

Figure 8: Periodic kernels

is the reproducing kernel of the **Sobolev space** of order $2m$ for univariate 2π–periodic functions.

From Anette Meyenburg's thesis [Mey96] we cite the infinitely differentiable periodic kernels

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{1}{n!} \cos(nx) &= \cos(\sin(x)) \cdot \exp(\cos(x)) \\ \sum_{n=0}^{\infty} \frac{1}{2^n} \cos(nx) &= \frac{1 - \frac{1}{2}\cos(x)}{1 - \cos(x) + \frac{1}{4}}. \end{aligned} \tag{4.14}$$

Their proofs are easy, if the cosines of the left–hand side are written as exponentials. They are plotted in Fig. 8 on the top right and bottom left, respectively. 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.

By standard Fourier analysis, one also gets [Mey96]

$$\begin{aligned}
\exp(-2|x|) &= \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{1-(-1)^n e^{-2\pi}}{4+n^2}\cos(nx),\ x\in[-\pi,\pi],\\
K(x) &:= \begin{cases} 4-x & 0\le x \le 2\pi-4\\ 8-2\pi & 2\pi-4< x <4\\ 4-2\pi+x & 4\le x\le 2\pi\end{cases}\\
&= \frac{16}{\pi}+\sum_{n=1}^{\infty}\frac{4\sin^2(2n)}{\pi n^2}\cos(nx).
\end{aligned}$$

The final one is plotted in the bottom right of Fig. 8.

We now want to apply Theorem 4.7 to Fourier series. We have to evaluate the errors of the truncated kernels. In case of $\lambda_n = n^{-2m}$ we have to bound

$$\begin{aligned}
K(x,x)-K_{N-1}(x,x) &= \sum_{n=N}^{\infty}\frac{1}{n^{2m}}\\
&= \sum_{p=1}^{\infty}\sum_{q=0}^{N-1}\frac{1}{(pN+q)^{2m}}\\
&\le \frac{1}{N^{2m-1}}\sum_{p=1}^{\infty}\frac{1}{p^{2m}}\\
&= \frac{\zeta(2m)}{N^{2m-1}}.
\end{aligned}\tag{4.15}$$

Now Theorem 4.7 implies

Theorem 4.16. *In the native Hilbert spaces*

$$\mathcal{H}_m := \left\{ f \ :\ (4.11),\ \sum_{n=0}^{\infty}(a_n^2+b_n^2)n^{2m}<\infty\right\}\tag{4.17}$$

for the expansion kernels of (4.12), the pointwise convergence rate of partial sums of trigonometric series (4.11) truncated at $n=N$ has the behavior $o(N^{-m+1/2})$ for $N\to\infty$ and fixed $m\ge 1$. The kernels of (4.14) lead to pointwise convergence rates $o(\frac{1}{\sqrt{N!}})$ and $o(2^{-N/2})$, respectively, in their associated native Hlbert spaces defined via (4.3). □

Note that all trigonometric kernels in cosine form can be transformed by the standard transformation $x=\cos(\varphi)$ into series of Chebyshev polynomials on $[-1,1]$. For instance, the kernel

$$K(x,y) := \sum_{n=0}^{\infty}\frac{1}{n!}T_n(x)T_n(y)$$

on $[-1,1]$ can be transformed by substitution $x = \cos(\varphi),\ y = \cos(\psi)$ into

$$\begin{aligned} & \sum_{n=0}^{\infty} \frac{1}{n!} \cos(n\varphi)\cos(n\psi) \\ = \ & \frac{1}{2} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\cos(n(\varphi+\psi)) + \cos(n(\varphi-\psi))\right) \\ = \ & \frac{1}{2} \left[\cos(\sin(\varphi+\psi)) \cdot \exp(\cos(\varphi+\psi)) + \cos(\sin(\varphi-\psi)) \cdot \exp(\cos(\varphi-\psi))\right]. \end{aligned}$$

4.4 Taylor Kernels

There is still another variation on the theme of periodic kernels. For all complex–valued functions which are holomorphic on the interior of the unit disc and still L_2 on the unit circle C, we can define an inner product by

$$(f,g)_H := \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\varphi}) \overline{g(e^{i\varphi})} d\varphi = \frac{1}{2\pi i} \int_C \frac{f(z)\overline{g(z)}}{z} dz.$$

This generates the **Hardy space**[2] H^2, and complex polynomials z^n are orthonormal in this space for $n \geq 0$. Consequently, we can consider expansion kernels

$$K(u,z) := \sum_{n=0}^{\infty} \lambda_n \overline{u}^n z^n =: \Phi(\overline{u}z). \tag{4.18}$$

with nonnegative and absolutely summable weights λ_n. All power series with nonnegative coefficients and convergence radius at least 1 provide examples of expansion kernels. If we confine the functions to the unit circle, we can set $u = \exp(i\psi)$ and $z = \exp(i\varphi)$ to get periodic complex–valued expansion kernels

$$K(\psi,\phi) := \sum_{n=0}^{\infty} \lambda_n \exp(in(\varphi-\psi)) =: \Phi(\varphi-\psi) \tag{4.19}$$

and if we focus on real–valued kernels, we are back to the cosine series we started with.

But there is no need to confine everything to the unit circle, since we know that the functions have unique extensions to the full disc, determined by their values on the circle. **Cauchy's integral formula**

$$f(z) = \frac{1}{2\pi i} \int_C \frac{f(\zeta)}{\zeta - z} d\zeta$$

[2] `http://en.wikipedia.org/wiki/Hardy_space`

then is a reproduction formula, and its kernel is the **Szegö kernel**

$$K(u,z) = \frac{1}{1-\overline{u}z} = \sum_{n=0}^{\infty} \overline{u}^n z^n.$$

This kernel cannot be evaluated when both arguments are on the circle, which is to be expected because functions in Hardy space have no continuous point evaluation on the circle itself. Interpolation of functions on point sets $Z = \{z_1, \ldots, z_N\}$ inside the unit disc are no problem, and the result is a linear combination of rational functions

$$K(z_j,z) = \frac{1}{1-\overline{z_j}z}$$

which have "mirror" singularities at $z = z_j/|z_j|^2$ outside the disc. The interpolant is an optimal recovery of functions from Hardy space, including being norm–minimal in L_2 on the circle among all other conceivable interpolants.

But this it not all we can say, by far. Clearly, the functions we want to consider here have expansions in **Taylor series**

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n$$

with at least convergence radius 1. We can view this as an expansion into the orthonormal basis, with

$$(f, z^n)_H = \frac{f^{(n)}(0)}{n!}.$$

If we define a weighted kernel by (4.18), Theorem 4.6 shows that it reproduces in the Hilbert subspace of functions with the summability condition

$$\sum_{n=0}^{\infty} \frac{|(f,z^n)_H|^2}{\lambda_n} = \sum_{n=0}^{\infty} \frac{|f^{(n)}(0)|^2}{(n!)^2\lambda_n}.$$

and the inner product

$$(f,g)_\lambda := \sum_{n=0}^{\infty} \frac{f^{(n)}(0)\overline{g^{(n)}(0)}}{\lambda_n (n!)^2}.$$

Theorem 4.20. *In all of these cases, independent of the weights used, the reproduction formula is the* **Taylor formula**. *The corresponding Hilbert spaces can be called* **Taylor spaces**.

Proof: We just work it out by first calculating

$$\frac{d^j}{dz^j}_{|0} K(u,z) = \lambda_j j! \overline{u}^j,\ j \geq 0$$

and then

$$\begin{aligned} f(u) &= (f, K(u,\cdot)) \\ &= \sum_{n=0}^{\infty} \lambda_n \frac{f^{(n)}(0)\lambda_n n! u^n}{\lambda_n (n!)^2} \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} u^n. \quad \square \end{aligned}$$

This gives a whole range of specific kernels and associated Hilbert subspaces of the Hardy space. See Table 1 for a list in the notation of (4.18), where we introduced subsets $\mathcal{N}$ of $\mathbb{N}$ in order to care for expansions in even and odd terms. If the kernels exist for $|z| = 1$, they can be brought to the unit circle and used for periodic functions via (4.19).

$\Phi(z) = \sum_{j \in \mathcal{N}} \lambda_j z^j$	$\mathcal{N}$	λ_j
$(1-z)^{-1}, -1 < \|z\| < 1$	$\mathbb{N}$	1
$(1-z^2)^{-1}, -1 < \|z\| < 1$	$2\mathbb{N}$	1
$(1-z)^{-\alpha}, \alpha \in \mathbb{N}, -1 < \|z\| < 1$	$\mathbb{N}$	$\frac{(\alpha+j-1)!}{(\alpha-1)!j!}$
$-\frac{\log(1-z)}{z}, -1 < \|z\| < 1$	$\mathbb{N}$	$\frac{1}{j+1}$
$\exp(z)$	$\mathbb{N}$	$1/j!$
$(1-z^2)^{-\alpha}, \alpha \in \mathbb{N}, -1 < \|z\| < 1$	$2\mathbb{N}$	$\frac{(\alpha+j-1)!}{(\alpha-1)!j!}$
$\sinh(z)$	$2\mathbb{N}+1$	$1/j!$
$\sinh(z)/z$	$2\mathbb{N}$	$\frac{1}{(j+1)!}$
$\cosh(z)$	$2\mathbb{N}$	$1/j!$
$z^{-\alpha} I_\alpha(z)$	$2\mathbb{N}$	$\frac{1}{j! 4^j \Gamma(j+\alpha+1)}$

Table 1: Kernels for Taylor Spaces

A more detailed analysis of Taylor spaces is in a preprint [ZS10].

4.5 Native Spaces of Expansion Kernels

We now want to turn the above situation upside down, starting with an expansion kernel and nothing else. We want to arrive at a Hilbert space

of functions on Ω with continuous point evaluation such that the expansion kernel is reproducing. Note that this is the standard case in Machine Learning with kernels.

Thus we start with an expansion kernel (1.6) on page 8 with the summability condition (1.7) 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 (1.8) under the inner product

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

The dual space then is $\ell_{2,1/\lambda,I}$, and the action of a functional $\mu \in \ell_{2,1/\lambda,I}$ on a sequence $c \in \ell_{2,\lambda,I}$ is defined as

$$\mu(c) := \sum_{i\in I} \mu_i c_i.$$

With this definition, the Riesz map is

$$R \,:\, \ell_{2,1/\lambda,I} \to \ell_{2,\lambda,I},\ R(\mu) = \{\overline{\mu_i}/\lambda_i\}_{i\in I}. \tag{4.21}$$

More details on sequence spaces are in Section 11.3 on page 220.

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 we expect is

$$K(x,y) = \sum_{i\in I} \lambda_i \overline{\varphi_i(x)} \varphi_i(y) = (\Phi(y), \Phi(x))_{\lambda,I} \text{ for all } x, y \in \Omega,$$

but we have no Hilbert space and no reproduction formula yet.

If the φ_n are linearly independent, one can go easily over to the native space via sequence spaces of expansion coefficients, using the spaces and inner products of Section 4.1. But if linear independence of the φ_i is not guaranteed, we cannot work with coefficients of expansions into the φ_i in a simple way. There is a workaround by R. Opfer [Opf06] using frames, but this is unnecessarily complicated. Instead, we directly work in subspaces of sequence spaces. The "functions" will be sequences in $\ell_{2,1/\lambda,I}$, while "functionals" will be in $\ell_{2,\lambda,I}$. We define point evaluations of "functions" $c = \{c_i\}_i \in \ell_{2,1/\lambda,I}$ via

$$\delta_x(c) := \sum_i c_i \varphi_i(x),$$

and they are continuous due to

$$|\delta_x(c)|^2 \leq \left(\sum_i \frac{|c_i|^2}{\lambda_i}\right)\left(\sum_i |\varphi_i(x)|^2\lambda_i\right) = \|c\|^2_{\ell_{2,1/\lambda,I}} K(x,x).$$

Their Riesz images are sequences

$$R(\delta_x) = \{\overline{\varphi_i(x)}\lambda_i\}_i$$

(note that we go in the reverse direction of (4.21)) as "functions", such that their evaluation at y is

$$\delta_y(R(\delta_x)) = \sum_i \varphi_i(y)\overline{\varphi_i(x)}\lambda_i = K(x,y)$$

as expected. The reproduction equation is

$$(c, R(\delta_x)) \quad = \quad \delta_x(c)$$

by definition. But not all of the sequence spaces come out to be allowed. The dual of the native space will be

$$\mathcal{L} := \text{ clos span } \{\delta_x \ : \ x \in \Omega\}$$

while the native space $\mathcal{H}$ is the closure of the span all sequences $R(\delta_x)$ for $x \in \Omega$. This makes perfect sense, but it is not so easy to evaluate these spaces for specific applications. In particular, it is not guaranteed that the φ_i are in the native space. And, they cannot be an orthonormal system like in our starting point in Section 4.1, because they need not even be nonzero or linearly independent at this point.

But there is a formal trick to come back to the orthonormal basis. We artificially extend Ω by the index set I and postulate function values

$$\varphi_k(j) = \delta_{jk}, \ j,k \in I$$

there, leaving the values on Ω unchanged, and doing no harm to the summability condition. Now we can use the reproduction equation for

$$(c, R(\delta_j)) = c_j \text{ for all } j \in I$$

and

$$\lambda_j\varphi_j(y) = \delta_y(R(\delta_j)) = K(j,y)$$

to see that the φ_j are in the native space now. Their inner product there is

$$(K(j,y), K(k,y))_{\mathcal{H}} = K(j,k) = \lambda_j\delta_{jk} \text{ for all } j,k \in I,$$

as expected from Section 4.1. In the unweighted space $\ell_{2,1,I}$ their sequence representations are the unit sequences, thus they are orthonormal there. We do not assume continuous point evaluation on "functions" in the full space $\ell_{2,1,I}$, since we have posed our special summability condition that forces us to use weights. We summarize:

Theorem 4.22. *The native space for an expansion kernel on Ω with weights $\lambda_i,\ i \in I$ and features $\varphi_i,\ i \in I$ is isometrically isomorphic to a closed subspace of $\ell_{2,1/\lambda,I}$, while its dual is a closed subspace of $\ell_{2,\lambda,I}$ in Riesz relation, being the closure of all point evaluation functionals. By a suitable formal extension of Ω one can come back to the situation in Theorem 4.6 on page 51.*

The problem is to characterize the spaces $\mathcal{H}$ and $\mathcal{L}$ in more detail. But this is dependent on the specific example.

4.6 Error Analysis of Expansion Kernels

RS: this is still under research, as of February 1, 2011.

Theorem 4.23. *For any finite set $X = \{x_1, \ldots, x_N\} \subseteq \Omega$ and an expansion kernel K with the summability condition (4.2), the pointwise norm of the error functional has the form*

$$\begin{aligned} & \left\| \delta_x - \sum_{j=1}^{N} u_j(x) \delta_{x_j} \right\|_{\mathcal{H}^*}^2 \\ = & \sum_n \lambda_n \left| \varphi_n(x) - \sum_{j=1}^{N} u_j(x) \varphi_n(x_j) \right|^2 \\ \geq & P_X^2(x) \end{aligned}$$

If this is minimized over all $u_j(x) \in \mathbb{K}$, the Power Function $P_X^2(x)$ results, and the (existing) optimal solution $u_j^*(x)$ satisfies the linear system

$$K(x_k, x) = \sum_{j=1}^{N} u_j^*(x) K(x_k, x_j),\ 1 \leq k \leq N.$$

Proof: The identity follows from direct calculation, the inequality via the definition of the Power Function, and the linear system follows from standard orthogonality properties of the optimum. □

If an interpolation process for a set X yields a small power function, and if the weights λ_n decrease rapidly, then there must necessarily be a good

recovery of the first φ_n from the data. More precisely,

$$\left| \varphi_n(x) - \sum_{j=1}^{N} u_j^*(x) \varphi_n(x_j) \right|^2 \leq \frac{P_X^2(x)}{\lambda_n}$$

for all n.

Theorem 4.24. *Assume that a set $X \subset \Omega$ is unisolvent for a space $P_M := span\ \{\varphi_1, \ldots, \varphi_M\}$ of functions on Ω. Then interpolation on X with values in P_M is possible by a linear process, and it recovers function from P_M exactly.*

Proof: This should be well–known, but we give a proof here. Consider the evaluation map E_X with

$$E_X(f) := (f(x_1), \ldots, f(x_N))^T \in \mathbb{K}^N.$$

By unisolvency, it is injective on P_M, and thus there is an inverse map back to P_M on the range $E_X(P_M)$. This means that interpolation on X can be written as a map

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

which is the identity on P_M, and where the functions u_j are in $\mathcal{P}_M$. □

In our context, this works like Shannon's theorem. If high–frequency parts of f are missing, recovery can be exact.

We add another triviality:

Theorem 4.25. *Oversampling can be used to stabilize linear interpolation processes on unisolvent sets for fixed trial spaces.*

Proof: If we work with the notation of the previous theorem, oversampling means that $N >> M$, and the functions u_j are not unique. Thus one can minimize norms like

$$\|u(x)\|_2^2 = \sum_{j=1}^{N} |u_j(x)|^2 \tag{4.26}$$

over N variables under the M constraints

$$\varphi_k(x) = \sum_{j=1}^{N} u_j(x) \varphi_k(x_j),\ 1 \leq k \leq M$$

for each fixed x. By standard arguments of quadratic optimization, the solution has the form

$$u_j(x) = \sum_{k=1}^{M} v_k(x)\overline{\varphi_k(x_j)},\ 1 \le j \le N$$

where the functions v_k satisfy the normal equations

$$\varphi_m(x) = \sum_{k=1}^{M} v_k(x) \sum_{j=1}^{N} \overline{\varphi_k(x_j)}\varphi_m(x_j),\ 1 \le m \le M.$$

Increasing the number N of data points yields more degreees of freedom for the minimization, and thus the optimal value of (4.26) gets smaller if N is increased while P_M is fixed. □

Theorem 4.27. *Let X be unisolvent for P_M, and let $u_j^M(x)$ be the recovery functions on P_M from values on X, possibly with quite some stability, i.e. a reasonably bounded value (4.26) due to oversampling. Then*

$$\begin{aligned} P_X^2(x) \ &\le \ \left(1 + \sum_{j=1}^{N} |u_j^M(x)|^2\right) \cdot \\ &\quad \cdot \left(K(x,x) - K_M(x,x) + \sum_{j=1}^{N} (K(x_j,x_j) - K_M(x_j,x_j))\right). \end{aligned}$$

Proof: Just consider

$$\begin{aligned} P_X^2(x) \ &\le \ \sum_{n=M+1}^{\infty} \lambda_n \left|\varphi_n(x) - \sum_{j=1}^{N} u_j^M(x)\varphi_n(x_j)\right|^2 \\ &\le \ \left(1 + \sum_{j=1}^{N} |u_j^M(x)|^2\right) \sum_{n=M+1}^{\infty} \lambda_n \left(|\varphi_n(x)|^2 + \sum_{j=1}^{N} |\varphi_n(x_j)|^2\right) \\ &= \ \left(1 + \sum_{j=1}^{N} |u_j^M(x)|^2\right) \left(K(x,x) - K_M(x,x)\right. \\ &\quad \left. + \sum_{j=1}^{N} (K(x_j,x_j) - K_M(x_j,x_j))\right). \end{aligned}$$

If the λ_n are decreasing quickly, one can have $K - K_M$ very small or even numerically zero for reasonably small M. Then the above result says that if X is large enough to be unisolvent on P_M and to allow enough oversampling to let the first factor in the bound be not too large, the interpolation error using the kernel K will be small.

Example 4.28

Consider the kernel

$$K(x,y) = \exp(2xy) = \sum_{n=0}^{\infty} \frac{2^n}{n!} x^n y^n$$

arising within the expansion (1.9) on $\Omega = [-1,1] \in \mathbb{R}$. It has a similar bound like (4.10). For point sets $X = X_M$, we choose the M zeros or extrema of the appropriate Chebyshev polynomials, and consider recovery of polynomials. Then it is well–known that the standard Lebesgue constants and thus also the absolute maxima of the u_j behave like $\log(M)$, leading to a $\mathcal{O}(\log(M))$ bound of (4.26). Consequently,

$$P^2_{X_M}(x) \le C \log(M) \frac{2^M}{M!} \text{ for all } x \in [-1,1].$$

RS: the plots are to be changed, they still belong to the Gaussian

Figure 9 shows the φ_n, the Power Function and its upper bound of Theorem 4.23 , the Lagrange bases using either kernel translates or polynomials, and the corresponding Lebesgue functions (4.26) for $M = 12$, in reading order. The actual bounds are

$$\begin{aligned} \lambda_{12} &\le 4.3 \cdot 10^{-6} \\ K - K_{12} &\le 1.0 \cdot 10^{-7} \\ P^2_{X_{12}} &\le 6.0 \cdot 10^{-13} \\ \|u(x)\|_2^2 &\le 1.3. \end{aligned}$$

The bound in Theorem 4.23 is quite sharp, while the upper bound of Theorem 4.27 has some leeway in the second inequality sign in the proof.

Example 4.29

Let us consider the Taylor spaces of Section 4.4 which have the Taylor series as a reproduction formula and complex–valued kernels

$$K(u,z) := \sum_{n=0}^{\infty} \lambda_n \overline{u}^n z^n$$

in the unit disc. We consider their restriction to a real interval $\Omega = [-R,R]$ with $0 < R < 1$ and assume that the λ_n decrease with increasing n. Then the truncated kernels can be bounded by

$$\begin{aligned} K(x,x) - K_M(x,x) &= \sum_{n>M} \lambda_n |x|^{2n} \\ &\le \lambda_{M+1} \sum_{n>M} |x|^{2n} \\ &\le \lambda_{M+1} \frac{R^{2M+2}}{1-R^2}. \end{aligned}$$

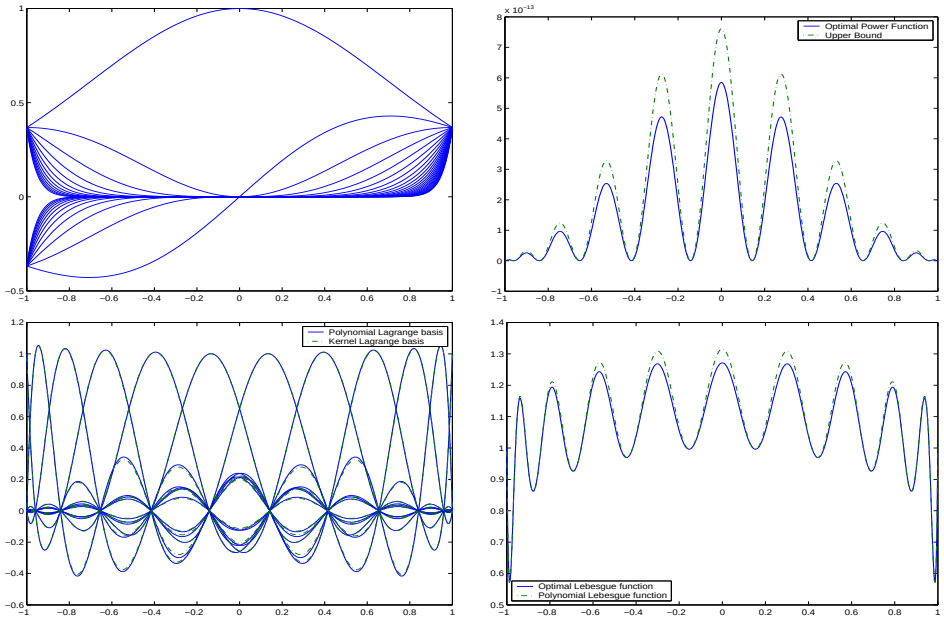

Figure 9: Gaussian Expansion Kernel Plots

Now we consider interpolation in $N+1 \geq M+1$ points of $[-R, R]$ such that polynomials up to degree M are recovered. The power function will then be bounded by

$$P_X^2(x) \quad \leq \quad \lambda_{M+1} \frac{(M+2)R^{2M+2}}{1-R^2} \left(1 + \sum_{j=0}^{N} |u_j^N(x)|^2 \right)$$

with the appropriate recovery functions $u_0, \ldots, u_N$. Note that this will yield geometric convergence for $M \to \infty$, if (4.26) can be kept under control, which is a nontrivial problem.

If we take $N+1 = M+1$ equidistant points, the standard Lebesgue constant will be

$$\sum_{j=0}^{M} |u_j^M(x)| \leq C \frac{2^{M+1}}{eM \log(M)}$$

such that

$$1 + \sum_{j=0}^{M} |u_j^M(x)|^2 \leq C \frac{2^{2M+2}}{M^2 \log^2(M)}$$

with a constant C which is independent of M. This yields

$$P_X^2(x) \leq C \lambda_{M+1} \frac{(M+2)R^{2M+2}}{1-R^2} \frac{2^{2M+2}}{M^2 \log^2(M)}$$

and leads to geometric convergence to zero if $R < 1/2$ even for the Szegö kernel.

If $M+1$ Chebyshev nodes are used, the standard Lebesgue constant will be of order $\log(M)$, and then

$$1 + \sum_{j=0}^{M} |u_j^M(x)|^2 \le C \log^2(M)$$

with a constant C which is independent of M. Thus

$$P_X^2(x) \le C\lambda_{M+1} \frac{(M+2)R^{2M+2}}{1-R^2} \log^2(M)$$

implies geometric convergence for all $R < 1$ and all kernels considered.

Let us rewrite this in terms of the **fill distance**

$$h := \max_{-R \le y \le R} \min_{x_j \in X} |y - x_j|$$

of a subset X. For $M+1$ Chebyshev zeros on $[-1,1]$, the fill distance is bounded above by $h = \pi/(M+1)$, but we can make it easier by picking $M = \lceil \frac{\pi}{h} \rceil$ for a given h. Then

$$\begin{aligned} R^{2M} &= (R^2)^{\lceil \frac{\pi}{h} \rceil} \\ &= \exp\left(2\lceil \tfrac{\pi}{h} \rceil \log(R)\right) \end{aligned}$$

and

$$P_X^2(x) \le C\lambda_{\lceil \frac{\pi}{h} \rceil + 1} \frac{(\lceil \frac{\pi}{h} \rceil + 2)R^2}{1-R^2} \log^2(\lceil \frac{\pi}{h} \rceil) \exp\left(2\lceil \frac{\pi}{h} \rceil \log(R)\right). \tag{4.30}$$

Even for the Szegö kernel, this is a convergence rate of the form

$$P_X^2(x) \le C \exp(-c/h) \text{ for } h \to 0$$

with certain positive constants C, c.

Theorem 4.31. *If a Taylor kernel with decaying λ_n is given and if one works in $[-R,R]$ for some $0 < R < 1$, one can find for all $h > 0$ a point set X with fill distance at most h such that (4.30) holds.* □

We now start with an arbitrary prescribed point set X with fill distance h and $N+1$ points, but now we employ oversampling. We pick the smallest M with $2M^2 \geq 1/h$, i.e.

$$M = \left\lceil \frac{1}{\sqrt{2h}} \right\rceil,$$

and then we know by Theorem 8.41 on page 151 that we get

$$\sum_{j=0}^{N} |u_j(x)| \leq 2,$$

leading to

$$1 + \sum_{j=0}^{N} |u_j(x)|^2 \leq 5.$$

Consequently, we get the bound

$$P_X^2(x) \quad \leq \quad 5\lambda_{M+1} \frac{(M+2)R^{2M+2}}{1-R^2}$$

and can insert our choice of M for

$$P_X^2(x) \quad \leq \quad 5\lambda_{1+\left\lceil \frac{1}{\sqrt{2h}} \right\rceil} \frac{\left(2 + \left\lceil \frac{1}{\sqrt{2h}} \right\rceil\right) R^2}{1-R^2} R^{2\left\lceil \frac{1}{\sqrt{2h}} \right\rceil} \tag{4.32}$$

which is a convergence result of the form

$$P_X^2(x) \leq C \exp(-c/\sqrt{h}) \tag{4.33}$$

with positive and explicitly obtainable constants.

Theorem 4.34. *For all kernels of Taylor spaces with decaying λ_n, and for all point sets $X \subset [-R, R]$ with fill distance h and $0 < R < 1$, the Power Function is bounded by (4.32) and has exponential convergence to zero for $h \to 0$ with a law like (4.33).* □

Note that the proof implies that such sets have $N+1$ points with at least $hN \geq 2R$, but the polynomial recovery used in the bound will only be exact up to degree M with $2hM^2 \geq 1$. This is a considerable amount of oversampling, but the proof requires it only for going to the general case $0 < R < 1$. The proof for the case $R < 1/2$ does not need oversampling.

RS: Open: what happens in practice? Is the proof still too weak?

RS: To Do: better convergence if λ_n decay fast

Let us consider the case of real–valued 2π–periodic trigonometric kernels from (4.12) and interpolate in the usual $2N+1$ equidistant points in $[-\pi, \pi]$ such that we have a fill distance of $h = \pi/(2N+1)$. Then we know that the Lebesgue constants behave like $\mathcal{O}(\log(N))$, and we can use (4.15) in Theorem 4.27 for

Theorem 4.35. *If the trigonometric kernels (4.12) are used whose native spaces (4.17) are of Sobolev type, and if interpolation in $2N+1$ equidistant points is performed, the pointwise error decays at least like $\mathcal{O}(N^{-m+1}\log(N)) = \mathcal{O}(h^{m-1}|\log(h)|)$ for $N \to \infty$ or $h \to 0$.*

4.7 Finite Case

We now specialize to the context of learning models on a **finite** set Ω consisting of $N = |\Omega|$ points and a **finite–dimensional** feature space. Instead of using point notation for Ω, we can identify Ω with the set $\Omega = \{1, \ldots, N\}$ and use index notation instead, and we assume the feature space to be $\mathbb{K}^L$ for simplicity. Expansion kernels (1.6)

$$K(j,k) := \sum_{\ell=1}^{L} \lambda_\ell \overline{\varphi_\ell(j)} \varphi_\ell(k)$$

then can be written as Hermitian positive semidefinite matrices K with entries $K(j,k),\ 1 \le j,k \le N$ as

$$K = \Phi^* \, \Lambda \, \Phi$$

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

The feature map $j \mapsto \Phi(j) := \{\varphi_\ell(j)\}_\ell \in \mathbb{K}^L$ maps to the N columns of Φ, and thus $\mathcal{L}$ is the subspace of $\mathbb{K}^L$ spanned by the columns of Φ. In most cases, it will be all of $\mathbb{K}^L$, but not necessarily so. Anyway, each element of $\mathcal{L}$ is of the form Φz with $z \in \mathbb{K}^N$. Then $\mu_z := R^{-1}(\Phi z)$ will be $\mu_z = \Lambda \overline{\Phi z}$ and the elements of $\mathcal{H}$ are vectors with elements

$$T(\mu_z)(j) = \sum_{\ell=1}^{L} \varphi_\ell(j) \mu_{z\ell} = \sum_{\ell=1}^{L} \varphi_\ell(j) \lambda_\ell \sum_{k=1}^{N} \overline{\varphi_\ell(k) z_k} = e_j^T \overline{K z},$$

i.e. linear combinations of rows of K. Thus $\mathcal{H}$ is the row span of K, which is also evident from the fact that the native space should be the closure of the $K(j,\cdot)$. Each function f in $\mathcal{H}$ thus is a linear combination of rows of K, and

thus it has the form $f_a := K^T a$ with a vector $a \in \mathbb{K}^N$. The inner product then is

$$(f_a, f_b)_{\mathcal{H}} = a^T K \overline{b} = a^T \Phi^* \Lambda \Phi \overline{b} \text{ for all } a, b \in \mathbb{K}^N.$$

The well–definedness of the inner product can here be checked easily, since for $K^T a = K^T \tilde{a}$ and $K^T b = K^T \tilde{b}$ we get

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

Also, the positive definiteness of the inner product is simple to see, because from $\|f_a\|_K^2 = a^T K \overline{a} = 0$ we first get $\Phi \overline{a} = 0$ from

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

with the nonsingular diagonal matrix $\sqrt{\Lambda}$ defined in an obvious way. But $\Phi \overline{a} = 0$ implies $f_a = K^T a = (\Phi^* \Lambda \Phi)^T a = \Phi^T \Lambda \overline{\Phi} a = 0$.

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 deal with square submatrices of equal row/column selections of the matrix K, i.e. with minors of it, while the kernel is unchanged..

A singular value decomposition splits the kernel matrix K into a product

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

with a unitary $N \times N$ matrix U and a real diagonal $N \times N$ matrix Σ of **singular values** of K, i.e. the nonnegative eigenvalues of $K^* K$. Note that this amounts to consider an equivalent setting with now $L = N$, $U = \Phi$, and $\Lambda = \Sigma$, but now the diagonal of Σ may contain zero entries. The unitary matrix U just is a coordinate change in the native space, and the new N feature functions are orthonormal, but only L of them are used. The kernel gets the equivalent form

$$K(j,k) = \sum_{\ell=1}^{L} \sigma_\ell \overline{u_\ell(j)} u_\ell(k)$$

with the u_ℓ being orthonormal vectors. If small singular values occur here, they can be ignored, thus reducing the kernel's complexity.

5 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. For certain reasons to become apparent later, we shall postpone completion as far as possible.

5.1 Unisolvency

To define a sufficiently general notion of conditional positive (semi–) definiteness, we fix a finite–dimensional space $\mathcal{P}$ of functions on a set Ω, denote its dimension by Q and select a basis $p_1, \ldots, p_Q$. The case of (unconditional) positive (semi–) definiteness, as in Definitions 2.37 and 3.4 refers to the special case $Q = 0$ and $\mathcal{P} = \{0\}$.

Definition 5.1. *A subset $X = \{x_1, \ldots, x_N\}$ of Ω is called* $\mathcal{P}$**–unisolvent**, *if zero is the only function in $\mathcal{P}$ that vanishes on X.*

Looking at the matrix P_X of values $p_j(x_k)$, $1 \le j \le Q$, $1 \le k \le N$, we see that it must have rank Q for unisolvency, thus $N \ge Q$ must hold. Therefore we assume Ω to have at least Q points and contain a unisolvent set. From now on, all subsets X of Ω we shall consider must be $\mathcal{P}$–unisolvent and thus have at least Q points. Later, we shall need

Theorem 5.2. *Each $\mathcal{P}$–unisolvent set X has a unisolvent subset of Q points.*

Proof: Just select a nonsingular $Q \times Q$ submatrix of P_X. □.

$\mathcal{P}$–unisolvency means that functions from $\mathcal{P}$ are completely determined by their values on X. Therefore we can have a recovery formula

$$p(x) = \sum_{j=1}^{N} p(x_j) u_j(x) \text{ for all } p \in \mathcal{P},\ x \in \Omega$$

with a suitable set of functions $u_1, \ldots, u_N$ spanning $\mathcal{P}$. If $|X| = Q$, the u_j will be a Lagrange basis with $u_j(x_k) = \delta_{jk}$, $1 \le j,k \le Q$.

It is instructive to consider minimal $\mathcal{P}$–unisolvent sets for spaces of polynomials over $\mathbb{R}$ in d variables and of degree n. In one dimension, each set of $n+1$ distinct points is unisolvent and minimal, due to the Fundamental Theorem of Algebra. In d real dimensions, minimal unisolvent sets for linear real–valued polynomials are the nondegenerate simplices consisting of $d+1$ points not on a hyperplane. In geometry, certain configurations of unisolvents sets are called "in general position". For instance, minimal unisolvent sets for quadratic polynomials on $\mathbb{R}^2$ consist of 6 points not lying on a conic. But the connection of unisolvent sets to geometry cannot be pursued here any further.

5.2 Conditional Positive Definiteness

Definition 5.3. *Let a finite–dimensional space $\mathcal{P}$ of real–valued functions on a set Ω be given. A Hermitian kernel $K : \Omega \times \Omega \to \mathbb{K}$ is called* $\mathcal{P}$**–conditionally positive (semi–) definite**, *if for all $\mathcal{P}$–unisolvent subsets $X = \{x_1, \ldots, x_N\}$ of Ω the kernel matrices with entries $K(x_j, x_k)$, $1 \le j, k \le N \ge Q$ are positive (semi–) definite on the subspace of $\mathbb{K}^N$ of vectors $a \in \mathbb{K}^N$ with the* **moment conditions**

$$\sum_{j=1}^{N} a_j p(x_j) = 0 \text{ for all } p \in \mathcal{P}. \tag{5.4}$$

If the space $\mathcal{P}$ consists of all polynomials of order (=degree -1) m on Ω, the kernel is conditionally positive (semi–) definite of **order** *m, if it is $\mathcal{P}$–conditionally positive (semi–) definite.*

There are some highly important conditionally positive definite kernels, 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 2. Like the

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 2: Orders of conditional positive definiteness

special univariate spline kernels we shall encounter later, 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 3.3 and 3.4 to see their connection to Hilbert spaces. Note that (unconditionally) positive (semi–) definite kernels are $\mathcal{P}$–conditionally positve (semi–) definite for all finite–dimensional spaces $\mathcal{P}$.

5.3 Interpolation Problems

The standard technique to set up an interpolation problem on a $\mathcal{P}$–unisolvent point set $X = \{x_1, \ldots, x_N\}$ in Ω for a $\mathcal{P}$–conditionally positive kernel K is to use a linear combination

$$s(y) := s_{X,a,b}(y) := \sum_{j=1}^{N} \overline{a_j} K(x_j, y) + \sum_{m=1}^{Q} b_m p_m(y) \text{ for all } y \in \Omega \tag{5.5}$$

using coefficient vectors $a = (a_1, \ldots, a_N)^T \in \mathbb{K}^N$ and $b = (b_1, \ldots, b_Q)^T \in \mathbb{K}^Q$, but with a satisfying the moment conditions (5.4). Interpolation of data $f_1, \ldots, f_N$ in X then poses the $(N+Q) \times (N+Q)$ linear system

$$\begin{array}{rcccl} s(x_k) = \sum_{j=1}^{N} \overline{a_j} K(x_j, x_k) & + & \sum_{m=1}^{Q} b_m p_m(x_k) & = & f_k, \quad 1 \le k \le N, \\ \sum_{j=1}^{N} \overline{a_j} \overline{p_n(x_j)} & + & 0 & = & 0, \quad 1 \le n \le Q \end{array} \tag{5.6}$$

with a Hermitian coefficient matrix.

Theorem 5.7. *If X is $\mathcal{P}$–unisolvent and K is $\mathcal{P}$–conditionally positive definite, the system (5.6) is uniquely solvable.*

Proof: Assume a homogeneous system of the same form and sum the first N equations up with weights a_k. Then

$$\sum_{j,k=1}^{N} \overline{a_j} a_k K(x_j, x_k) + \sum_{m=1}^{Q} b_m \sum_{k=1}^{N} a_k p_m(x_k) = 0,$$

and by $\mathcal{P}$–conditional positive definiteness, the vector a must vanish. But then the first N equations are

$$\sum_{m=1}^{Q} b_m p_m(x_k) = 0, \ 1 \le k \le N,$$

and $\mathcal{P}$–unisolvency of X implies $b = 0$. □

In view of Theorem 2.33 on page 21, we would like to prove that for $\mathcal{P}$–conditional positive **semi**definite kernels, the system is solvable if the data come from a function in the native space, but we have no native space yet. Readers will have to wait for Section 5.7. But there is something simpler:

Theorem 5.8. *If X is $\mathcal{P}$–unisolvent and K is $\mathcal{P}$–conditionally positive semidefinite, the system (5.6) is solvable with $a = 0$ if the data come from functions in $\mathcal{P}$.*

Proof: We set $a = 0$ from the start, and use that, by $\mathcal{P}$–unisolvency of X, each function from $\mathcal{P}$ can be uniquely recovered from its values on X. □

5.4 Inner Product

We now proceed like in Section 3.3, fix Ω, $\mathcal{P}$, and K and define the set

$$M := \left\{(a, X) \,:\, X \subseteq \Omega,\ \mathcal{P}\text{–unisolvent},\ |X| =: N,\ a \in \mathbb{K}^N,\ P_X^T a = 0\right\}$$

of vector/set pairs that satisfy the moment condition (5.4) in the form

$$P_X^T a = 0 \text{ with the matrix } P_X^T = (p_j(x_k))_{1 \le j \le Q,\, 1 \le k \le N}.$$

In particular, we have to assume that Ω has at least one $\mathcal{P}$–unisolvent set in order to let M be nonempty.

Then we define the space of functions

$$H := \{\overline{\lambda_{a,X}^y K(\cdot, y)},\ (a, X) \in M\} \tag{5.9}$$

and the space of functionals

$$L := \{f \mapsto \lambda_{a,X}(f) := \sum_{j=1}^{N} a_j f(x_j) \,:\, (a, X) \in M,\ f \in H\}.$$

It is easy to see that L is a linear space, since we already know this without using the moment conditions, and adding two functionals vanishing on $\mathcal{P}$ will yield a functional vanishing on $\mathcal{P}$. We can argue similarly for H.

We can now follow the pattern of Section 3.3 to define a sesquilinear form (3.12) on L, where we just have to additionally obey the moment conditions. Theorems 3.14 and 3.15 carry over verbatim, but we cannot use functionals $\delta_x = \lambda_{1,x}$ for providing continuous point evaluation, because they are not necessarily in L. We are left with the Riesz map

$$R \,:\, L \to H,\ R(\lambda_{a,X})(y) = \overline{\lambda_{a,X}^x K(y, x)} =: f_{a,X}(y) \tag{5.10}$$

and the identities

$$(\lambda_{a,X}, \lambda_{b,Y})_L = (f_{b,Y}, f_{a,X})_H = \lambda_{a,X}(f_{b,Y}) \tag{5.11}$$

for all (a, X), $(b, Y) \in M$.

Theorem 5.12. *The sum of spaces $\mathcal{P} + H$ is direct, if the kernel K is $\mathcal{P}$–conditionally positive semidefinite.*

Proof: Consider a function $p \in \mathcal{P}$ and a functional $\lambda_{b,Y} \in L$ with

$$\overline{p(x)} = \lambda_{b,Y}^y K(x, y) \text{ for all } x \in \Omega.$$

Then

$$\lambda_{a,X}(p) = 0 = (\lambda_{a,X}, \lambda_{b,Y})_L$$

for all $\lambda_{a,X} \in L$, in particular for $\lambda_{b,Y}$. Thus $\lambda_{b,Y} = 0$ as a functional on H, but $b = 0$ holds only in case of definiteness. By antilinearity of the Riesz map, we conclude in the general case that $f_{b,Y}$ is zero and thus also p. □

5.5 Projections

To see how interpolation works under the inner product of the previous section, we fix a finite $\mathcal{P}$–unisolvent set $X \subseteq \Omega$ and define a subspace

$$H_X := \left\{ \overline{\lambda_{a,X}^y K(\cdot, y)} \ : \ (a, X) \in M, X \text{ fixed} \right\} \tag{5.13}$$

of H for each fixed X.

Theorem 5.14. *The orthogonal complement of H_X in H is*

$$H_X^\perp = \{ f \in H \ : \ \lambda_{a,X}(f) = 0 \text{ for all } (a, X) \in M, X \text{ fixed} \}. \quad \square$$

Now we can define a projector Π_X onto H_X. Note that we let H_X to be finite–dimensional here. Exactly like in Theorem 2.22 on page 18 we get

Theorem 5.15. *For each $f \in H$ and each fixed $\mathcal{P}$–unisolvent set $X \subset \Omega$ there is an interpolant $\Pi_X(f) \in H_X$ with the somewhat nonstandard interpolation conditions*

$$\lambda_{b,X}(f) = \lambda_{b,X}(\Pi_X(f)) \text{ for all } (b, X) \in M.$$

It is the best approximation to f from H_X and attains the minimal norm in H under all functions in H satisfying the same interpolation problem. □

To see the connection to what we did in Section 5.3, we should interpolate an abstract element $f = p + g$ from $\mathcal{P} + H$ on X. Then the above theory applies only to g, and the interpolant in the above sense must have the form

$$s_0(x) = (\Pi_X(g))(x) = \sum_{j=1}^{N} \overline{a_j} K(x_j, x)$$

with a vector a satisfying the moment condition, i.e. $(a, X) \in M$. The interpolation conditions are

$$\sum_{k=1}^{N} b_k s_0(x_k) = \lambda_{b,X}(s_0) = \lambda_{b,X}(g)$$

for all $(b, X) \in M$. Of course, adding functions from $\mathcal{P}$ to either g or s_0 will not change these interpolation conditions, due to the moment conditions.

Theorem 5.16. *Under the above conditions, there is a function $p \in \mathcal{P}$ with*

$$p(x_k) = g(x_k) - (\Pi_X(g))(x_k),\ 1 \le k \le N.$$

Proof: Clearly, $h := g - \Pi_X(g)$ lies in $H_X^\perp$, and we know that $\lambda_{b,X}(h) = 0$ for all $(b, X) \in M$. Splitting $\mathbb{K}^N$ into the subspaces

$$T_X := \{(p(x_1), \ldots, p(x_N))^T \ :\ p \in \mathcal{P}\} \text{ and } T_X^\perp,$$

we see that the vector $(h(x_1), \ldots, h(x_N))^T$ is orthogonal to all $b \in T_X^\perp$, thus in T_X. □.

The data of $f = g + p$ can now be interpolated on X by $s = s_0 + q$, where q interpolates the data of $g - s_0 + p = g - \Pi_X(g) + p$ on X. In fact,

$$\begin{aligned} f(x_k) &= g(x_k) + p(x_k) \\ &= g(x_k) - s_0(x_k) + p(x_k) + s_0(x_k) \\ &= q(x_k) + s_0(x_k) \\ &= s(x_k),\ 1 \le k \le N. \end{aligned}$$

Definition 5.17. *For a $\mathcal{P}$–conditionally positive definite Kernel K, the* **pre–native space** *is*

$$N_K := \mathcal{P} + H$$

where H is the pre–Hilbert space of (5.9) under the inner product defined via (5.11).

Of course, we can extend the inner product on H to a semi–inner product on N_K by defining it to be zero if one of the arguments is in $\mathcal{P}$. This is the first part of

Theorem 5.18. *The pre–native space N_K carries a semi–inner product which vanishes if one of the arguments is in $\mathcal{P}$, and it is an inner product on H. Each function in N_K can be interpolated on any $\mathcal{P}$–unisolvent set X by a function $s_{f,X}$ from $\mathcal{P} + H_X$ with H_X from (5.13). The construction can be performed along the lines of Section 5.3. The function $s_{f,X}$ minimizes the seminorm under all other interpolants from N_K, and it is the best approximation to f in the seminorm from all functions in $\mathcal{P} + H_X$.*

Proof: We only need to prove the final sentence, and we can use Theorem 5.15 for that purpose. In the context of Section 5.3 and the splitting $N_K = \mathcal{P} + H$ we can split $s_{f,X} = p_{f,X} + f_{a,X}$ with $p_{f,X} \in \mathcal{P}$ and $f_{a,X} \in H$. Then $f_{a,X}$ .....
RS: gap here, to be done...

Theorem 5.18 extends Theorem 5.7 to let the system (5.6) be solvable, if the data come from functions in the pre–native space.

5.6 Conditional Lagrange Basis

We now proceed towards a Lagrange–type basis. We need the system (5.6) for this, but we know solvability only if the data on the right–hand side are from a function in the pre–native space, or if the kernel is definite. We shall go for the first case.

Starting with a $\mathcal{P}$–unisolvent set X, we can add another point $y \in \Omega$ and still know that the set $X \cup \{y\}$ is $\mathcal{P}$–unisolvent. This means that we can have a formula

$$(Pu)(y) := \sum_{j=1}^{N} u(x_j) p_j(y) \tag{5.19}$$

which reproduces functions from $\mathcal{P}$, but yields functions in $\mathcal{P}$ if applied to general functions. With this at hand, the functional

$$\mu_y \ : \ f \mapsto f(y) - (Pf)(y) = f(y) - \sum_{j=1}^{N} p_j(y) f(x_j) \tag{5.20}$$

lies in L and the function

$$R(\mu_y) = \overline{\mu_y^z K(\cdot, z)} = K(y, \cdot) - \sum_{j=1}^{N} \overline{p_j(y)} K(x_j, \cdot)$$

lies in H. If we use its values on X as a right–hand side in the system (5.6) and write the coefficients as functions of y, we get

$$\sum_{j=1}^{N} a_j(y)p_n(x_j) \;=\; 0,\ 1 \le n \le Q$$

and

$$\begin{aligned} &\sum_{j=1}^{N} \overline{a_j(y)}K(x_j,x_k) + \sum_{m=1}^{Q} b_m(y)p_m(x_k) \\ =\;& K(y,x_k) - \sum_{j=1}^{N} \overline{p_j(y)}K(x_j,x_k) \end{aligned}$$

which turns into

$$\sum_{j=1}^{N} \overline{u_j(y)}K(x_j,x_k) + \sum_{m=1}^{Q} b_m(y)p_m(x_k) = K(y,x_k),\ 1 \le k \le N.$$

with

$$u_j(y) = a_j(y) + p_j(y),\ 1 \le j \le N$$

and

$$\sum_{j=1}^{N} u_j(y)p(x_j) = \sum_{j=1}^{N} a_j(y)p(x_j) + \sum_{j=1}^{N} p_j(y)p(x_j) = 0 + p(y) \text{ for all } p \in \mathcal{P}.$$

This is usually combined into the system

$$\begin{array}{rcrcll} \displaystyle\sum_{j=1}^{N} K(x_k,x_j)u_j(y) & + & \displaystyle\sum_{m=1}^{Q} \overline{b_m(y)p_m(x_k)} & = & K(x_k,y), & 1 \le k \le N \\ \displaystyle\sum_{j=1}^{N} u_j(y)p(x_j) & + & 0 & = & p(y), & p \in \mathcal{P} \end{array} \tag{5.21}$$

which is solvable and has the same coefficient matrix as (5.6), but a right–hand side that is not necessarily in the native space. If the kernel is definite, the functions u_j are a Lagrange basis, and the functions b_m satisfy $b_m(x_j) = 0,\ 1 \le j \le N,\ 1 \le m \le Q$. This follows from uniqueness, setting $y = x_i$.

Theorem 5.22. *The overall solution of the interpolation for data $f(x_j)$ for a function f from $\mathcal{P} + H$ takes the form*

$$s(y) = \sum_{j=1}^{N} u_j(y)f(x_j).$$

Proof: We write s in the form (5.5) and insert the two above equations into the right–hand side. Then

$$\begin{aligned} s(y) &= \sum_{k=1}^{N} \overline{a_k} K(x_k, y) + \sum_{n=1}^{Q} b_n p_n(y) \\ &= \sum_{k=1}^{N} \overline{a_k} \left(\sum_{j=1}^{N} K(x_k, x_j) u_j(y) + \sum_{m=1}^{Q} \overline{b_m(y) p_m(x_k)} \right) + \sum_{n=1}^{Q} b_n \sum_{j=1}^{N} u_j(y) p_n(x_j) \\ &= \sum_{j=1}^{N} u_j(y) \left(\sum_{k=1}^{N} \overline{a_k} K(x_k, x_j) + \sum_{n=1}^{Q} b_n p_n(x_j) \right) + 0 \\ &= \sum_{j=1}^{N} u_j(y) f(x_j), \ 1 \le j \le N. \quad \square \end{aligned}$$

5.7 Native Space

We now use the technique of the previous section to arrive at reproduction formulae and ar a proper definition of the netive space. Again, X is $\mathcal{P}$–reproducing, and we use the functionals $\mu_y \in L$ of (5.20) applied to functions $f \in H$. This yields

$$\mu_y(f) = (f, R(\mu_y)) = \left(f, K(y, \cdot) - \sum_{j=1}^{N} \overline{p_j(y)} K(x_j, \cdot) \right)_{\mathcal{H}}$$

where we cannot split the inner product, because terms like $(f, K(x, \cdot))_{\mathcal{H}}$ are undefined. For functions $f \in H$ we can split the left–hand side as

$$\mu_y(f) = f(y) - \sum_{j=1}^{N} f(x_j) p_j(y),$$

and this gives us a reproduction formula

$$f(y) - \sum_{j=1}^{N} f(x_j) p_j(y) = \left(f, K(y, \cdot) - \sum_{j=1}^{N} \overline{p_j(y)} K(x_j, \cdot) \right)_{\mathcal{H}} \tag{5.23}$$

that holds on H, while the right–hand term can be dropped for $f \in \mathcal{P}$. Both sides have limits when we go to the completion of H.

But the assignment of single function values to elements of the completion is still to be done. We want to complete H into an abstract Hilbert space $\mathcal{H}$ and define the native space formally as $\mathcal{P} \times \mathcal{H}$ and assign values to a pair (p, f) on Ω consistently, i.e. without dependence on $\mathcal{P}$–unisolvent sets X.

To start with a unique value assignment, we fix a minimal $\mathcal{P}$–unisolvent set $\Xi = \{\xi_1, \ldots, \xi_Q\}$ with exactly Q Lagrange basis functions of $\mathcal{P}$ we call $\pi_1, \ldots, \pi_Q$. Then, given $f \in \mathcal{H}$ and $p \in \mathcal{P}$, we use (5.23) for $X = \Xi$ to assign function values as

$$(p, f)_\Xi(y) := \sum_{m=1}^{Q} p(\xi_m)\pi_m(y) + \left(f, K(y, \cdot) - \sum_{m=1}^{Q} \overline{\pi_m(y)} K(\xi_m, \cdot) \right)_{\mathcal{H}}. \quad (5.24)$$

To show that this definition is consistent with what we have in (5.23), we first note that the inner product above vanishes for $y = \xi_n$, such that there we have $(p, f)_\Xi(\xi_n) = p(\xi_n)$, $1 \le n \le Q$. This shows that the above formula is consistent with (5.23) and makes sense for the completion.

To compare this with the situation on a general set, we denote the values needed in (5.23) by f_X and first evaluate

$$(p, f)_\Xi(x_j) \;=\; \sum_{m=1}^{Q} p(\xi_m)\pi_m(x_j) + \left(f, K(x_j, \cdot) - \sum_{m=1}^{Q} \overline{\pi_m(x_j)} K(\xi_m, \cdot) \right)_{\mathcal{H}}.$$

Then

$$\begin{aligned}
&\sum_{j=1}^{N} p_j(y)(p, f)_\Xi(x_j) \\
=\;& \sum_{j=1}^{N} p_j(y) \sum_{m=1}^{Q} p(\xi_m)\pi_m(x_j) \\
&+ \left(f, \sum_{j=1}^{N} \overline{p_j(y)} \left(K(x_j, \cdot) - \sum_{m=1}^{Q} \overline{\pi_m(x_j)} K(\xi_m, \cdot) \right) \right)_{\mathcal{H}} \\
=\;& \sum_{j=1}^{N} p_j(y) \sum_{m=1}^{Q} p(\xi_m)\pi_m(x_j) \\
&+ \left(f, \sum_{j=1}^{N} \overline{p_j(y)} K(x_j, \cdot) - K(y, \cdot) + K(y, \cdot) - \sum_{m=1}^{Q} K(\xi_m, \cdot)\overline{\pi_m(y)} \right)_{\mathcal{H}} \\
=\;& \sum_{m=1}^{Q} p(\xi_m)\pi_m(y) + (p, f)_\Xi(y) - \sum_{m=1}^{Q} p(\xi_m)\pi_m(y) \\
&+ \sum_{j=1}^{N} p_j(y) f_X(x_j) - f_X(y)
\end{aligned}$$

proves

$$(p, f)_\Xi(y) - \sum_{j=1}^{N} p_j(y)(p, f)_\Xi(x_j) = f_X(y) - \sum_{j=1}^{N} p_j(y) f_X(x_j),$$

i.e. our value assignment is consistent. From (5.24) we also see that

$$(p, f)_\Xi(y) = (p, 0)_\Xi(y) + (0, f)_\Xi(y) \text{ for all } f \in \mathcal{H},\ p \in \mathcal{P},\ y \in \Omega$$

implies that we can form $\mathcal{P} + \mathcal{H}$ in a consistent way. We summarize:

Definition 5.25. *The* **native space** *for a $\mathcal{P}$–conditionally positive semidefinite kernel K can be defined as $\mathcal{P} \times \mathcal{H}$ or $\mathcal{P} + \mathcal{H}$ with a consistent assignment of function values on Ω to make all reproduction equations (5.23) meaningful. A consistent definition of values needs fixing a minimal $\mathcal{P}$–unisolvent set Ξ, and the definition of values will depend on that set. The native space carries a seminorm with kernel $\mathcal{P}$ which is an inner product on $\mathcal{H}$.*

We now look back at Section 5.5 and consider the Hilbert space completion. We now can go over to the completion in the pre–native space $\mathcal{P} + H$, which was impossible until we knew how to add these spaces. Thus we get after short inspection of what we did there,

Theorem 5.26. *The interpolation problem for $\mathcal{P}$–conditionally positive definite kernels on $\mathcal{P}$–unisolvent sets is solvable, if the data come from a function in the native space of the kernel.* □

The extension of Theorems 5.15 and 5.16 to the native space is

Theorem 5.27. *If $s_{f,X}$ is the interpolant to a function f of the native space of a $\mathcal{P}$–conditionally positive semidefinite kernel K on a $\mathcal{P}$–unisolvent set X, then $s_{f,X}$ minimizes the seminorm under all other interpolants from the native space. Furthermore, it provides the best approximation to f from H_X in the seminorm.* □

5.8 Power Function

To generalize the Power Function to the conditionally positive semidefinite case, let K be a $\mathcal{P}$–conditionally positive semidefinite kernel on Ω and let $X = \{x_1, \ldots, x_N\}$ be a $\mathcal{P}$–unisolvent subset of Ω. We consider general reproduction formulae of the form

$$P(f)(x) := \sum_{j=1}^{N} u_j(x) f(x_j) \tag{5.28}$$

which should recover functions from $\mathcal{P}$, i.e. the functionals

$$\mu_x(f) := f(x) - \sum_{j=1}^{N} u_j(x) f(x_j) = \left(\delta_x - \sum_{j=1}^{N} u_j(x) \delta_{x_j} \right) (f)$$

are in L. Then we define the **Power Function** as

$$P_X(x) := \inf_{u(x)\in\mathbb{K}^N} \left\{ \left\| \delta_x - \sum_{j=1}^N u_j(x)\delta_{x_j} \right\|_L \; : \; p(x) = \sum_{j=1}^N u_j(x)p(x_j) \text{ for all } p \in \mathcal{P} \right\}. \tag{5.29}$$

The norms can be explicitly evaluated via

$$\begin{aligned} \|\mu_x\|_L^2 &= \mu_x^u \overline{\mu_x^v K(x,v)} \\ &= \overline{\mu_x^v K(x,v)} - \sum_{j=1}^N u_j(x)\overline{\mu_x^v K(x_j,v)} \\ &= K(x,x) - \sum_{k=1}^N \overline{u_k(x)} K(x_k,x) - \sum_{j=1}^N u_j(x)K(x,x_j) \\ &\quad + \sum_{j=1}^N \sum_{k=1}^N u_j(x)\overline{u_k(x)}K(x_k,x_j) \end{aligned} \tag{5.30}$$

and are upper bounds for $P_X^2(x)$. This will be useful for error bounds, since we have

Theorem 5.31. *If interpolation of a function f in the native space $\mathcal{N}_K$ for a $\mathcal{P}$–conditionally positive semidefinite kernel K on Ω is performed on a $\mathcal{P}$–unisolvent set X and solved by some function s of the form (5.5), then there is an error bound*

$$|f(x) - s(x)| \le P_X(x)\|f\|_{\mathcal{N}_K} \text{ for all } x \in \Omega, \; f \in \mathcal{N}_K. \tag{5.32}$$

For each recovery formula (5.28) which is exact on $\mathcal{P}$, the Power Function has the upper bound

$$P_X^2(x) \le \|\mu_x\|_L^2$$

with (5.30). Equality is attained if the $u_j(x)$ are the Lagrange–type reconstruction functions of Section 5.6 defined by the system (5.21).

Proof: We only have to prove the final statement. For the $u_j(x)$ of (5.21) and the associated functional μ_x we then have to prove that

$$(\mu_x, \lambda_{b,X})_L = 0 \text{ for all } \lambda_{b,X} \in L,$$

because these $\lambda_{b,X}$ are the admissible perturbations. This is

$$\begin{aligned}
(\mu_x, \lambda_{b,X})_L &= \mu_x(f_{b,X}) \\
&= f_{b,X}(x) - \sum_{j=1}^{N} u_j(x) f_{b,X}(x_j) \\
&= \sum_{k=1}^{N} \overline{b_k} K(x_k, x) - \sum_{j=1}^{N} u_j(x) \sum_{k=1}^{N} \overline{b_k} K(x_k, x_j) \\
&= \sum_{k=1}^{N} \overline{b_k} \left(K(x_k, x) - \sum_{j=1}^{N} u_j(x) K(x_k, x_j) \right) \\
&= \sum_{m=1}^{Q} \overline{b_m(y)} \sum_{k=1}^{N} \overline{b_k p_m(x_k)} \\
&= 0
\end{aligned}$$

due to (5.21). □

5.9 Reduced Kernels

Here, we describe a practical trick that allows a simple transition from a $\mathcal{P}$–conditionally positive (semi–) definite kernel to an unconditionally positive (semi–) definite kernel.

Repeating what we needed in the previous section to define function values for the abstract elements of the native space, we fix a minimal $\mathcal{P}$–unisolvent set $\Xi \subset \Omega$ of size $|\Xi| = Q$. Every $p \in \mathcal{P}$ can then be reproduced by a Lagrange basis $\pi_1, \ldots, \pi_Q$ with $\pi_j(\xi_k) = \delta_{jk}$, $1 \leq j, k \leq Q$, i.e. we can without loss of generality assume that

$$p(x) = \sum_{m=1}^{Q} p(\xi_m)\pi_m(x) =: (\Pi_\Xi(p))(x) \text{ for all } x \in \Omega,\ p \in \mathcal{P}$$

after changing to the Lagrange basis. This defines a linear projector Π_Ξ onto P that extends to general functions on Ω as

$$(\Pi_\Xi(f))(x) := \sum_{m=1}^{Q} f(\xi_m)\pi_m(x) \text{ for all } x \in \Omega,\ f : \Omega \to \mathbb{K}. \tag{5.33}$$

This implies that the functionals

$$\mu_x := \delta_x - \sum_{m=1}^{Q} \pi_m(x)\delta_{\xi_m}$$

satisfy the moment conditions, and we define the **reduced kernel**

$$\begin{aligned}\tilde{K}(x,y) &:= (\mu_x,\mu_y)_L \\ &= \mu_y^s\overline{\mu_x^t K(s,t)} \\ &= K(x,y) - \sum_{m=1}^{Q}\overline{\pi_m(x)}K(\xi_m,y) \\ &\quad -\sum_{n=1}^{Q}\pi_n(y)K(x,\xi_n) + \sum_{m=1}^{Q}\sum_{n=1}^{Q}\overline{\pi_m(x)}\pi_n(y)K(\xi_m,\xi_n)\end{aligned} \tag{5.34}$$

for all $x, y \in \Omega$.

Theorem 5.35. *The reduced kernel is Hermitian and unconditionally positive semidefinite on Ω. It vanishes, if one of the arguments is in Ξ. If K is $\mathcal{P}$–conditionally positive definite on Ω, then $\tilde{K}$ is unconditionally positive definite on $\Omega \setminus \Xi$. Quadratic forms with moment conditions will be the same for K and $\tilde{K}$.*

Proof: The first statement follows from the definition via an inner product of functionals. The second follows from $\mu_x = 0$ for $x \in \Xi$. Since the μ functionals annihilate functions in $\mathcal{P}$, quadratic forms for K and $\tilde{K}$ must coincide, if they satisfy moment conditions.

If we have a set $X = \{x_1, \ldots, x_N\} \subseteq \Omega \setminus \Xi$ and a vector $a \in \mathbb{K}^N$, then we have to look at the quadratic form

$$\begin{aligned}&\sum_{j,k=1}^{N}\overline{a_j}a_k\tilde{K}(x_j,x_k) \\ =\ &\sum_{j,k=1}^{N}\overline{a_j}a_k K(x_j,x_k) \\ &-\sum_{k=1}^{N}a_k\sum_{m=1}^{Q}K(\xi_m,x_k)\sum_{j=1}^{N}\overline{a_j}\overline{\pi_m(x_j)} \\ &-\sum_{j=1}^{N}\overline{a_j}\sum_{n=1}^{Q}K(x_j,\xi_n)\sum_{k=1}^{N}a_k\pi_n(x_k) \\ &+\sum_{m=1}^{Q}\sum_{n=1}^{Q}K(\xi_m,\xi_n)\sum_{j=1}^{N}\overline{a_j}\overline{\pi_m(x_j)}\sum_{k=1}^{N}a_k\pi_n(x_k).\end{aligned}$$

If we define

$$\beta_n := -\sum_{k=1}^{N}a_k\pi_n(x_k),\ 1 \le n \le Q,$$

then this turns into

$$\begin{aligned}
&\sum_{j,k=1}^{N} \overline{a_j} a_k \tilde{K}(x_j, x_k) \\
= &\sum_{j,k=1}^{N} \overline{a_j} a_k K(x_j, x_k) + \sum_{k=1}^{N} a_k \sum_{m=1}^{Q} K(\xi_m, x_k)\overline{\beta_m} \\
&+ \sum_{j=1}^{N} \overline{a_j} \sum_{n=1}^{Q} K(x_j, \xi_n)\beta_n + \sum_{m=1}^{Q}\sum_{n=1}^{Q} K(\xi_m, \xi_n)\overline{\beta_m}\beta_n
\end{aligned}$$

which is a Hermitian quadratic form for the set $X \cup \Xi$. Checking moment conditions on this set for $\pi_m \in \mathcal{P}$, we get

$$\begin{aligned}
&\sum_{k=1}^{N} a_k \pi_m(x_k) + \sum_{n=1}^{Q} \beta_n \pi_m(\xi_n) \\
&\sum_{k=1}^{N} a_k \sum_{n=1}^{Q} \pi_m(\xi_n)\pi_n(x_k) - \sum_{n=1}^{Q}\sum_{k=1}^{N} a_k \pi_n(x_k)\pi_m(\xi_n) = 0.
\end{aligned}$$

If K is conditionally positive definite on Ω, and if the form for $\tilde{K}$ on $X \subset \Omega \backslash \Xi$ vanishes for a coefficient vector a, then the form for K vanishes on $X \cup \Xi$ and we get $a = 0$. □

We now can resort to the unconditionally positive semidefinite case and know from Theorem 3.17 on page 47 that $\tilde{K}$ is the reproducing kernel of a Hilbert space $\tilde{\mathcal{H}}$ of functions vanishing on Ξ. Each function $\tilde{f} \in \tilde{\mathcal{H}}$ has the reproduction formula

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

Since the functions of $\tilde{\mathcal{H}}$ vanish on Ξ, and the functions of $\mathcal{P}$ are determined by their values on Ξ, we can add the space $\mathcal{P}$ to $\tilde{\mathcal{H}}$ without overlap. We define the sum

$$\hat{\mathcal{H}} := \mathcal{P} + \tilde{\mathcal{H}}$$

which is direct, and we use the projector Π_Ξ from (5.33). Then

$$f - \Pi_\Xi(f) \in \tilde{\mathcal{H}} \text{ for all } f \in \hat{\mathcal{H}}.$$

Inserting this as $\tilde{f}$ into the reproduction formula in $\tilde{\mathcal{H}}$, we get the Taylor–type reproduction formula

$$f(x) = \Pi_\Xi(f)(x) + (f - \Pi_\Xi(f), \tilde{K}(x, \cdot))_{\tilde{\mathcal{H}}} \text{ for all } f \in \hat{\mathcal{H}},\ x \in \Omega. \qquad (5.36)$$

By setting

$$(f,g)_{\hat{\mathcal{H}}} := (f - \Pi_\Xi(f), g - \Pi_\Xi(g))_{\tilde{\mathcal{H}}} \text{ for all } f,g \in \hat{\mathcal{H}}$$

we get a semi–inner product on $\hat{\mathcal{H}}$ that vanishes if one of the arguments is in $\mathcal{P}$. And we can use

$$\Pi_\Xi(\tilde{K}(x,\cdot)) = 0 \text{ for all } x \in \Omega$$

for

$$(\tilde{K}(x,\cdot),\tilde{K}(y,\cdot))_{\hat{\mathcal{H}}} \;=\; \tilde{K}(x,y) \text{ for all } x,y \in \Omega. \tag{5.37}$$

Theorem 5.38. *The native space of the given $\mathcal{P}$–conditionally positive semidefinite kernel K coincides as a space of functions with $\hat{\mathcal{H}} := \mathcal{P} + \tilde{\mathcal{H}}$, where $\tilde{\mathcal{H}}$ is the native space for the reduced kernel $\tilde{K}$.*

Proof: By Theorem 5.35 we can use $\tilde{K}$ instead of K when we do the construction of sections 5.4 and 5.7, since the quadratic forms will not change. The change is only modulo $\mathcal{P}$. This implies that the resulting space $\mathcal{H}$ will automatically consist of functions vanishing on Ξ, and the equations (5.36) and (5.24) coincide. Thus we get a special instance of value assignment that is consistent with Definition 5.25. □

In Theorem 5.7 of Section 5.3 we saw that we can do interpolation on $\mathcal{P}$–unisolvent sets $X = \{x_1,\ldots,x_N\}$, if the kernel K is $\mathcal{P}$–conditionally positive definite. We did this without using Hilbert space structure, but now we want to link this with our Hilbert space background. Given a function f in $\mathcal{H}$ and a $\mathcal{P}$–unisolvent set X on which we want to interpolate f by a function $s \in \mathcal{H}$, we cannot expect f and s to coincide on Ξ. Formally, we take the data $\mu_{x_j}(f),\ 1 \le j \le N$ and interpolate these data first by some

$$s_0(x) := \sum_{j=1}^{N} \alpha_j \tilde{K}(x_j,x)$$

that vanishes on Ξ. We then have

$$\begin{aligned}
\mu_{x_k}(f) &= f(x_k) - (\Pi_\Xi)(f)(x_k) \\
&= f(x_k) - \sum_{m=1}^{Q} \pi_m(x_k) f(\xi_m) \\
&= \mu_{x_k}(s_0) \\
&= s_0(x_k),\ 1 \le k \le N
\end{aligned}$$

and see that we should use

$$s \;:=\; (\Pi_\Xi)(f) + s_0$$

to get a full interpolant on X.

But we have to check in which function span this interpolant lies. We see that

$$\begin{aligned} s_0(x) &= \sum_{j=1}^{N} \alpha_j \tilde{K}(x_j,x) \\ &= \sum_{j=1}^{N} \alpha_j K(x_j,x) - \sum_{m=1}^{Q} K(\xi_m,x) \sum_{j=1}^{N} \alpha_j \overline{\pi_m(x_j)} \\ &\quad -\sum_{n=1}^{Q} \pi_n(x) \sum_{j=1}^{N} \alpha_j K(x_j,\xi_n) + \sum_{m=1}^{Q}\sum_{n=1}^{Q} \overline{\pi_m(x)} K(\xi_m,\xi_n) \sum_{j=1}^{N} \alpha_j \pi_n(x_j) \end{aligned}$$

still contains functions of the form $K(\xi_m,\cdot)$ if there are no moment conditions with respect to Ξ. This means that we have an interpolant that is possibly different from what we had before.

But if Ξ is a subset of a $\mathcal{P}$–unisolvent data set on which we want to interpolate, we have no problem at all and just perform the previous algorithms in a different way. We use the reduced kernel on $X \setminus \Xi$ for data of $f - \Pi_\Xi(f)$ there, and then add the interpolant in $\mathcal{P}$ to $\Pi_\Xi(f)$, which is $\Pi_\Xi(f)$ itself.

We now show how this can be done in practice. Assume $\Xi \subseteq X$ and define $Y := X \setminus \Xi$ with $|Y| = N - Q$. We sort the points such that the first Q points of X are the points of Ξ. Then we write down the linear system (5.6) in $(Q, N-Q, Q)$ block form as

$$\begin{pmatrix} A_{\Xi,\Xi} & A_{\Xi,Y} & I_{Q\times Q} \\ A_{\Xi,Y}^* & A_{Y,Y} & B \\ I & B^* & 0_{Q\times Q} \end{pmatrix} \begin{pmatrix} \overline{a_\Xi} \\ \overline{a_Y} \\ b \end{pmatrix} = \begin{pmatrix} f_\Xi \\ f_Y \\ 0_{Q\times 1} \end{pmatrix}$$

using the fact that the matrix of values $\pi_m(x_j)$ consists of $I_{Q\times Q}$ for the Ξ part and a $(N-Q)\times Q$ matrix B for the points of Y, since the π_m are a Lagrangian basis on Ξ. We know that the system is solvable, and we use the third and first part to get

$$\begin{aligned} \overline{a_\Xi} &= -B^* \overline{a_Y}, \\ b &= f_\Xi - A_{\Xi,\Xi}\overline{a_\Xi} - A_{\Xi,Y}\overline{a_Y} \\ &= f_\Xi + (A_{\Xi,\Xi}B^* - A_{\Xi,Y})\,\overline{a_Y}. \end{aligned} \tag{5.39}$$

We insert this into the second part. Then

$$\begin{aligned} f_Y &= A_{\Xi,Y}^* \overline{a_\Xi} + A_{Y,Y}\overline{a_Y} + Bb \\ &= -A_{\Xi,Y}^* B^* \overline{a_Y} + \overline{a_Y} + B\left(f_\Xi + (A_{\Xi,\Xi}B^* - A_{\Xi,Y})\,\overline{a_Y}\right), \end{aligned}$$

gives the solvable system

$$f_Y - Bf_\Xi = \left(A_{Y,Y} - A^*_{\Xi,Y}B^* - BA_{\Xi,Y} + BA_{\Xi,\Xi}B^*\right)\overline{a_Y}.$$

Once this is solved, we use (5.39) to get the remaining parts of the solution. Writing the above system in detail, we get

$$\begin{aligned} & f(x_k) - \sum_{m=1}^{Q} \pi_m(x_k) f(\xi_m) \\ = & \sum_{j=Q+1}^{N} \left(K(x_j,x_k) - \sum_{m=1}^{Q} K(\xi_m,x_k)\overline{\pi_m(x_j)} - \sum_{n=1}^{Q} \pi_n(x_j) K(x_k,\xi_n) \right. \\ & \left. + \sum_{m,n=1}^{Q} \pi_n(x_j)\overline{\pi_m(x_j)} K(\xi_m,\xi_n) \right) \overline{a_j} \end{aligned}$$

for $Q+1 \le k \le N$. This is

$$\mu_{x_k} = \sum_{j=Q+1}^{N} \tilde{K}(x_j,x_k)\overline{a_j},\ Q+1 \le k \le N,$$

i.e. a reduced system using the reduced kernel.

Readers should note that this is a variation of a Schur complement argument.

For applications, we need to know which functionals can be used for generalized interpolation, provided that we have a $\mathcal{P}$–conditionally positive semidefinite kernel K that is explicitly known as a function on $\Omega \times \Omega$. This will amount to generalize Theorem 2.11 to the $\mathcal{P}$–conditionally positive semidefinite case. We can avoid a new theorem by applying Theorem 2.11 to the reduced kernel.

5.10 Extended Kernels

The previous sections showed that in the $\mathcal{P}$–conditionally positive (semi)–definite case we have a direct sum $\mathcal{P} + \mathcal{H}$ as a native space, with an inner product only on the $\mathcal{H}$ part. But in order to arrive at an inner product on the whole space, we can extend the reduced kernel (5.34) to get the **extended kernel**

$$\begin{aligned} K^\dagger(x,y) &:= \tilde{K}(x,y) + \sum_{m=1}^{Q} \overline{\pi_m(x)}\pi_m(y) \\ &= K(x,y) + \sum_{m=1}^{Q} \overline{\pi_m(x)}\pi_m(y) - \sum_{m=1}^{Q} \overline{\pi_m(x)} K(\xi_m,y) \\ &\quad - \sum_{n=1}^{Q} \pi_n(y) K(x,\xi_n) + \sum_{m=1}^{Q}\sum_{n=1}^{Q} \overline{\pi_m(x)}\pi_n(y) K(\xi_m,\xi_n) \end{aligned} \tag{5.40}$$

for all $x, y \in \Omega$, using the Lagrange basis $\pi_1, \ldots, \pi_Q$ of $\mathcal{P}$ again. The additional part clearly is an unconditionally positive semidefinite kernel itself. Regarding the results of Section 5.7, we here fix $X = \Xi$ like in Section 5.9 and get

Theorem 5.41. *If K is a $\mathcal{P}$–conditionally positive semidefinite kernel on Ω, the extended kernel $K^\dagger$ is an unconditionally positive semidefinite kernel on Ω whose native space $\mathcal{H}^\dagger$ coincides with the native space of K as a vector space. The subspace*

$$\mathcal{H}^\dagger_\Xi := \{f \in \mathcal{H}^\dagger \ : \ f(\Xi) = \{0\}\}$$

is isometrically isomorphic to $\mathcal{H}$ and orthogonal to $\mathcal{P}$ in the new inner product induced by $K^\dagger$. If K is definite, so is $K^\dagger$.

Proof: The kernel $K^\dagger$ clearly is Hermitian and unconditionally positive semidefinite on Ω due to its definition and Theorem 5.35. We note in passing that

$$K^\dagger(x, \xi_m) = \overline{\pi_m(x)}, \ K^\dagger(\xi_m, y) = \pi_m(y) \text{ for all } 1 \le m \le Q, \ x, y \in \Omega.$$

The projector Π_Ξ onto $\mathcal{P}$ is (5.33), and thus

$$\begin{aligned} K^\dagger(x,y) - \Pi_\Xi(K^\dagger(x,\cdot))(y) &= K^\dagger(x,y) - \sum_{m=1}^{Q} \pi_m(y) K^\dagger(x,\xi_m) \\ &= K^\dagger(x,y) - \sum_{m=1}^{Q} \overline{\pi_m(x)} \pi_m(y) \\ &= \tilde{K}(x,y) \text{ for all } x, y \in \Omega. \end{aligned}$$

The translates $K^\dagger(x,\cdot)$ are in $\mathcal{P} + H$, and to each translate $\tilde{K}(x,\cdot)$ we can add an element of $\mathcal{P}$ to get $K^\dagger(x,\cdot)$. Thus, without going to closures, we now work in $\mathcal{P} + H$ without loss of generalization, unless we go to closures.

We first assert that the inner product $(.,.)_\dagger$ on the native space of $K^\dagger$, necessarily satisfying (2.3), is the same as

$$(f,g)_{\dagger\dagger} := \sum_{m=1}^{Q} f(\xi_m)\overline{g(\xi_m)} + (f - \Pi_\Xi(f), g - \Pi_\Xi(g))_{\mathcal{H}} \text{ for all } f, g \in \mathcal{P} + H.$$

We only have to check this on translates of $K^\dagger$. This is

$$\begin{aligned}
& (K^\dagger(x,\cdot), K^\dagger(y,\cdot))_{\dagger\dagger} \\
= & \sum_{m=1}^{Q} K^\dagger(x,\xi_m)\overline{K^\dagger(y,\xi_m)} \\
& + \Big(K^\dagger(x,\cdot) - \Pi_\Xi(K^\dagger(x,\cdot)), K^\dagger(y,\cdot) - \Pi_\Xi(K^\dagger(y,\cdot))\Big)_{\mathcal{H}} \\
= & \sum_{m=1}^{Q} \overline{\pi_m(x)}\pi_m(y) + (\tilde{K}(x,\cdot), \tilde{K}(y,\cdot))_{\mathcal{H}} \\
= & K^\dagger(x,y) - \tilde{K}(x,y) + (\tilde{K}(x,\cdot), \tilde{K}(y,\cdot))_{\mathcal{H}} \\
= & K^\dagger(x,y)
\end{aligned}$$

due to (5.37). In this inner product, the spaces $\mathcal{P}$ and H are orthogonal. Furthermore, we have the reproduction equation

$$\begin{aligned}
f(x) &= (\Pi_\Xi(f))(x) + \Big(f - \Pi_\Xi(f), \tilde{K}(x,\cdot)\Big)_{\mathcal{H}} \\
&= \sum_{m=1}^{Q} f(\xi_m)\pi_m(x) + \Big(f - \Pi_\Xi(f), \tilde{K}(x,\cdot)\Big)_{\mathcal{H}} \\
&= \sum_{m=1}^{Q} f(\xi_m)\overline{K^\dagger(x,\xi_m)} + \Big(f - \Pi_\Xi(f), K^\dagger(x,\cdot) - \Pi_\Xi(K^\dagger(x,\cdot))\Big)_{\mathcal{H}} \\
&= (f, K^\dagger(x,\cdot))_{\dagger\dagger} \\
&= (f, K^\dagger(x,\cdot))_\dagger \text{ for all } f \in \mathcal{P} + H,\ x \in \Omega.
\end{aligned}$$

Going to the closure now is no problem at all. Since it is clear that

$$(f,g)_{\mathcal{H}} = (f,g)_\dagger$$

for all f, g in $\mathcal{P} + \mathcal{H}$ that vanish on Ξ, the Hilbert space topologies on $\mathcal{H}$ for the reduced kernel and on the subspace $\mathcal{H}_\Xi^\dagger$ of $\mathcal{P} + \mathcal{H}$ for the extended kernel are isometric. We now are exactly in the situation of Theorem 2.19 on page 17. The space $\mathcal{P}$ is orthogonal to $\mathcal{H}$ in the new inner product, and the extended kernel reflects this orthogonal decomposition. □

Using the extended kernel, one can bypass all the hassles induced by conditional positive definiteness, when it comes to numerical calculations. But everything will depend on Ξ, and if interpolations on different $\mathcal{P}$–unisolvent sets X are performed and compared, one must keep Ξ fixed throughout. However, results will be different from what we did in Section 5.3, if Ξ is not contained in the set X used for interpolation. The reason is that translates of $\tilde{K}$ and $K^\dagger$ will usually contain translates at the points of Ξ.

6 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 $\mathcal{P}_k$. In the d–variate case we shall use the notation $\mathcal{P}_k^d$.

6.1 Semi–inner product

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

$$|f|_k^2 := \int_a^b \left(\frac{d^k f(t)}{dt^k}\right)^2 dt \tag{6.1}$$

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

Equation (6.1) 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 6.2. *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. □

6.2 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**

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

the above equation takes the form

$$\begin{aligned} f(x) &= \underbrace{\sum_{j=0}^{k-1} \frac{f^{(j)}(a)}{j!}(x-a)^j}_{=:(P_{k,a}f)(x)} + (f, K_{k,a}(x,\cdot))_k \\ &= (P_{k,a}f)(x) + (f, K_{k,a}(x,\cdot))_k,\ x \in [a,b]. \end{aligned} \tag{6.3}$$

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.

6.3 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

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

with

$$K_{k,b}(x,y) = (-1)^k \frac{(y-x)_+^{2k-1}}{(2k-1)!}.$$

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

$$\begin{aligned} f(x) &= \tfrac{1}{2}(P_{k,a}f)(x) + \tfrac{1}{2}(P_{k,b}f)(x) + \tfrac{1}{2}(f, K_{k,a}(x,\cdot) + K_{k,b}(x,\cdot))_k \\ &=: (R_k f)(x) + (f, K_k(x,\cdot))_k \end{aligned} \tag{6.4}$$

with

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

Note that the two reproduction formulae (6.3) and (6.4) can both be used to our convenience. The different kernels are linked to different polynomial projectors.

We have three reproduction formulas and three kernels, so far. But we also want to have (2.3), and this will not hold for either of these kernel. We boldly define

$$\Phi_k(x,y) := (K_k(x,\cdot), K_k(y,\cdot))_k \text{ for all } x,y \in [a,b]$$

and see what this new symmetric kernel is. We use (6.4) for $f(y) := K_k(x,y)$ to find

$$\begin{aligned} & \Phi_k(x,y) \\ = \ & (K_k(x,\cdot), K_k(y,\cdot))_k \\ = \ & K_k(x,y) - R_k(K_k(x,\cdot))(y) \\ = \ & K_k(x,y) - \tfrac{1}{2}(P_{k,a}K_k(x,\cdot))(y) - \tfrac{1}{2}(P_{k,b}K_k(x,\cdot))(y) \\ = \ & K_k(x,y) - \frac{1}{2}(-1)^k \sum_{j=0}^{k-1} \frac{(x-a)^{2k-1-j}}{(2k-1-j)!}(-1)^j (y-a)^j \\ & -\frac{1}{2}(-1)^k \sum_{j=0}^{k-1} \frac{(b-x)^{2k-1-j}}{(2k-1-j)!}(y-b)^j. \end{aligned}$$

Since both Φ_k and K_k are symmetric, so is the polynomial on the right–hand side. This means that it must be a polynomial of degree at most $k-1$ in **both** x and y by some hidden cancellation. But as these kernels differ only by functions in $\mathcal{P}_k$, we have

$$\Phi_k(x,y) = (\Phi_k(x,\cdot),\Phi_k(y,\cdot))_k \text{ for all } x,y \in [a,b] \tag{6.5}$$

and the reproduction equation

$$f(x) =: (R_k f)(x) + (f,\Phi_k(x,\cdot))_k \text{ for all } f \in \mathcal{C}^k[a,b],\ x \in [a,b]. \tag{6.6}$$

To illustrate the hidden cancellation, we consider $k=1$. Then

$$\begin{aligned} & \Phi_1(x,y) \\ = \ & K_1(x,y) + \frac{1}{2}(x-a) + \frac{1}{2}(b-x) \\ = \ & -\tfrac{1}{2}|x-y| + \tfrac{a+b}{2}. \end{aligned}$$

Readers might check the cancellation for $k=2$.

6.4 Conditional Positive Definiteness

As readers will already expect, we have

Theorem 6.7. *The kernel Φ_k is unconditionally positive semidefinite. All spline kernels we have considered so far are $\mathcal{P}_k$–conditionally positive definite.*

Proof: The first statement follows immediately from (6.5), but note that it does not hold for the other kernels. For the conditional positive definiteness, we take the kernel Φ_k, a $\mathcal{P}_k$–unisolvent set X and a coefficient vector a with moment conditions, and then we have to prove that if the function

$$f(t) := \sum_{j=1}^{N} a_j \Phi_k(x_j,t) \tag{6.8}$$

is in $\mathcal{P}_k$, then all coefficients vanish. For these coefficients, we look at

$$\begin{aligned} \lambda_{a,X}(v) &= \sum_{j=1}^{N} a_j v(x_j) \\ &= \sum_{j=1}^{N} a_j (v(x_j) - (R_k v)(x_j)) \\ &= \left(v, \sum_{j=1}^{N} a_j \Phi_k(x_j,\cdot) \right)_k \\ &= 0 \end{aligned}$$

for all functions v that we can insert into the Taylor formula. By choosing a Lagrange polynomial basis for interpolation on X, we get that all coefficients must vanish. This finishes the proof for Φ_k, but the same argument works for all other combinations of kernels and projectors that we have seen so far and that lead to a Taylor formula. □

6.5 Native Space for Spline Kernels

Theorem 6.7 allows for two paths towards a native space:

1. for Φ_k as an unconditionally positive semidefinite kernel,
2. for Φ_k or other kernels as $\mathcal{P}_k$–conditionally positive definite kernels.

We shall see that we get different results. If we use Φ_k as an unconditionally positive semidefinite kernel, the native space will consist of the closure of translates $\Phi_k(x,\cdot)$ under an inner product that also allows (6.5). But then this inner product, since it coincides with $(.,.)_k$ on the "generators" $\Phi_k(x,\cdot)$ must be identical to $(.,.)_k$. This is puzzling at first sight, because in general $(.,.)_k$ vanishes on $\mathcal{P}_k$, i.e. it is not positive definite. But if a linear combination (6.8) without moment conditions is in $\mathcal{P}_k$, we can plug it into (6.6) to get

$$(f,\Phi_k(x,\cdot))_k = \sum_{j=1}^{N} a_j \left(\Phi_k(x_j,\cdot),\Phi_k(x,\cdot)\right)_k = \sum_{j=1}^{N} a_j \Phi_k(x_j,x) = f(x) = 0$$

for all $x \in \Omega$, proving that $(.,.)_k$ is positive definite on the span of the translates of Φ_k. If we take another look at (6.5) and (6.6), we see that all translates of Φ_k necessarily are in the kernel of the projector R_k, and we have the standard reproduction formula

$$f(x) = (f,\Phi_k(x,\cdot))_k$$

for all f in the native space $\mathcal{H}$ generated by the translates of Φ_k, as is to be expected. But this space consists of functions f with $R_k(f) = 0$, i.e. it is not exactly what we want, since the space $\mathcal{P}_k$ has fallen out. Of course, we could get it back in by going over to an extended kernel, but this would yield another native space.

We now check what we get if we consider Φ_k as a $\mathcal{P}_k$–conditionally positive definite kernel. The native space will then consist of the direct sum of $\mathcal{P}_k$ with a Hilbert space $\mathcal{H}$ generated by functions

$$f_{a,X}(y) := \sum_{j=1}^{N} a_j \Phi_k(x,y)$$

where the coefficients satisfy moment conditions and X is $\mathcal{P}_k$–unisolvent. The inner product on $\mathcal{H}$ is

$$\begin{aligned}(f_{a,X}, f_{b,Y})_{\mathcal{H}} &= \sum_{j=1}^{N}\sum_{i=1}^{M} a_i b_j \Phi_k(x_i, y_j) \\ &= \sum_{j=1}^{N}\sum_{i=1}^{M} a_i b_j (\Phi_k(x_i, \cdot), \Phi_k(y_j, \cdot))_k \\ &= (f_{a,X}, f_{b,Y})_k\end{aligned}$$

by (6.5), and it is positive definite there, as we know. Clearly, all $f_{a,X}$ are in the **Beppo–Levi space**

$$BL^k[a,b] := \{f \ : \ [a,b] \to \mathbb{R} \ : \ f^{(k)} \in L_2[a,b]\}$$

because their k–th derivatives are piecewise continuous or even smoother, and they lie in the subspace of the f with $R_k(f) = 0$. This is clear, because for each function $f \in BL^k[a,b]$ we have the reproduction formula (6.6) as an instance of a symmetrized Taylor formula. The semi–inner product $(.,.)_k$ is positive definite on that subspace, and thus we can decompose $BL^k[a,b]$ into $BL^k[a,b] = \mathcal{P}_k + \ker R_k$ with $\ker R_k$ being a Hilbert space isometrically isomorphic to $L_2[a,b]$ via the map $f \mapsto f^{(k)}$ from $BL^k[a,b]$ to $L_2[a,b]$.

We now assert

Theorem 6.9. *The native space for Φ_k as a $\mathcal{P}_k$–conditionally positive definite kernel is the Beppo–Levi space $BL^k[a,b]$.*

Proof: We know that all $f_{a,X}$ are in $\ker R_k$, and thus we only need to prove that the closure of these functions under $(.,.)_k$ is not larger than $\ker R_k$. Assume that some $g \in \ker R_k$ is orthogonal to all $f_{a,X}$. This implies

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

and we use the argument at the start of the proof of Theorem 5.16 on page 79 to conclude that the vector of values $g(x_1), \ldots, g(x_N)$ can be viewed as a

vector of values of a polynomial in $\mathcal{P}_k$. Since this holds for all $\mathcal{P}_k$–unisolvent sets X, the function g must itself be a polynomial, and vanish since it is $\ker R_k$. □

We now can apply everything we know from Chapter 5 about conditionally positive definite kernels. The interpolation systems of Section 5.3 are nonsingular, and we have the same optimality results.

7 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.

7.1 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.
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. [ADR94, CFB97, NFN00, CBC+01, OBS03, RTSD03, WK05, BK05]. In particular, one can use such transformations to couple incompatible finite element codes [ABW06].
Furthermore, interpolation of functions has quite some impact on methods solving partial differential equations.

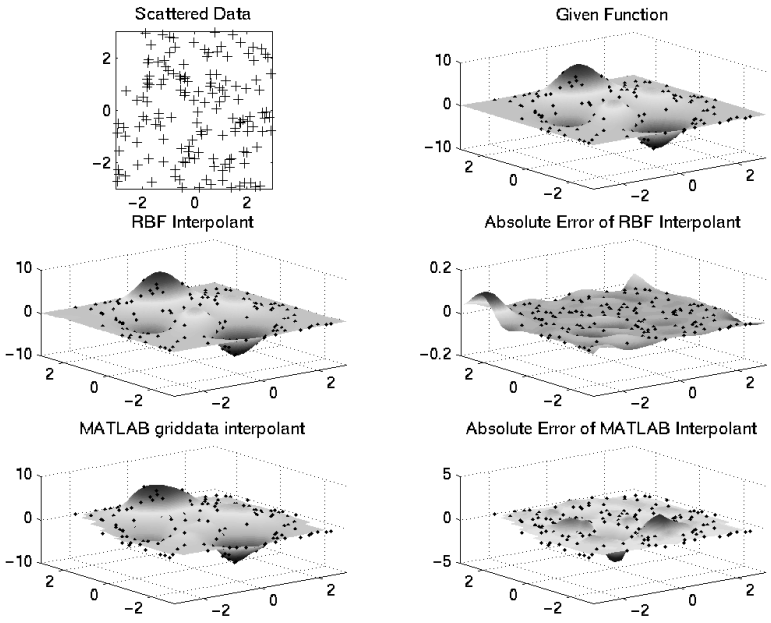

Figure 10: Interpolation by radial basis functions

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*" [BKO+96]. Since kernel interpolants approximate higher derivatives well, local function values can be used to provide good estimates for derivative data [WHW05]. This has connections to **p**seudospectral methods [Fas06].

7.2 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 [Wu92], is to use special trial functions

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

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 (3.1) 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 (3.3) turns into

$$\lambda_j u = \sum_{k=1}^{n} a_k \lambda_k^t \lambda_j^x \phi(\|x - t\|_2),\ 1 \le j \le n \tag{7.1}$$

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.

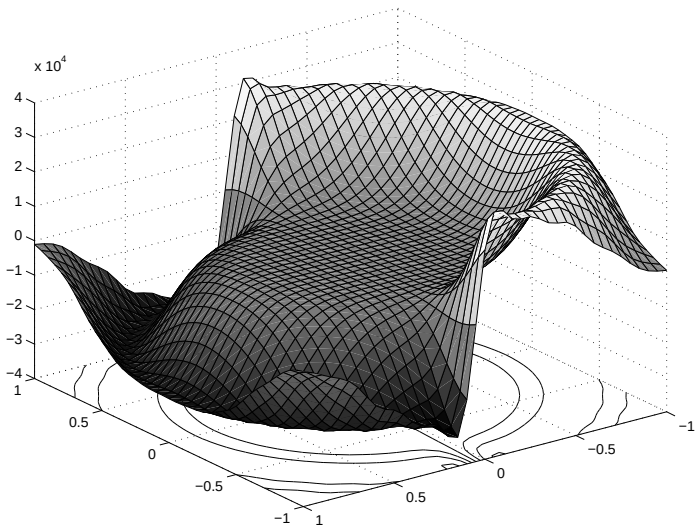

Figure 11: Generalized interpolant to Neumann data

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 (??) or (??) 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 (5.4) 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 (5.4) and $\mathcal{P}$–unisolvency.
Another example of recovery from non-Lagrange data is the construction of **Lyapounov basins** from data consisting of **orbital derivatives** [GW07].
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.

7.3 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 [Buh03, Wen05] and recent papers [NWW06]. They are completely *local*, and they are in terms of the **fill distance**

$$h := h(X, \Omega) := \sup_{y \in \Omega} \min_{x \in X} \|x - y\|_2 \tag{7.2}$$

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 [MN92, Yoo01] 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. This is particularly important when applications require good reproductions of derivatives, e.g. velocity fields or stress tensors.

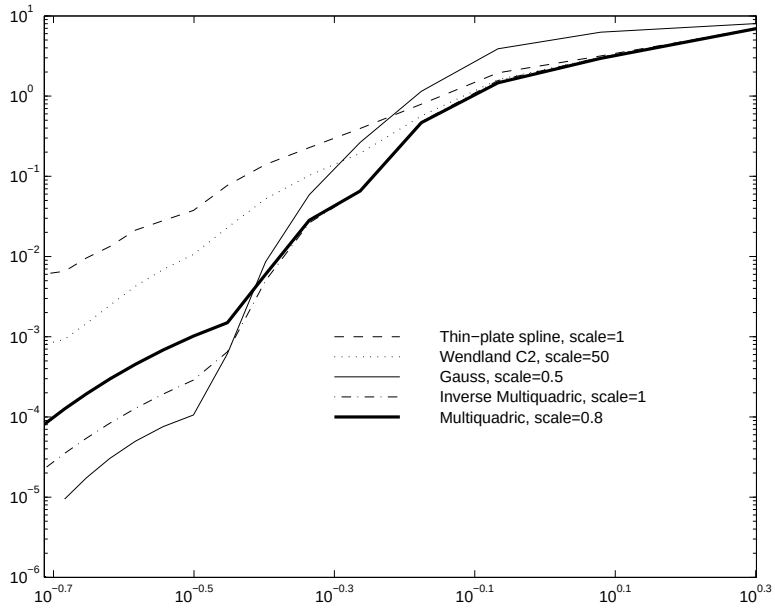

Figure 12: Nonstationary interpolation to a smooth function as a function of fill distance

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

$$f(x, y) = 0.03 * \max(0, 6 - x^2 - y^2)^2, \tag{7.3}$$

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.

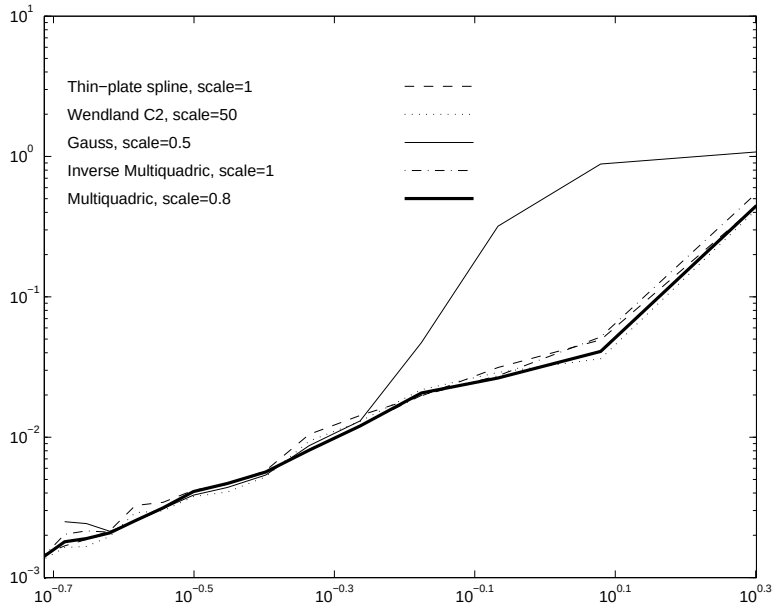

Figure 13: Nonstationary interpolation to a nonsmooth function as a function of fill distance

7.4 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 (3.3) and (7.1) 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 (7.2). 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,

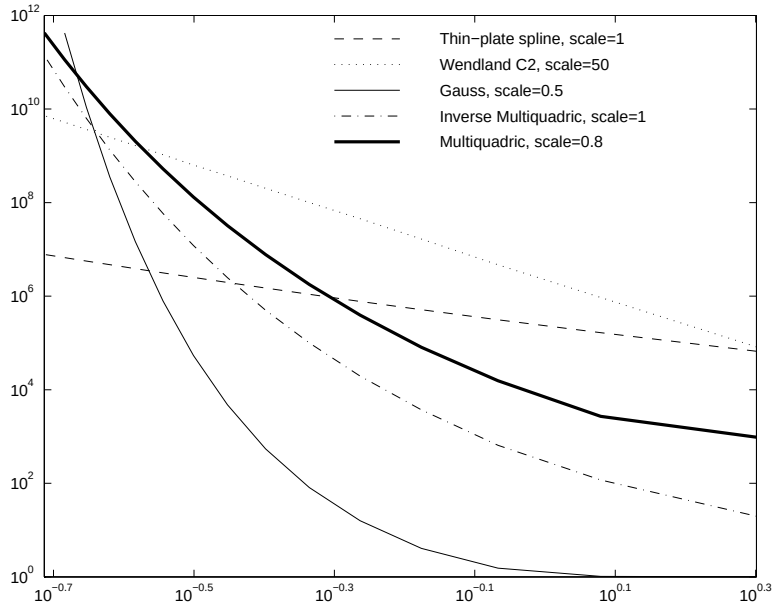

Figure 14: Condition as function of separation distance

but with at least two points of X on the boundary. Thus for convex domains 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

$$\gamma \, h(X, \Omega) \leq S(X) \leq h(X, \Omega). \tag{7.4}$$

To maintain quasi-uniformity, it suffices in most cases to delete "duplicates". Furthermore, there are sophisticated "thinning" techniques [FI98, DDFI05, WR05] to keep fill and separation distance proportional, i.e. to assure quasi-uniformity at multiple scaling levels.

7.5 Uncertainty Principle

Unless radial basis functions are rescaled in a data-dependent way, it can be proven [Sch95] 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.

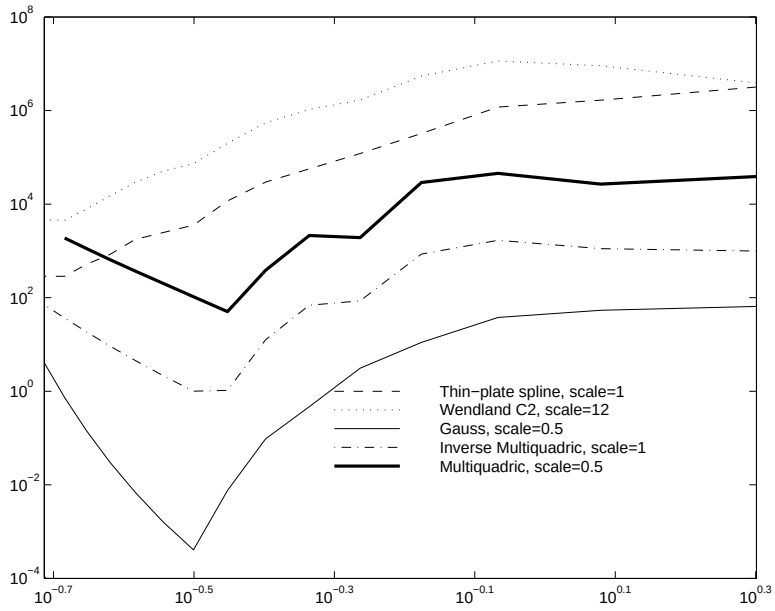

Figure 15: Squared L_∞ error times condition as a function of fill distance

We illustrate this by an example. Since [Sch95] 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, 22, 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 reproduction quality, smooth radial basis functions can well recover nonsmooth functions, as proven by papers concerning error bounds [NWW06]. On the other hand, non-smooth radial basis functions will not achieve high convergence rates when approximating smooth functions [SW02]. 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 7.6.

7.6 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** [DF02]. 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.

$$\phi_c(\|x\|_2) := \sqrt{c^2 + \|x\|_2^2} = c \cdot \sqrt{1 + \frac{\|x\|_2^2}{c^2}} = c \cdot \phi_1\left(\frac{\|x\|_2}{c}\right) \tag{7.5}$$

for multiquadrics, because the outer factor c is irrelevant when forming trial spaces from functions (3.1). 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 (3.3), but the error behavior is disastrous, because there is no reproduction quality between the spikes. The opposite case of 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 [DF02, Sch05, LF05, LYY05, Sch06a]. 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

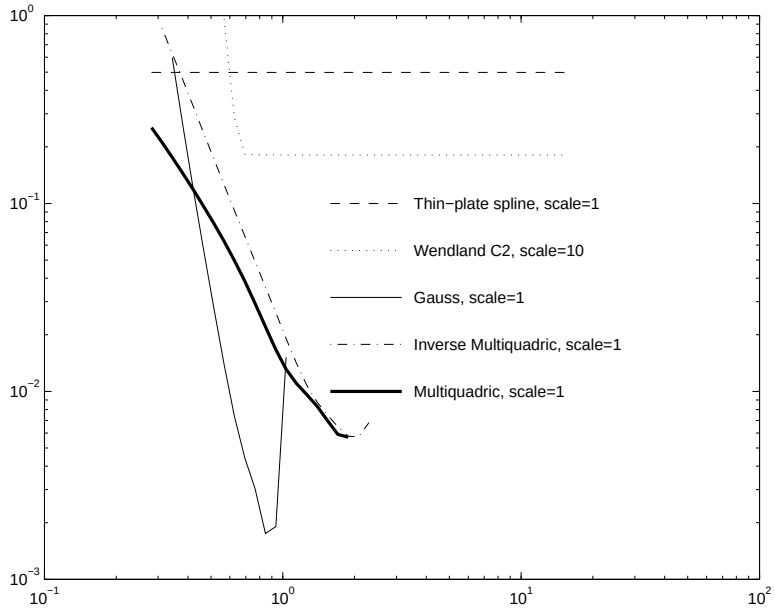

Figure 16: Error as function of relative scale, smooth case

techniques [LF03, FW04] 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 (7.3) 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 [LF03, FW04].

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 (7.2). This is often called a **stationary** setting, e.g. in the context of wavelets and quasi-interpolation.

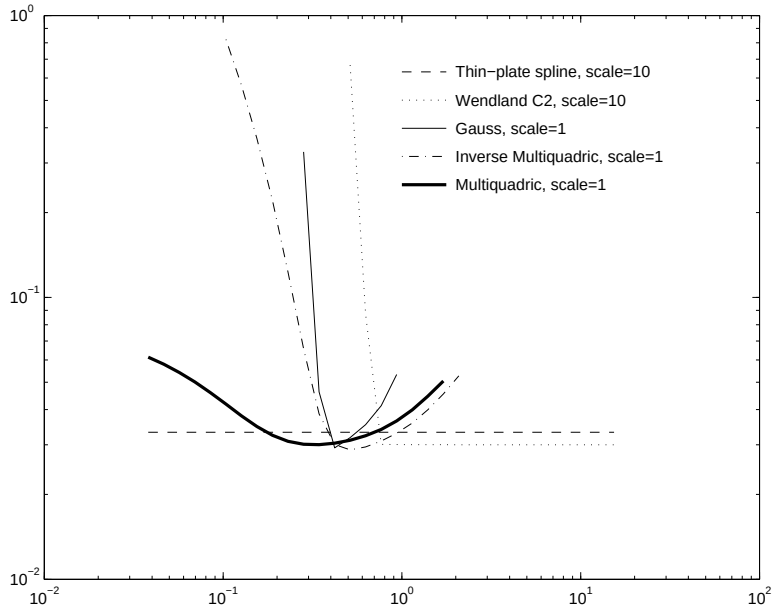

Figure 17: Error as function of relative scale, nonsmooth case

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 [Buh88, Buh90]. 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 [Sch97a], and for special situations there is a sound theoretical analysis called **approximate approximation** [MS96]. 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,

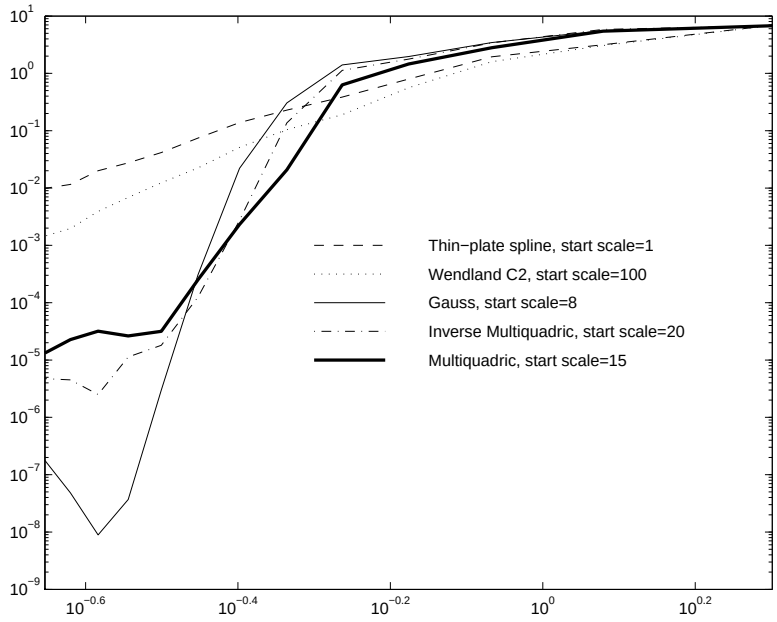

Figure 18: Stationary interpolation to a smooth function at small starting scales

as shown in Figure 17, but this is an experimental observation which still is without proper theoretical explanation.

7.7 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.
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.

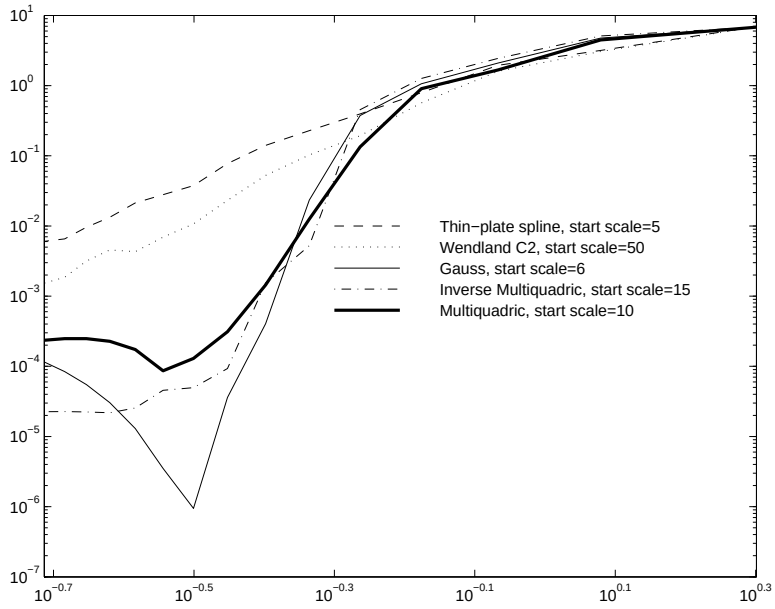

Figure 19: Stationary interpolation to a smooth function at wider starting scales

- 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.

- 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.

7.8 Sensitivity to Noise

So far, the discussion focused on noiseless data, with the exception of Figure 23. 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 (3.3) 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 [Sch06b, Sch07].
A third possibility is the old Levenberg-Marquardt trick of adding a positive value λ into the diagonal of the kernel matrix of (3.3) with entries $\phi(\|x_j - x_k\|_2)$. As is well-known from 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 [WR05], because it is able to shift the stability from dependence on the separation distance to dependence on the fill distance (see section 7.4).

A fourth possibility is **regularization**, for example using a singular-value decomposition as described in section 7.10.
In general, one can replace the system (3.3) 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 (3.1) or (**??**), the quadratic form

$$\|u\|_\phi^2 := \sum_{j,k=1}^{n} \alpha_j \alpha_k \phi(\|x_j - x_k\|_2) \tag{7.6}$$

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 [Buh03, Wen05] 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** [HW03, Li04, TWN04, CC05b, CC05a, HW05, JZ05, Li05, Sai05, Nas06]. The above strategy of minimizing the quadratic form (7.6) also is central for modern methods of **machine learning**, but we cannot pursue this subject in detail here [CST00, SS02, STC04].
Let us use minimization of the quadratic form (7.6) 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 (7.6) under the constraints

$$-\epsilon \le \sum_{j=1}^{n} \alpha_j \phi(\|x_j - x_k\|_2) - f(x_k) \le \epsilon,\ 1 \le k \le n \tag{7.7}$$

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

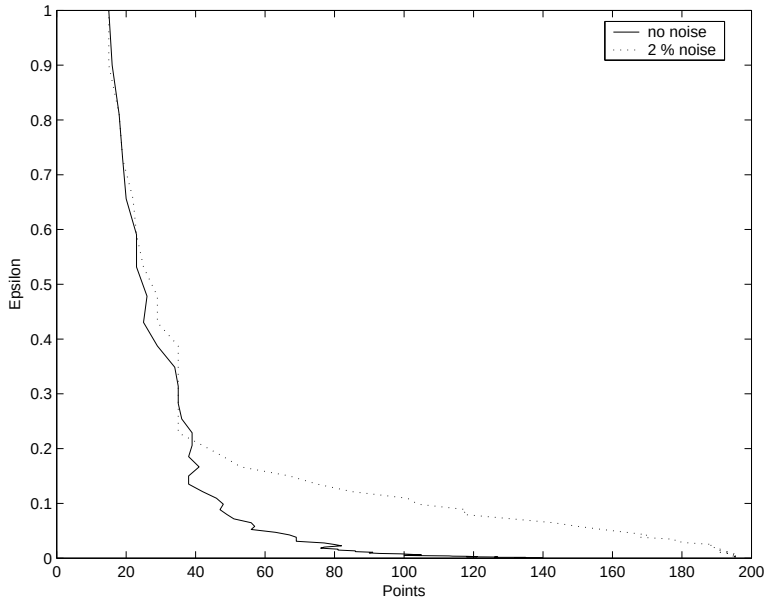

Figure 20: Connection between ϵ and the number $n(\epsilon)$ of necessary points

conditions imply that only those points x_k are actually used where one of the bounds in (7.7) 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 [HO93].
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.

.... incomplete...

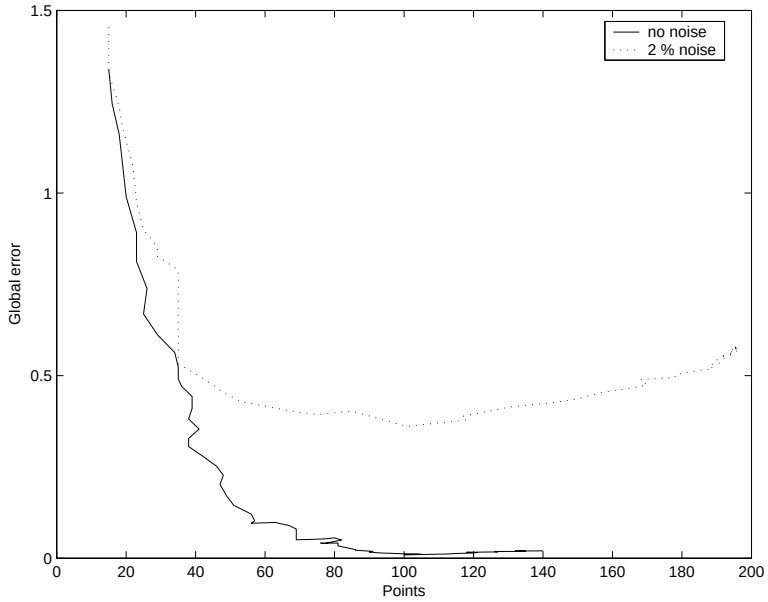

Figure 21: Error $E(\epsilon)$ as a function of the number $n(\epsilon)$ of necessary points

Demos on power functions and on point selection

.... incomplete...

7.9 Calculation

We now want to take a closer look at the systems (**??**) or (**??**). 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 (**??**) as minimization of a quadratic form. The latter is

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

and can be rewritten as

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

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

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

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 (**??**) and the Lagrange type formula (**??**). 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 (??) 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 (??) with the $u_j^*(x)$ constructed as above. This will not necessarily interpolate the data, but probably be a good reconstruction strategy anyway.

7.10 Regularization

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

$$Ax = b \in \mathbb{R}^m \tag{7.8}$$

which is to be solved for a vector $x \in \mathbb{R}^n$. The system may arise from any method using kernels, including (??) and (??), 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 [GvL96] should do the job, i.e. produce a numerical solution $\tilde{x}$ which solves

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 \tag{7.9}$$

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** [GvL96], the matrix A can be decomposed into

$$A = U\Sigma V^T \tag{7.10}$$

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 $\wr(mn^2 + nm^2)$ 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 (7.10), the solution of either the minimization problem (7.9) or, in the case $m = n$, the solution of (7.8) can be obtained and analyzed as follows. We first introduce new vectors

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

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

$$\sum_{j=1}^{n}(\sigma_j\hat{y}_j - c_j)^2 + \sum_{j=n+1}^{m} c_j^2 \leq \eta^2. \tag{7.11}$$

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

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

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

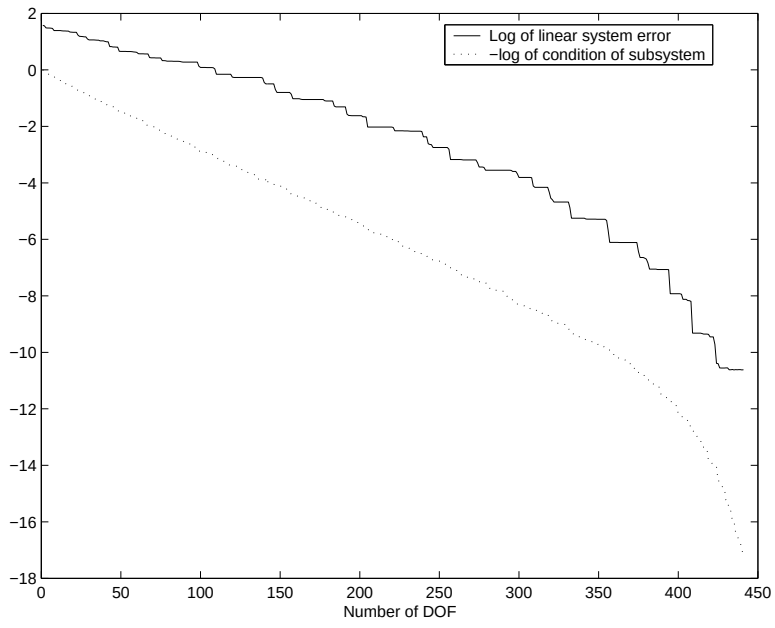

Figure 22: Error and condition of linear subsystems via SVD

Figure 22 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 (**??**). 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 $1 \leq k \leq 441$ show that there are low-rank subsystems which already provide good approximate solutions.

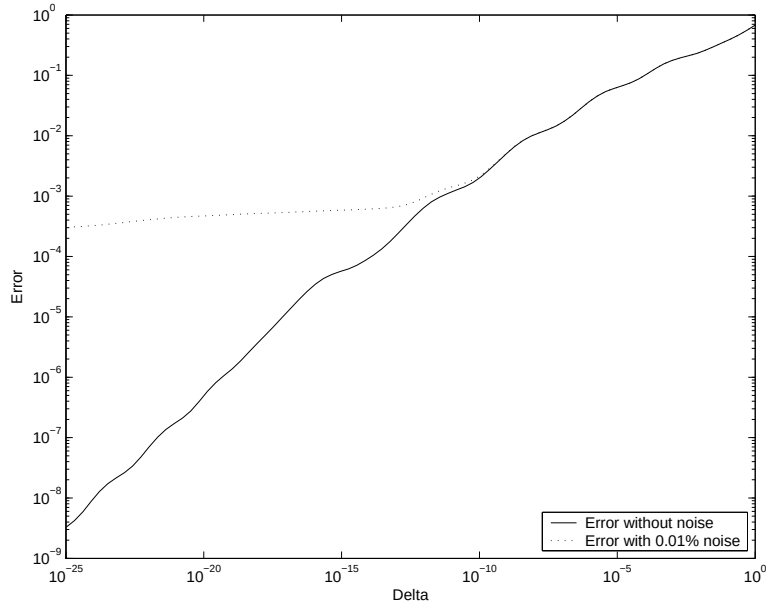

Figure 23: Error as function of regularization parameter δ^2

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 (7.11), 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

$$\sum_{j=1}^{n} (\sigma_j y_j - c_j)^2 + \sum_{j=n+1}^{m} c_j^2 \leq \eta^2, \tag{7.12}$$

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

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

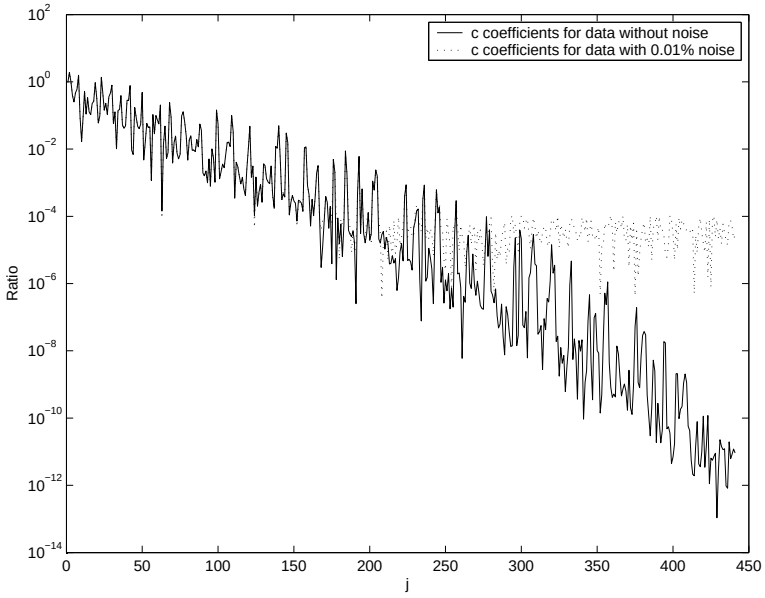

Figure 24: Coefficients $|c_j|$ as function of j

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

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

depending on the positive parameter δ of the Tikhonov form, and for $x^\delta := Vy^\delta$ 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 (7.12) 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 22, replacing the truncation strategy by the above regularization. Figure 23 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 $1.46 \cdot 10^{19}$,

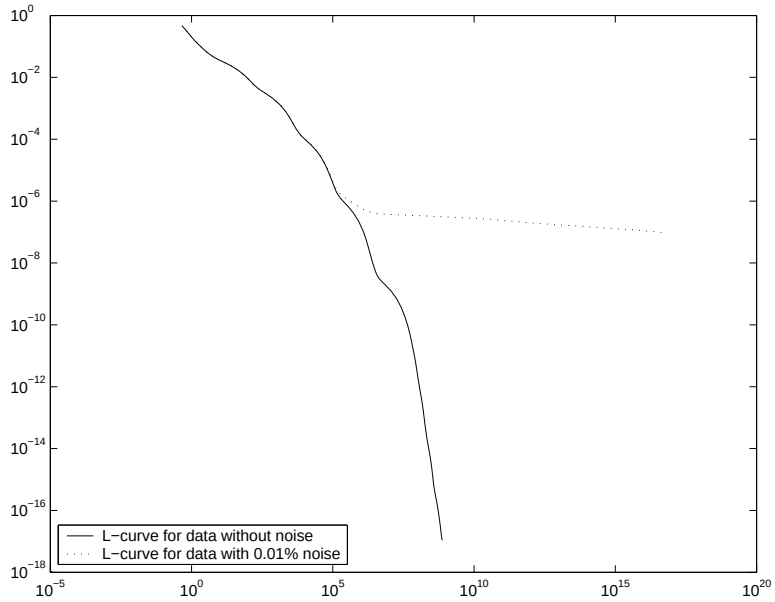

Figure 25: The L-curve for the same problem

but it may actually be higher. Figure 24 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 24 and 23 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 [Han92, Han94, Han00] including a MATLAB software package.
Figure 25 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.

8 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 stable local polynomial approximation. The second alternative is more modern, but, as the other one, it is hardcore mathematics and involves quite some work. Since the current forms of sampling inequalities contain a good deal of concealed oversampling, it will hopefully worthwhile to start slowly and exhibit the places where oversampling kicks in. Thus we shall first focus on motivating and explaining the important ingredients to error bounds in general, before we reproduce the current state–of–the art.

8.1 General Considerations

For simplicity, we start with an unconditionally or $\mathcal{P}$–conditionally real–valued positive definite symmetric kernel K on a s set Ω. This means that we rule out the truly complex–valued and the positive **semi**definite case. The reason is that we want to work with a true Lagrange basis $u_1^*, \ldots, u_N^*$ whenever we have a $\mathcal{P}$–unisolvent set $X = \{x_1, \ldots, x_N\}$ of points of Ω, and we want to postpone extensions to a later version of the text.

We assume that we want to recover an unknown function f from the native space $\mathcal{N}$ of K from its given data $f(x_1), \ldots, f(x_N)$ on X, and we shall focus on interpolation by kernel translates and functions from $\mathcal{P}$. This means that the interpolant $s = s_{X,f,K}$ exists and is uniquely defined by its representation

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

in terms of the Lagrange basis. This is the same for the unconditionally and the $\mathcal{P}$–conditionally positive definite case, though the constructions of the Lagrange bases in (2.41) and (5.21) are different.

The error bounds we are looking for are of the form

$$|f(x) - s(x)| \leq ?$$

for arbitrary functions $f \in \mathcal{N}$ and arbitrary points $x \in \Omega$. We have the standard bound

$$|f(x) - s(x)| \leq P_X(x)\|f\|_{\mathcal{N}}$$

from (2.30) and (5.32) via the Power Function at our disposal. This nicely splits the effect of f and X into two independent factors, and we can use the optimality property of the Power Function from (2.31) and (5.29) for getting upper bounds on it. We shall pursue this line of argument later, since it gives a lot of information. Furthermore, in the case of kernels coming from covariances, the natural notion of an error bound is given by the Power Function itself, because it describes the variance of the Kriging estimation error, and leads to confidence bands in the case of Gaussian processes.

But in may cases, e.g. for the Gaussian kernel, the native space is far too small to be useful, and thus one wants to extend these error bounds to larger function spaces. This is where the quest for sampling inequalities started.

8.2 Sampling Inequalities

The basic idea of **sampling inequalities** is to forget about kernels and their sometimes exotic native spaces. 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 C, we can write

$$|f(x) - s(x)| \quad \leq \quad |f(x_j) - s(x_j)| + 2C \cdot \|x - x_j\|_2$$

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

$$\|f - s\|_{\infty,\Omega} \leq 2C \cdot h,$$

but we need to have C 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.

This can be cast into a more general bound like

$$\|f\|_{\infty,\Omega} \leq F(h(X,\Omega))|f|_{\mathcal{F}} + C \cdot \|f\|_{\infty,X} \tag{8.1}$$

with $F(h) \to 0$ for $h \to 0$, and holding for all f in some function space $\mathcal{F}$ with a (semi–) norm $|f|_{\mathcal{F}}$. This is a special case of a **sampling inequality**.

Its application to error functions $f - s$ of interpolants on X works whenever f and s are in $\mathcal{F}$ via

$$\begin{aligned} \|f-s\|_{\infty,\Omega} &\leq F(h(X,\Omega))|f-s|_{\mathcal{F}} + C \cdot \|f-s\|_{\infty,X} \\ &= F(h(X,\Omega))|f-s|_{\mathcal{F}} \\ &\leq F(h(X,\Omega))(|f|_{\mathcal{F}} + |s|_{\mathcal{F}}). \end{aligned}$$

At this point, one has to take into account that $|s|_{\mathcal{F}}$ will still depend on X and thus also on $h(X,\Omega)$, but in many cases one can infer a stability bound of the form

$$|s|_{\mathcal{F}} \leq C|f|_{\mathcal{F}}$$

with a constant independent of X, e.g. when we have the usual optimality principle for interpolants in native spaces. This leads to

$$\|f-s\|_{\infty,\Omega} \leq (1+C)F(h(X,\Omega))|f|_{\mathcal{F}} \tag{8.2}$$

with $F(h) \to 0$ for $h \to 0$ and works for general function spaces, circumventing the restriction to native Hilbert spaces.

8.3 Simple Bounds for Power Functions

In the unconditional case, we can use Theorem 2.29 on page 20 for some simple upper bounds. For $X \subseteq Y \subseteq \Omega$ we have $P_Y(x) \leq P_X(x)$ for all $x \in \Omega$. And if we have two native spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ with associated kernels K_1 and K_2, respectively, such that for the unit balls we have an inclusion

$$f \in \mathcal{H}_1,\ \|f\|_{\mathcal{H}_1} \leq 1 \Rightarrow f \in \mathcal{H}_2,\ \|f\|_{\mathcal{H}_2} \leq 1,$$

then

$$P_{X,K_1}(x) \leq P_{X,K_2}(x) \text{ for all } x \in \Omega.$$

This generalizes to the case of bounded inclusions. Assume two unconditional kernels K_1, K_2 with native spaces $\mathcal{H}_1$, $\mathcal{H}_2$ such that

$$f \in \mathcal{H}_1 \Rightarrow f \in \mathcal{H}_2,\ \|f\|_{\mathcal{H}_2} \leq C\|f\|_{\mathcal{H}_1}.$$

Then

$$\begin{aligned} P_{X,\mathcal{H}_1}(x) &= \sup_{f\in\mathcal{H}_1, f(X)=\{0\}, \|f\|_{\mathcal{H}_1}\leq 1} f(x) \\ &\leq \sup_{f\in\mathcal{H}_2, f(X)=\{0\}, \|f/C\|_{\mathcal{H}_2}\leq 1} f(x) \\ &\leq C \sup_{g\in\mathcal{H}_2, g(X)=\{0\}, \|g\|_{\mathcal{H}_2}\leq 1} g(x) \\ &= C\, P_{X,\mathcal{H}_2}(x) \end{aligned}$$

allows to carry all upper bounds on $P_{X,\mathcal{H}_2}$ over to upper bounds on $P_{X,\mathcal{H}_1}$ up to a constant factor.

Roughly speaking: larger native spaces in the sense of unit ball inclusion or bounded inclusion lead to larger Power Functions.

Other upper bounds for the Power Function are based on (2.31) and (5.29). There, a set of functions $u_1, \ldots, u_N$ occur which in case of $\mathcal{P}$–conditional positive definiteness must additionally recover functions from $\mathcal{P}$ in the sense of the second set of equations in (5.21). Then the upper bounds are of the form

$$\begin{aligned} P_X^2(x) \leq\ & K(x,x) - 2\sum_{j=1}^{N} u_j(x)K(x_j,x) \\ & + \sum_{j,k=1}^{N} u_j(x)u_k(x)K(x_j,x_k). \end{aligned} \tag{8.3}$$

The simplest case uses **nearest–neighbor reconstruction**. Assume that for each $x \in \Omega$ we pick a single $x_{k(x)} \in X$ and define

$$u_j(x) := \left\{ \begin{array}{ll} 1 & j = k(x) \\ 0 & \text{else} \end{array} \right\}.$$

Then

$$P_X^2(x) \leq K(x,x) - 2K(x_{k(x)},x) + K(x_{k(x)},x_{k(x)}) = d(x,x_{k(x)})^2$$

with the distance defined in (2.17). This shows that one should pick $x_{k(x)} \in X$ closest to x in that distance. Since this recovery process reproduces constants, we have

Theorem 8.4. *If K is unconditionally positive semidefinite or conditionally positive semidefinite with respect to the space of constant functions, the Power Function on nonempty sets X of interpolation points satisfies*

$$P_X(x) \leq \min_{x_j \in X} d(x, x_j)$$

with the distance defined in (2.17). □

Note that this simple result does not assume any smoothness of K or any structure on Ω.

If the domain Ω lies in $\mathbb{R}^d$, we can use barycentric coordinates $u_0(x), \ldots, u_d(x)$ if the point x lies in a nondegenerate simplex with vertices $x_0, \ldots, x_d$. This yields a process that recovers all linear polynomials. We set $u_j(x) = 0$ for all other indices j. Then by standard arguments on the "linear precision" of barycentric coordinates, and for twice continuously differentiable functions f,

$$|f(x) - \sum_{j=0}^{d} u_j(x) f(x_j)| \leq C(f)\epsilon(x)^2$$

if $\epsilon(x)$ is the diameter of the simplex, and with a constant $C(f)$ that involves the second derivatives of f in such a way that it acts like a seminorm that vanishes on all polynomials of degree at most 1.

We now assume that the native space $\mathcal{N}$ of the kernel K is contained in the space of twice differentiable functions in the sense that there is a bounded immersion, i.e. there is a bound

$$|f(x) - \sum_{j=0}^{d} u_j(x) f(x_j)| \leq C(f)\epsilon(x)^2 \leq c\,\epsilon(x)^2 |f|_{\mathcal{N}} \text{ for all } f \in \mathcal{N} \qquad (8.5)$$

with a constant c independent of f and X. By Theorem 5.31 we then get

$$\begin{aligned} P_X(x) &\leq \|\delta_x - \sum_{j=0}^{d} u_j(x)\delta_{x_j}\|_L \\ &\leq c\,\epsilon(x)^2 \end{aligned}$$

with an even easier argument in case of unconditional positive semidefiniteness via (2.31).

Theorem 8.6. *Assume that Ω is a compact domain in $\mathbb{R}^d$ and the kernel K is*

1. *unconditionally positive semidefinite or*

2. *conditionally positive semidefinite with respect to polynomials of degree at most one,*

3. *and has a native space which is continuously embedded in the space of twice continuously differentiable functions in the sense of (8.5).*

If an arbitrary point $x \in \Omega$ lies in a nondegenerate simplex of diameter $\epsilon(x)$ spanned by $d+1$ data points of some interpolation set X, then the Power Function can be bounded at x by

$$P_X(x) \leq c\epsilon(x)^2.$$

□

Note that this applies to smooth unconditionally positive semi–definite kernels, but the connection to thin–plate splines is not obvious at this point. It mimics the error bounds for piecewise linear interpolation on simplices, i.e. for the simplest finite element spaces.

Clearly, this argument generalizes to orders larger than 2 in (8.5) and Theorem 8.6, provided that $\epsilon(x)$ is replaced by some other useful quantity $E(x,h)$ which is small if the point x is surrounded by sufficiently many well–placed points of X. Note that it uses local error bounds for local recovery processes to prove local error bounds for **global** processes of an order that is not worse than the order of the best possible local recovery. But it turns out to be not so easy to find the right quantity $E(x,h)$ and the right notions for "sufficiently many" and "well–placed" in more general situations.

Let us have a short look at necessary conditions for good bounds on the Power Function. Assume that we can prove something like

$$P_X(x) \leq CE(x,h)$$

for all data sets X with fill distance at most h. This implies

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

for the Lagrange–type basis associated to the kernel and the data set X. Thus there is a continuously embedded subspace H of $\mathcal{H}$ (here: $\mathcal{H}$ itself) and a set of reproduction functions u_j such that

$$\left| f(x) - \sum_{j=1}^{N} u_j(x) f(x_j) \right| \leq CE(x,h)\|f\|_{H} \text{ for all } f \in H,\, x \in \Omega..$$

If we try to prove upper bounds for the Power Function via finding instances where the above inequality holds, we have not departed from the main road. We just have to find the best of all such reconstruction processes, and issues like oversampling or stability are not necessarily on our way.

To do the more general case, we can simplify the upper bound (8.3) by introducing the error operator

$$E_x^y(f(y)) := f(x) - \sum_{j=1}^{N} u_j(x) f(x_j)$$

to get

$$\begin{aligned} P_X^2(x) &\leq K(x,x) - 2\sum_{j=1}^{N} u_j(x) K(x_j,x) \\ &\quad + \sum_{j,k=1}^{N} u_j(x) u_k(x) K(x_j,x_k) \\ &= K(x,x) - \sum_{j=1}^{N} u_j(x) K(x_j,x) \\ &\quad + \sum_{j=1}^{N} u_j(x) \left(\sum_{k=1}^{N} u_k(x) K(x_j,x_k) - K(x_j,x) \right) \\ &= E_x^z K(z,x) - \sum_{j=1}^{N} u_j(x) E_x^z K(z,x_j) \\ &= E_x^y E_x^z K(y,z). \end{aligned}$$

Our basic technique will be to use a bound of the form

$$|E_x^y(f(y))| := \left| f(x) - \sum_{j=1}^{N} u_j(x) f(x_j) \right| \leq \epsilon_{X,K}(h) \|Lf\| \tag{8.7}$$

with some linear differential opperator L with values on some normed space. We then can bound the Power Function by

$$\begin{aligned} P_X^2(x) &\leq |E_x^y E_x^z K(y,z)| \\ &\leq \epsilon_{X,K}(h) \|L^y E_x^z K(y,z)\| \\ &\leq \epsilon_{X,K}^2(h) \|L^y\| L^z K(y,z)\|\| \end{aligned} \tag{8.8}$$

if the final expression makes sense.

8.4 Univariate Case

Let us look at the univariate case first, using a compact interval $\Omega = [a, b]$ and a finite subset $X = \{x_1, \ldots, x_N\}$ thereof. Of course, we would fix a point $x \in [a, b]$ and then select a "local" subset

$$X_x := \{x_j \in X \ : \ j \in N(x) \subseteq \{1, \ldots, N\}\}$$

of points of X which are "sufficiently many" and "well–placed" near x. Of course, to keep things elementary, we would like to work with local polynomial recoveries. Thus we fix a positive integer k and work locally with polynomials of order at most k. The simplest idea would be to pick the k closest neighbors to x within X and to perform local Lagrange interpolation by some polynomial p_x of order at most k at these points. If we go for an error bound of the form (8.5), we can take the error formula for interpolation in Newton form as

$$f(y) - p_x(y) = [y, X_x]f \prod_{x_j \in X_x} (y - x_j) \text{ for all } y \in [a, b]$$

where $[y, X_x]f$ is the divided difference on the points of $X_x \cup \{y\}$ applied to f. If we assume f to be continuously k–times differentiable, we get the local error bound

$$|f(x) - p_x(x)| \leq \frac{\|f^{(k)}\|_{\infty,[a,b]}}{k!} \prod_{x_j \in X_x} |x - x_j|.$$

This is of the form (8.7), if we use the fact that

1. the first nearest neighbor to x is at distance at most h,
2. the second nearest neighbor to x is at distance at most $3h$,
3. the third nearest neighbor to x is at distance at most $5h$,
4. the k–th nearest neighbor to x is at distance at most $(2k-1)h$

and thus

$$\prod_{x_j \in X_x} |x - x_j| \leq h^k \frac{(2k)!}{2^k k!}.$$

leading to

$$|E_x^y f(y)| \leq h^k \frac{(2k)!}{2^k (k!)^2} \|f^{(k)}\|_{\infty,[a,b]}.$$

Now we use a univariate kernel K which has k continuous and independent derivatives with respect to both variables. Then we can use (8.8) to get

$$P_X^2(x) \leq \left(h^k \frac{(2k)!}{2^k (k!)^2} \right)^2 \sup_{a \leq z \leq b} \sup_{a \leq y \leq b} \left| \frac{\partial^k}{\partial z^k} \frac{\partial^k}{\partial y^k} K(z,y) \right|$$

which altogether leads to

Theorem 8.9. *Assume a positive semidefinite kernel K on $[a,b] \times [a,b]$ which is k times continuously and independently differentiable with respect to both arguments. Then, with the constant*

$$c_k = \frac{(2k)!}{2^k (k!)^2} \sqrt{\sup_{a \leq z \leq b} \sup_{a \leq y \leq b} \left| \frac{\partial^k}{\partial z^k} \frac{\partial^k}{\partial y^k} K(z,y) \right|},$$

for every point set $X \subset [a,b]$ consisting of at least k points and with fill distance at most h, the Power Function can be bounded in the form

$$P_X(h)(x) \leq c_k h^k \text{ for all } x \in \Omega.$$

Example 8.10

Let us check this for the Gaussian

$$K(x,y) = \exp(-(x-y)^2/2)$$

on the interval $[-1,1]$. The derivatives are linked to Hermite polynomials H_n by

$$\frac{d^n}{dt^n} \exp(-t^2/2) = (-1)^n \exp(-t^2/2) H_n(t).$$

To take derivatives with respect to both arguments, we set $t := x - y$ and use

$$\begin{aligned} \frac{\partial^k}{\partial x^k} \frac{\partial^k}{\partial y^k} K(x,y) &= (-1)^k \frac{d^{2k}}{dt^{2k}}_{|t=x-y} \exp(-t^2/2) \\ &= (-1)^k \exp(-(x-y)^2/2) H_{2k}(x-y) \end{aligned}$$

This is a Hermite function with a well–known, but complicated extremal behavior. For $x, y \in [-1,1]$, we can proceed crudely by bounding $H_{2k}(x-y)$ via bounds on Hermite polynomials on $[-2,2]$ using the recursion

$$H_{n+1}(x) = x H_n(x) - n H_{n-1}(x)$$

and $H_0 = 1,\ H_1(t) = t$. We assert

$$\|H_n\|_{\infty,[-2,2]} \leq 3^n \, n!$$

and get inductively

$$\begin{aligned}\|H_{n+1}\|_{\infty,[-2,2]} &\leq 2\|H_n\|_{\infty,[-2,2]} + n\|H_{n-1}\|_{\infty,[-2,2]} \\ &\leq 2 \cdot 3^n\, n! + n3^{n-1}\,(n-1)! \\ &= 7 \cdot 3^{n-1}\, n! \\ &= \frac{7}{9(n+1)} \cdot 3^{n+1}\,(n+1)! \\ &\leq 3^{n+1}\,(n+1)!\end{aligned}$$

Thus

$$\left|\frac{\partial^k}{\partial x^k}\frac{\partial^k}{\partial y^k}K(x,y)\right| \leq 3^{2k}\,(2k)!$$

on $[-1,1]$ and

$$c_k^2 \leq \frac{((2k)!)^3 3^{2k}}{2^{2k}(k!)^4}.$$

Up to fixed multiplicative constants independent of k, we can apply Stirling's asymptotics

$$n! \approx n^n e^{-n}\sqrt{n}$$

to get the bound

$$\begin{aligned}c_k^2 &\leq C\frac{((2k)^{2k}e^{-2k}\sqrt{2k})^3 3^{2k}}{2^{2k}(k^k e^{-k}\sqrt{k})^4} \\ &\approx C\frac{(2k)^{6k}e^{-6k}k^{3/2}3^{2k}}{2^{2k}k^{4k}e^{-4k}k^2} \\ &\leq C\left(\frac{2^3k^3e^{-3}3}{2k^2e^{-2}}\right)^{2k} \\ &= C\,(12e^{-1}k)^{2k}.\end{aligned}$$

This combines into

$$P_X \leq C\left(12he^{-1}k\right)^k$$

We now couple k to h via requiring

$$k = \left\lfloor \frac{1}{12h} \right\rfloor$$

such that we get

$$12he^{-1}k \leq e^{-1} < 1$$

and the exponential rate

$$P_X \leq C\exp\left(k\log e^{-1}\right) \approx C\exp(-1/(12h)).$$

But we should check if we need oversampling here. For a fill distance of h with N points in $[-1,1]$, we need at least $N \approx 1/h$ points, placing the inner ones at distance $2h$ from each other and the outer ones at $-1+h$ and $1-h$. Thus we need no oversampling for the exponential rate.

Theorem 8.11. *If a kernel on domains in $\mathbb{R}^d$ is radial, i.e. it is a function*

$$K(x,y) = \phi(\|x-y\|_2) = g(\|x-y\|_2^2/2),\ x,y \in \mathbb{R}^d,$$

and if in the above representation the function g is k–times continuously differentiable, then K has continuous partial derivatives up to order k, irrespective which partial derivatives are taken.

Proof: If we set $r := \|x-y\|_2$, then

$$\frac{\partial}{\partial x_j} r = \frac{x_j}{r},\ \frac{\partial}{\partial x_j}(r^2/2) = x_j,\ 1 \le j \le d.$$

Consequently,

$$\frac{\partial}{\partial x_j} g(r^2/2) = g'(r^2/2) x_j,$$

and repeated application of this simple rule shows that each partial derivative requires a derivative of g and produces a factor x_j or y_k. Altogether, a mixed derivative of total order m is a linear combination of derivatives of g at $r^2/2$ up to order m, multiplied by polynomials in x and y of degree up to m. □.

This result can be applied easily to various kernels. Examples are the Wendland kernels or the Whittle/Matern/Sobolev kernels.

Example 8.12

Let us look at the latter, i.e. $r^\nu K_\nu(r) =: g_\nu(r^2/2)$ for $\nu > 0$ with the Bessel function of second kind. It has the property

$$K_\nu'(z) = -K_{\nu+1}(z) + \frac{\nu}{z} K_\nu(z) = -K_{\nu-1}(z) - \frac{\nu}{z} K_\nu(z)$$

and we need

$$g_\nu(s) := \phi_\nu(\sqrt{2s}) = K_\nu(\sqrt{2s})(\sqrt{2s})^\nu$$

and $(\sqrt{2s})' = \frac{1}{\sqrt{2s}}$. Then

$$\begin{aligned}
g_\nu'(s) &= K_\nu'(\sqrt{2s}) \frac{1}{\sqrt{2s}} (\sqrt{2s})^\nu + K_\nu(\sqrt{2s}) \nu (\sqrt{2s})^{\nu-1} \frac{1}{\sqrt{2s}} \\
&= K_\nu'(\sqrt{2s})(\sqrt{2s})^{\nu-1} + \nu K_\nu(\sqrt{2s})(\sqrt{2s})^{\nu-2} \\
&= \left(-K_{\nu-1}(\sqrt{2s}) - \frac{\nu}{\sqrt{2s}} K_\nu(\sqrt{2s}) \right) (\sqrt{2s})^{\nu-1} + \nu K_\nu(\sqrt{2s})(\sqrt{2s})^{\nu-2} \\
&= -K_{\nu-1}(\sqrt{2s})(\sqrt{2s})^{\nu-1} \\
&= -g_{\nu-1}(s).
\end{aligned}$$

Thus the radial kernel $r^\nu K_\nu(r) =: g_\nu(r^2/2)$ has partial derivatives of total order $2k$ for $\nu > 2k$, and then we have convergence of order h^k in the univariate case. □

Example 8.13

Another case is the kernel r^3, which is conditionally positive definite of order 2. We have $g(s) = s^{3/2}$ up to a factor, and at first sight we cannot take two derivatives. But in 1D and in general we have

$$\frac{\partial^2}{\partial x \partial y} g(r^2/2) = x\, y\, g''(r^2/2)$$

and in our case

$$\frac{\partial^2}{\partial x \partial y} g(r^2/2) = \frac{3}{4} xy (r^2/2)^{-1/2} = \frac{3\sqrt{2}}{4} \frac{xy}{r}$$

which has no singularity at zero. Thus the convergence of interpolation is at least like $\mathcal{O}(h)$. With better methods we get $\mathcal{O}(h^{3/2})$. □

It is tempting to generalize all of this directly to the multivariate setting. But the main problem occurs right at the beginning, since it is a problem to prove error bounds for multivariate polynomial interpolation on irregular data sets. Even more, it is highly nontrivial to find good sufficient conditions for unisolvency.

8.5 Conditions for Unisolvency

We now want to derive sufficient and (if possible) necessary conditions for sets $X = \{x_1, \ldots, x_N\}$ to be unisolvent on sets $\Omega \subset \mathbb{R}^d$ with respect to linear spaces $\mathcal{P} = \text{ span } \{p_1, \ldots, p_Q\}$ of dimension $Q \leq N$ on Ω.

To begin with, it is clear that unisolvency is equivalent to the existence of functions $u_1, \ldots, u_N$ on Ω such that the reproduction equations

$$p_k(x) = \sum_{j=1}^{N} u_j(x) p_k(x_j), \ x \in \Omega, \ 1 \leq k \leq Q \tag{8.14}$$

hold. But note that at this point it is not clear whether we have $u_j \in \mathcal{P}$.

For what follows, we should introduce the *sampling map*

$$T \ : \ C(\Omega) \to \mathbb{K}^N, \ f \mapsto (f(x_1), \ldots, f(x_N))^T$$

and the vector

$$u(x) := (u_1(x), \ldots, u_N(x))^T \in \mathbb{K}^N.$$

Then we have a **stability inequality** of the form

$$|p(x)| \leq \|u(x)\|_q \|T(p)\|_r \text{ for all } p \in \mathcal{P},\ x \in \Omega$$

with q, r–norms on $\mathbb{K}^N$ satisfying $1/q + 1/r = 1$. For all $p \in \mathcal{P} \setminus \{0\}$, all $u(x) \in \mathbb{K}^N$ and all $x \in \Omega$ this implies

$$\frac{|p(x)|}{\|T(p)\|_r} \leq \sup_{p \neq 0} \frac{|p(x)|}{\|T(p)\|_r} \leq \inf_{u(x) \in \mathbb{K}^N, (8.14)} \|u(x)\|_q \leq \|u(x)\|_q.$$

This suggests that the two inner optimization problems are in weak duality.

Theorem 8.15. *Assume that $X = \{x_1, \ldots, x_N\} \subseteq \Omega$ is unisolvent for $\mathcal{P} = span\ \{p_1, \ldots, p_Q\}$ of dimension $Q \leq N$ on Ω. and let $x \in \Omega$ be fixed. Then the middle inequality in the above display is an equality, i.e.*

$$\sup_{p \neq 0} \frac{|p(x)|}{\|T(p)\|_r} = \inf_{u(x) \in \mathbb{K}^N, (8.14)} \|u(x)\|_q =: C^*(x)$$

and

$$\begin{array}{rcl} |p(x)| & \leq & C^*(x)\|T(p)\|_r \text{ for all } x \in \Omega,\ p \in \mathcal{P}, \\ \|u(x)\|_q & \geq & C^*(x) \text{ for all } x \in \Omega,\ u(x) \text{ with } (8.14) \end{array}$$

and there is an optimal recovery defined by some $u^*(x) \in \mathbb{K}^N$ with $\|u^*(x)\|_q = C^*(x)$ satisfying (8.14).

Proof: For each $u(x) \in \mathbb{K}^N$ with (8.14) we consider the linear functional

$$\mu_x\ :\ T(p) \mapsto \sum_{j=1}^{N} u_j(x) p(x_j) \text{ for all } p \in \mathcal{P}.$$

This is well–defined on the subspace $T(\mathcal{P}) \subseteq \mathbb{K}^N$ in the r–norm, and by the Hahn–Banach theorem there is an extension to all of $\mathbb{K}^N$ with the same norm in the dual. Thus there is a vector $\hat{u}(x) := (\hat{u}_1(x), \ldots, \hat{u}_N(x))^T \in \mathbb{K}^N$ extending the functional, i.e. it also satisfies (8.14), and its norm satisfies

$$\|\hat{u}(x)\|_q = \|\mu_{x|_{T(\mathcal{P})}}\|_{(\mathbb{K}^N)^*} = \sup_{p \in \mathcal{P},\, p \neq 0} \frac{|\mu_x(T(p))|}{\|T(p)\|_r} = \sup_{p \in \mathcal{P},\, p \neq 0} \frac{|p(x)|}{\|T(p)\|_r}$$

proving the assertion. □

RS: ToDo: do this as duality in convex optimization...

This implies that finding a recovery via $u(x)$ with (8.14) and the smallest possible Lebesgue function value $\|u(x)\|_q$ is the same as finding the smallest possible constant $C(x)$ in a stability inequality

$$|p(x)| \leq C(x)\|T(p)\|_r \text{ for all } p \in \mathcal{P},$$

and the minimal constant $C^*(x)$ is equal to the smallest possible value $\|u^*(x)\|_q$ of the Lebesgue function.

Towards Moving Least Squares, we can specialize the above theorem to weighted ℓ_2 norms.

Corollary 8.16. *Assume that $X = \{x_1, \ldots, x_N\} \subseteq \Omega$ is unisolvent for $\mathcal{P} = \text{span}\,\{p_1, \ldots, p_Q\}$ of dimension $Q \leq N$ on Ω. and let $x \in \Omega$ be fixed. Then the minimization of*

$$\|u(x)\|_{2,w}^2 := \sum_{j=1}^{N} |u_j^2(x)|^2 w_j$$

with positive weights $w_1, \ldots, w_N$ is equivalent to solving

$$\sup_{p \neq 0} \frac{|p(x)|}{\|T(p)\|_{2,1/w}},$$

and the resulting optimal recovery $u^*(x)$ satisfies the stability inequality

$$|p(x)| \leq \|u^*(x)\|_{2,w} \|T(p)\|_{2,1/w}. \quad \square$$

In the classical theory of Lebesgue functions, users will choose $r = \infty$ and $q = 1$. This leads to a very useful result:

Corollary 8.17. *Let $\Omega \subset \mathbb{R}^d$ be compact, and let the functions of $\mathcal{P}$ be continuous. If for all $p \in \mathcal{P}$ with $\|p\|_{\infty,\Omega} = 1$ we have $\|p\|_{\infty,X} \geq 1/C$ with $C \geq 1$, then X is $\mathcal{P}$–unisolvent and there is a recovery with $\|u(x)\|_1 \leq C$ for all $x \in \Omega$.*

Proof: Take an arbitrary $p \in \mathcal{P}$ with $\|p\|_{\infty,\Omega} = 1$. Then

$$\|p\|_{\infty,X} \geq \frac{1}{C} = \frac{1}{C}\|p\|_{\infty,\Omega}$$

implies via rescaling that for all $x \in \Omega$ and all $p \in \mathcal{P}$ we have

$$|p(x)| \leq \|p\|_{\infty,\Omega} \leq C\|p\|_{\infty,X}.$$

This implies unisolvency, and the rest follows from Theorem 8.15. $\square$

There also is a converse:

Corollary 8.18. *Let $\Omega \subset \mathbb{R}^d$ be compact, and let the functions of $\mathcal{P}$ be continuous. If X is $\mathcal{P}$–unisolvent, then there is a constant $C \geq 1$ such that for all $p \in \mathcal{P}$ with $\|p\|_{\infty,\Omega} = 1$ we have $\|p\|_{\infty,X} \geq 1/C$. Furthermore, there is a recovery with $\|u(x)\|_1 \leq C$ with the same constant.*

Proof: If we parametrize functions in P via a basis, we see that the values at maxima are continuously dependent on the coefficients, and thus

$$\sup_{p\in\mathcal{P},p\neq 0}\frac{\|p\|_{\infty,\Omega}}{\|p\|_{\infty,X}}$$

can be calculated via coefficients on a sphere, and thus the supremum is a finite number $C\geq 1$. All the rest follows from what we know already. □

To illustrate this for the Euclidean case, we provide

Theorem 8.19. *Assume that $X=\{x_1,\ldots,x_N\}\subseteq\Omega$ is unisolvent for $\mathcal{P}=$ span $\{p_1,\ldots,p_Q\}$ of dimension $Q\leq N$ on Ω. and let $x\in\Omega$ be fixed. In the case $r=q=2$ the recovery*

$$p(x)=\sum_{j=1}^{N}u_j(x)p(x_j)\ \textit{for all}\ p\in P \tag{8.20}$$

with minimal

$$\|u(x)\|_2^2:=\sum_{j=1}^{N}|u_j(x)|^2 \tag{8.21}$$

has the form

$$u_j^*(x)=\sum_{k=1}^{Q}a_k^*(x)\overline{p_k(x_j)}$$

with uniquely defined coefficients $a_k^*(x)$ satisfying

$$p_m(x)=\sum_{k=1}^{Q}a_k^*(x)\sum_{j=1}^{N}\overline{p_k(x_j)}p_m(x_j)=\sum_{j=1}^{N}u_j^*(x)p_m(x_j),\ 1\leq m\leq Q. \tag{8.22}$$

Both the a_k^* and the u_j^* are in $\mathcal{P}$.

Proof: By unisolvency, the system (8.22) is uniquely solvable, the $u_j^*(x)$ are well-defined and provide the reproduction. To see that they minimize (8.21), we have to prove that the vector with components $u_j^*(x)$ is orthogonal to the affine subspace defined by (8.20). This means that we have to take $v_j(x)$ with

$$0=\sum_{j=1}^{N}v_j(x)p(x_j)\ \text{for all}\ p\in P$$

and prove

$$\begin{aligned}&\sum_{j=1}^{N}u_j^*(x)\overline{v_j(x)}\\ =\ &\sum_{k=1}^{Q}a_k^*(x)\sum_{j=1}^{N}\overline{p_k(x_j)v_j(x)}=0.\end{aligned}$$ □

Now we define

$$p_x^* := \sum_{k=1}^{Q} \overline{a_k^*(x)} p_k \in P$$

and take an arbitrary $p \in P$ to get

$$\begin{aligned} |p(x)|^2 &= \left| \sum_{j=1}^{N} u_j^*(x) p(x_j) \right|^2 \\ &\leq \left(\sum_{j=1}^{N} |u_j^*(x)|^2 \right) \left(\sum_{j=1}^{N} |p(x_j)|^2 \right). \end{aligned} \tag{8.23}$$

We assert that equality is attained for p_x. For this, look at (8.22) and see that there is a reproduction

$$p_m(x) = \sum_{j=1}^{N} \overline{p_x^*(x_j)} p_m(x_j),\ 1 \leq m \leq Q.\ x \in \Omega,$$

which in particular yields

$$p_x(x) = \sum_{j=1}^{N} \overline{p_x^*(x_j)} p_x(x_j) = \sum_{j=1}^{N} |p_x(x_j)|^2$$

and by optimality of the $u_j^*(x)$ we have

$$p_x(x) = \sum_{j=1}^{N} |p_x^*(x_j)|^2 \geq \sum_{j=1}^{N} |u_j^*(x)|^2.$$

Now we check (8.23) for p_x and get that it is attained with equality, and in particular

$$p_x(x) = \sum_{j=1}^{N} \overline{p_x^*(x_j)} p_x(x_j) = \sum_{j=1}^{N} |p_x(x_j)|^2 = \sum_{j=1}^{N} |u_j^*(x)|^2. \quad \square$$

For later use with Moving Least Squares, we add another property of reconstructions which are optimally oversampled in the ℓ_2 sense. It turns out that the ℓ_2–optimally stable reproduction guarantees an optimal least–squares data error at the same time.

Theorem 8.24. *Let X be P–unisolvent, and consider a weighted minimization of*

$$\|u(x)\|_w^2 := \sum_{j=1}^{N} |u_j(x)|^2 w_j$$

with positive weights w_j under the reproduction constraints (8.20). Denote the solution by $u^*(x)$. For all weighted least–squares problems minimizing

$$\sum_{j=1}^{N} |f(x_j) - p(x_j)|^2 / w_j$$

under all $p \in P$, the solution $p_f^* \in P$ satisfies

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

Proof: Denoting the diagonal matrix with weights w_j by D_w, and the $Q \times N$ matrix of values $p_m(x_j)$ by A, and $P(x) := (p_1(x), \ldots, p_Q(x))^T$, the solution $u^*(x)$ of the first problem minimizes the penalized quadratic form

$$u^T(x) D_w \overline{u(x)} + z^T \overline{(P(x) - Au(x))}$$

leading to

$$(u^*)^T(x) D_w = z(x)^T \overline{A}, \text{ or } D_w u^*(x) = A^* z(x).$$

Equation (8.20) then leads to

$$\begin{aligned} Au^*(x) &= AD_w^{-1} A^* z(x) = P(x), \\ z(x) &= (AD_w^{-1} A^*)^{-1} P(x) \\ D_w u^*(x) &= A^* (AD_w^{-1} A^*)^{-1} P(x) \end{aligned}$$

and ends up with

$$u^*(x) = D_w^{-1} A^* (AD_w^{-1} A^*)^{-1} P(x).$$

For the second problem, write

$$p(x) = (p_1(x), \ldots, p_Q(x))\, c$$

and use $T : F \mapsto (f(x_1), \ldots, f(x_N))^T$ again, to let the problem take the form

$$\min \|T(f) - T(p)\|_{1/w}^2 = \min \|T(f) - A^T c\|_{1/w}^2.$$

Then by similar reasoning, we get

$$\begin{aligned} (T(f) - A^T c_f^*)^T D_w^{-1} A^* &= 0, \\ T(f)^T D_w^{-1} A^* &= (A^T c_f^*)^T D_w^{-1} A^* \\ \overline{A} D_w^{-1} T(f) &= \overline{A} D_w^{-1} A^T c_f^* \\ c_f^* &= (\overline{A} D_w^{-1} A^T)^{-1} \overline{A} D_w^{-1} T(f) \end{aligned}$$

and

$$\begin{aligned} T(f)^T u^*(x) &= T(f)^T D_w^{-1} A^* (A D_w^{-1} A^*)^{-1} P(x) \\ &= ((\overline{A} D_w^{-1} A^T)^{-1} \overline{A} D_w^{-1} T(f))^T P(x) \\ &= (c_f^*)^T P(x) = p_f^*(x). \quad \square \end{aligned}$$

RS: There seems to be a similar theorem for general q, r–norms.

We now proceed towards finding $\mathcal{P}$–unisolvent subsets of sets $X = \{x_1, \ldots, x_N\}$ of scattered points in a compact domain $\Omega \subset \mathbb{R}^d$ with some fill distance h. We assume $\mathcal{P}$ to consist of continuous functions. The goal is to prove $\mathcal{P}$–unisolvency for arbitrary sets with sufficiently small h. We might later restrict ourselves to small subsets of Ω, but this is a later issue, since at this point Ω might be just a small ball.

We want to use Corollary 8.17 for our purpose. This means that we start with some $p \in \mathcal{P}$ with $\|p\|_{\infty,\Omega} = 1$, and we can select an $x \in \Omega$ with $p(x) = 1$ without loss of generalization. If we can manage to show that there is an $x_j \in X \cap \Omega$ with $p(x_j) \geq 1/C > 0$, we are done after application of Corollary 8.17. If h is small enough, we can surely find such points x_j, and they will be close to x, but we do not know how fast p falls when we go from x to x_j. To get this under control, we must control differences of p–values at different points. This means that we have to control derivatives.

8.6 Stable Polynomial Reproduction

To do this for spaces $\mathcal{P}_m^d$ of d–variate polynomials of order m, we focus first on polynomials in $\mathcal{P}_m^1$ on $[-1, 1]$. There, the classical **Bernstein–Markov inequality** is

$$\|q'\|_{\infty,[-1,1]} \leq (m-1)^2 \|q\|_{\infty,[-1,1]} \text{ for all } q \in \mathcal{P}_m^1.$$

This turns into the form we need in 1D, if we work along

$$\begin{aligned} |q(s) - q(t)| &\leq |q'(\tau)||s-t| \\ &\leq (m-1)^2 \|q\|_{\infty,[-1,1]} |s-t| \end{aligned}$$

for all $s, t \in [-1, 1]$, $q \in \mathcal{P}_m^1$. This looks fine, but remember that we have q defined and bounded on all of $[-1, 1]$ while we use it only between s and t. In what follows, we can assume $m \geq 2$ throughout, because the case $m = 0$ is empty and the case $m = 1$ is trivial.

To apply this argument for a generalized Bernstein–Markov inequality on a convex set, we are tempted to connect x and x_j by a line, but in view of

the 1D case we will need that we have to be able to evaluate and bound the polynomials on more than just the section of the line between x and x_j. In a more explicit form, the drawback is that if we rewrite the classical Bernstein–Markov inequality for a small interval $[-\alpha, \alpha]$, we get

$$\|q'\|_{\infty,[-\alpha,\alpha]} \leq \frac{(m-1)^2}{\alpha} \|q\|_{\infty,[-\alpha,\alpha]} \text{ for all } q \in \mathcal{P}^1_m$$

by plugging $q(\alpha\, t)$ into the inequality on $[-1, 1]$. Thus it will not pay off to have something like $\alpha = \|x - x_j\|_2$. We need that we can draw a sufficiently long line from all points $x \in \Omega$ through at least one point $x_j \in X$, the length of the line segment contained in Ω being larger than h or $\|x - x_j\|_2$. The way out is the **interior cone condition** that we now describe.

Definition 8.25. *A cone with vertex x, axis direction z with $\|z\|^2 = 1$, height H, and angle $\theta \in (0, \pi/2)$ is the set*

$$\{x + \lambda y \ : \ 0 \leq \lambda \leq H,\ y \in \mathbb{R}^d,\ \|y\|_2 = 1,\ \ z^T y \geq \cos\theta\}.$$

To understand this, note that the angle between z and y should be at most θ. Since the cosine is decreasing, this means $\cos \angle(y, z) = z^T y \geq \cos\theta$.

Definition 8.26. *A set $\Omega \subset \mathbb{R}^d$ has an* **interior cone condition** *of angle $\theta \in (0, \pi/2)$ and height $H > 0$, if for each $x \in \Omega$ there is a cone with height at least H, angle at least θ and arbitrary axis $z(x)$ which is completely contained in Ω. This means that the cones*

$$\{x + \lambda y \ : \ 0 \leq \lambda \leq H,\ y \in \mathbb{R}^d,\ \|y\|_2 = 1,\ \ y^T z(x) \geq \cos\theta\}.$$

are all contained in Ω for all $x \in \Omega$ and a selection of unit axis vectors $z(x)$.

These cones contain balls of the form $B(x + tz(x), r(t))$ with radii $r(t) \leq t\, \sin(\theta)$ for $t \leq \frac{H}{1+\sin(\theta)}$. This follows easily from a litte drawing of the situation. Furthermore, the above balls do not contain x if $t > 0$. Finally, if some $v = x + \lambda y$ lies in the cone above, then the whole ray from x to $x + Hy$ lies in the cone.

Theorem 8.27. *Assume that a compact domain $\Omega \subset \mathbb{R}^d$ satisfies an interior cone condition of angle $\theta \in (0, \pi/2)$ and height $H > 0$. Then for all finite sets $X \subset \Omega$ with fill distance h satisfying*

$$h \ \ \leq \ \ h_0 := \frac{H \sin(\theta)}{1 + \sin(\theta)} \frac{1}{4(m-1)^2}$$

the set X is $\mathcal{P}^d_m$–unisolvent, and there is a recovery $u(x)$ with $\|u(x)\|_1 \leq 2$.

Proof: We again start with a $p \in \mathcal{P}_m^d$ with $\|p\|_{\infty,\Omega} = 1$ and $p(x) = 1$ for some $x \in \Omega$. Then we use the cone at x to work on. If we take

$$t := \frac{h}{\sin\theta} \leq \frac{H}{1+\sin(\theta)}$$

and consider the ball $B(x+tz(x), r(t))$, we can take $r(t) := t\,\sin(\theta) = h$ and can find an $x_j \in X$ in that ball. The ray from x through x_j can be extended up to the point $z_j := x + H(x-x_j)/\|x-x_j\|$, and we have $x \neq x_j$ since we have $t > 0$. Furthermore,

$$\begin{aligned}\|x-x_j\|_2 &\leq t + r(t)\\ &= t(1+\sin(\theta))\\ &= h\frac{1+\sin(\theta)}{\sin(\theta)} \leq H.\end{aligned}$$

We now apply the 1D Bernstein–Markov inequality along the line section from x to z_j via x_j, setting

$$q(t) := p(x + t(x_j - x)/\|x_j - x\|),\ 0 \leq t \leq H.$$

Then the scaled version of the Bernstein–Markov inequality is

$$\|q'\|_{\infty,[0,H]} = \frac{2(m-1)^2}{H}\|q\|_{\infty,[0,H]}.$$

This yields

$$\begin{aligned}|p(x) - p(x_j)| &= |q(0) - q(\|x-x_j\|_2)|\\ &\leq \|x-x_j\|_2 \frac{2(m-1)^2}{H}\|q\|_{\infty,[0,H]}\\ &\leq h\frac{1+\sin(\theta)}{\sin(\theta)}\frac{2(m-1)^2}{H}\\ &\leq \frac{1}{2},\end{aligned}$$

leading to

$$\begin{aligned}p(x_j) &= p(x_j) - p(x) + 1\\ &\geq \frac{1}{2}.\end{aligned}$$ □

This result still is not local. We would like to apply it in sets $\Omega \cap B(x,r)$ for arbitrary $x \in \Omega$ and certain ball radii r.

Theorem 8.28. *Assume that a compact domain $\Omega \subset \mathbb{R}^d$ satisfies an interior cone condition of angle $\theta \in (0, \pi/2)$ and height $H > 0$. Then for all r with $0 < r \leq H$ and all $x \in \Omega$ the set $\Omega \cap B(x,r)$ contains a ball of radius at least $r\sin(\theta)/(1+\sin(\theta))$.*

Proof: Let $x \in \Omega$ be arbitrary. We can have a cone $\mathcal{C} \subset \Omega$ of height H and angle θ with vertex x. Intersecting it with $B(x,r)$ with some $r \leq H$, we have that the ball $B(x + z(x)r/(1+\sin\theta), r\sin\theta/(1+\sin\theta))$ is in the cone and in the ball $B(x,r)$. □

Then we just need unisolvency conditions on balls. Via a small drawing, we get

Theorem 8.29. *Assume $\Omega = B(0,r)$ with some radius $r > 0$. Then Ω satisfies an interior cone condition with height r and angle θ with $\theta = \pi/3$.*

Proof: On the unit circle, pick the point $x = (-1,0)$ and its cone pointing to the right with a 60 degree angle and radius 1. Moving that cone with x to the right shows that a cone of this size works for all x between $(-1,0)$ and $(0,0)$, thus for all points in the circle. □

Theorem 8.30. *Assume that a compact domain $\Omega \subset \mathbb{R}^d$ satisfies an interior cone condition of angle $\theta \in (0, \pi/2)$ and height $H > 0$. Define*

$$c_0 := \frac{4(2+\sqrt{3})(m-1)^2(1+\sin(\theta))}{\sqrt{3}\sin(\theta)}.$$

Then for all finite sets $X \subset \Omega$ with fill distance h satisfying

$$h \quad \leq \quad h_0 := \frac{H}{c_0}$$

the set $X \cap B(x, c_0 h)$ is $\mathcal{P}_m^d$–unisolvent, and there is a recovery $u(x)$ with $\|u(x)\|_1 \leq 2$ based only on points in $X \cap B(x, c_0 h)$.

Proof: We start with the cone condition at some arbitrary $x \in \Omega$. Thus we can find a ball with radius $r\sin(\theta)/(1+\sin(\theta))$ that is still in $B(x,r) \cap \Omega$ for all $0 \leq r \leq H$. For this ball, we know by Theorem 8.29 that it satisfies an interior cone condition of height $\frac{r\sin(\theta)}{1+\sin(\theta)}$ and angle $\alpha = \pi/3$ with $\sin(\pi/3) = \sqrt{3}/2$. Then the h_0 of Theorem 8.27 for this ball is

$$h_0(r) = \frac{r\sin(\theta)\sqrt{3}}{(2+\sqrt{3})(1+\sin(\theta))} \frac{1}{4(m-1)^2} = \frac{r}{c_0}$$

and we now know that every set X with fill distance $h \leq h_0(r)$ has a unisolvent subset in that ball. We now turn this upside down, starting with some set X with fill distance $h \leq \frac{H}{c_0}$. We then take $r(h) := c_0 h \leq H$ and see that $h_0(r(h)) = h$, leading to unisolvency. □

Definition 8.31. *A compact domain $\Omega \subset \mathbb{R}^d$ allows* **uniformly stable local polynomial reproduction** *of order $m \geq 1$, if there are positive constants c, C, h_0 such that for all finite sets $X = \{x_1, \ldots, x_N\}$ with fill distance $h \leq h_0$ there are scalars $u_1(x), \ldots, u_N(x)$ such that*

$$\begin{aligned} \sum_{j=1}^{N} u_j(x) p(x_j) &= p(x) \text{ for all } p \in \mathcal{P}_m^d, x \in \Omega, \\ \sum_{j=1}^{N} |u_j(x)| &\leq C \text{ for all } x \in \Omega, \\ u_j(x) &= \text{ for all } x_j \text{ with } \|x - x_j\| \leq c_0\, h. \end{aligned}$$

Note that we have proven this for compact domains in $\mathbb{R}^d$ satisfying an interior cone condition with height H and angle θ, where the constants are given above.

To arrive at an error bound for these reproductions on functions f with continuous derivatives up to order m on Ω, we apply the Taylor formula from Theorem 12.1 on page 234 of Section 12.1. We fix a point $x \in \Omega$ with the cone $\mathcal{C} \subset \Omega$ of the proof of Theorem 8.28 and work on points $y \in \mathcal{C} \cap B(x, c_0 r)$ with c_0 from Theorem 8.30. Both the Taylor polynomial $T_x(f)$ of f at $x \in \Omega$ and the local polynomial reproduction

$$s_x(f)(y) := \sum_{\substack{j=1 \\ \|x - x_j\| \leq c_0 h}}^{N} u_j(y) f(x_j),$$

are in $\mathcal{P}_m^d$, and for the latter we are using only those data at the points $x_j \in X$ with $\|x - x_j\|_2 \leq c_0 h$ that we used in the proof of Theorem 8.30. These have the property that all line segments from x to x_j lie completely in Ω, together with the line segments from x to all $y \in \mathcal{C} \cap B(x, c_0 h)$. Thus we can apply the Taylor bound

$$|f(z) - T_x(f)(z)| \leq \|z - x\|_2^m \underbrace{\sum_{|\alpha|=m} \frac{1}{\alpha!} \|D^\alpha f\|_{\infty,\Omega}}_{=:|f|_m} \leq c_0 h^m |f|_m$$

for $z = y$ and all $z = x_j$ for these j. This yields

$$\begin{aligned}
|f(y) - s_x(f)(y)| &\leq |f(y) - T_x(f)(y)| + |T_x(f)(y) - s_x(f)(y)| \\
&= |f(y) - T_x(f)(y)| + |s_x(T_x(f))(y) - s_x(f)(y)| \\
&\leq |f(y) - T_x(f)(y)| + \left| \sum_{j=1}^{N} u_j(y)(T_x(f)(x_j) - f(x_j)) \right| \\
&\leq \left(1 + \sum_{j=1}^{N} |u_j(y)| \right) c_0^m h^m |f|_m \\
&\leq 3 c_0^m h^m |f|_m.
\end{aligned}$$

We now can go back to (8.8) and get

Theorem 8.32. *Under the assumptions and notations of Theorem 8.30, and if we have a conditionally positive semidefinite kernel of order at most m with continuous derivatives up to order m independently in both variables, the Power Function for recovery on $\mathcal{P}_m^d$–unisolvent sets X with fill distance $h \leq h_0$ is bounded by*

$$P_X^2(x) \leq 9 c_0^2 h^{2m} \sup_{y,z \in \Omega} \sum_{|\alpha|,|\beta|=m} \frac{1}{\alpha!\beta!} |D^{\alpha,y} D^{\beta,z} K(y,z)|.$$

RS: This is still open–ended. The final goal is to prove more general stability results in connection with oversampling.
RS; Incomplete here, as of February 1, 2011

8.7 Univariate Sampling Inequalities

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 $I := [a, b] \subset \mathbb{R}$, and there we shall define the (semi–) inner products and (semi–) norms

$$\begin{aligned}
(f,g)_j &:= \int_I f^{(j)}(t) g^{(j)}(t) dt \text{ for all } f, g \in C^j(I) \\
|f|_{\infty,X} &:= \sup_{t \in X} |f(t)| \text{ for all } f \in C(I),\ X \subset I, \\
|f|_{2,X}^2 &:= h(X,\Omega) \sum_{x \in X} f^2(x) \text{ for all } f \in C(I),\ X \subset I,\ |X| < \infty.
\end{aligned}$$

Throughout, we shall confine ourselves to subsets $X \subset [a, b]$ with fill distance $h = h(X, [a, b])$.

Lemma 8.33. *Given $g \in C^1(I)$ and a subset X of $[a,b]$ with fill distance h. Then we have*

$$\begin{array}{rcll} \|g\|_0 & \leq & h\,|g|_1 & +\sqrt{2}|g|_{2,X}, \\ \|g\|_0 & \leq & h\,|g|_1 & +\sqrt{2(b-a)}|g|_{\infty,X}, \\ \|g\|_{\infty,I} & \leq & h\,\|g'\|_{\infty,I} & +|g|_{\infty,X}, \\ \|\ g\|_{\infty,I} & \leq & \sqrt{h}\,|g|_1 & +|g|_{\infty,X}. \end{array} \tag{8.34}$$

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 $\sqrt{a^2+b^2} \le a+b$ for $a, b > 0$. If we sum up the discrete values, we arrive at

$$\begin{aligned} \|g\|_0^2 &\le h^2\ |g|_1^2 + 2h\sum_{j=1}^{N} g(x_j)^2 \\ \|g\|_0 &\le h\ |g|_1 + \sqrt{2}|g|_{2,X}. \end{aligned}$$

□

8.8 Example: Univariate Splines

In the notation of the text on splines, we have

Theorem 8.35. *Let $f \in \mathcal{C}^k[a,b]$ be interpolated by s^* in $N \ge k$ 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

$$\begin{aligned} \|f - s^*\|_{L_2[a,b]} &\le c_k h^k |f - s^*|_k \le 2c_k h^k |f|_k, \\ \|f - s^*\|_{L_\infty[a,b]} &\le c_k h^{k-1/2} |f - s^*|_k \le 2c_k h^{k-1/2} |f|_k. \end{aligned}$$

Proof: Note that the zeros of $f - s^*$ have a distance of at most $2h$ between each other and of at most $h \le 2h$ 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

$$\|f - s^*\|_{L_2[a,b]} \le h \cdot 4h \cdots 4^{k-1}h |f - s^*|_k =: c_k h^k |f - s^*|_k$$

and the left–hand parts of the assertions follow.

For the right-hand parts we use the optimality condition $|s^*|_k \le |f|_k$. □

If some additional boundary conditions are satisfied, the convergence order doubles.

Theorem 8.36. *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*

$$\begin{aligned} \|f - s^*\|_{L_2[a,b]} &\le c_k^2 h^{2k} |f|_{2k}. \\ \|f - s^*\|_{L_\infty[a,b]} &\le \tilde{c}_k^2 h^{2k-1} |f|_{2k}. \end{aligned}$$

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.

8.9 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 $p \in \mathcal{P}_k$ we have

$$\|p\|_{\infty,[a,b]} \leq C |p|_{\infty,X}. \tag{8.37}$$

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 (8.37). If we extend (8.37) 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 (8.37), let us fix a polynomial $p \in \mathcal{P}_k$ 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 8.38. *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 8.39. *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 8.39 is satisfied. Then X clearly is unisolvent, and we know that we can reproduce all polynomials $p \in \mathcal{P}_n$ by a nonunique formula like

$$p(x) = \sum_{x_j \in X} u_j(x) p(x_j). \tag{8.40}$$

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 8.41. *Under all possibilities to satisfy (8.40) under the conditions of Theorem 8.39 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**

$$T_X \ : \ f \mapsto (f(x_1), \ldots, f(x_N))^T \in \mathbb{R}^N,$$

which is continuous on $C[-1, 1]$ with the $\|.\|_\infty$ norm and invertible on $T(\mathcal{P}_n) \subset \mathbb{R}^N$. It has a bounded inverse

$$S \ : \ T(\mathcal{P}_n) \to \mathcal{P}_n \subset C[-1, 1].$$

For each vector $y \in T(\mathcal{P}_n) \subset \mathbb{R}^N$ there is a unique $p \in \mathcal{P}_n$ 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 $T(\mathcal{P}_n)$. By the **Hahn–Banach theorem** it has a **norm–preserving** extension to all of $\mathbb{R}^N$, and this is the abstract and non-constructive part of the argument. As a functional on all of $\mathbb{R}^N$ 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

$$\|\lambda_x\|_1 := \sum_{x_j \in X} |u_j(x)| = \sup_{y \in \mathbb{R}^N \setminus \{0\}} \frac{\left| \sum_{x_j \in X} u_j(x) y_j \right|}{\|y\|_\infty}$$

because the dual of $\mathbb{R}^N$ under the $\|.\|_\infty$ norm is $\mathbb{R}^N$ 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 $T(\mathcal{P}_n)$. 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

$$\|S\| := \sup_{T(p) \in T(\mathcal{P}_n)\setminus\{0\}} \frac{|p(x)|}{\|T(p)\|_\infty} \leq 2.$$

Thus we get the assertion. □

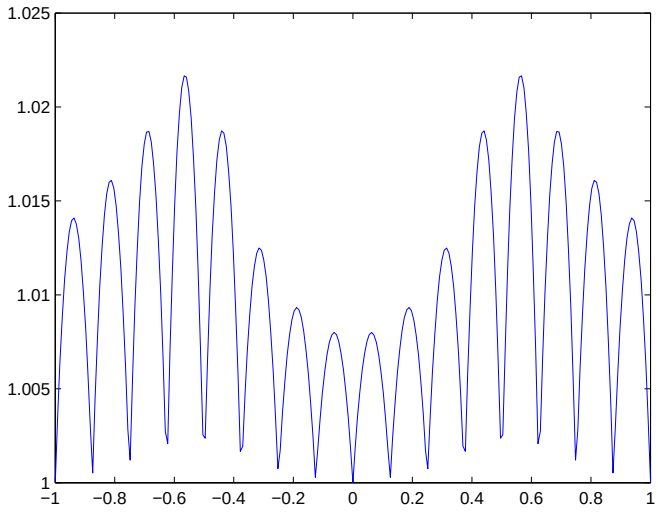

Figure 26: Minimal Lebesgue function for 17 equidistant points in $[-1, 1]$ and polynomial degree $n = 3$

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

$$\text{Minimize } 1_N^T(a+b) \text{ under } P_X(a-b) = p(x)$$

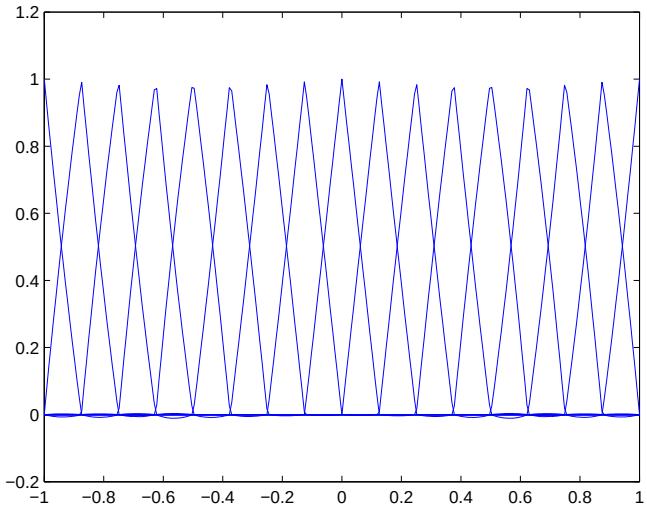

Figure 27: Functions u_j for the same case as in the previous figure

in normal form with nonnegative variables $a, b \in \mathbb{R}^N$. This can be solved pointwise, but the solution is rather strange, see figures 27 and 26. 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 28 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.

8.10 Norming Sets

As a little digression, we generalize the above construction, following an idea of Jetter, Stöckler, and Ward.

... incomplete here.....

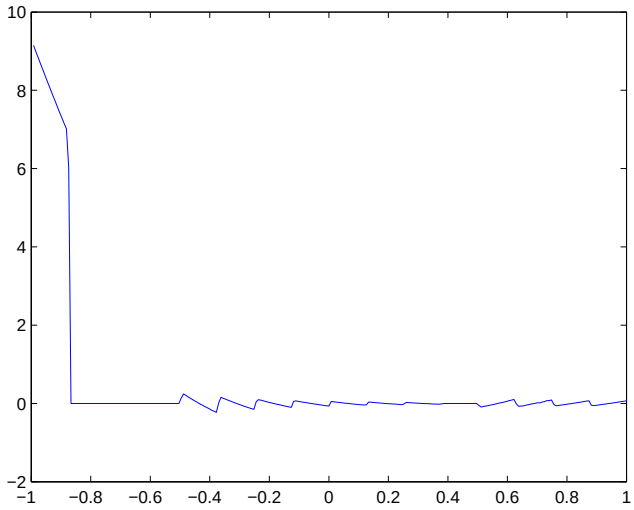

Figure 28: Derivative of u_1 for the same case as in the previous figure

8.11 Multivariate Polynomial Reproduction

In the multivariate setting, we should go for sampling inequalities of the form

$$\begin{array}{rcl} \|u\|_{W_2^m(\Omega)} & \leq & C\left(h^{M-m}|u|_{W_2^M(\Omega)} + h^{-m}\|u\|_{\infty,X_h}\right) \\ \|D^\alpha u\|_{L_\infty(\Omega)} & \leq & C\left(h^{M-|\alpha|-d/2}|u|_{W_2^M(\Omega)} + h^{-m}\|u\|_{\infty,X_h}\right) \end{array} \tag{8.42}$$

for all functions in the Sobolev space $W_2^M(\Omega)$ with the inner product

$$\begin{array}{rcl} (f,g)_{W_2^M(\Omega)} & := & \sum_{j=0}^{M}(f,g)_j \\ (f,g)_j & := & \sum_{|\alpha|=j}\int_\Omega D^\alpha f D^\alpha g \end{array}$$

and (semi–)norms

$$\begin{array}{rcl} \|f\|^2_{W_2^M(\Omega)} & := & (f,f)_{W_2^M(\Omega)} \\ |f|^2_{W_2^M(\Omega)} & := & (f,f)_M = \sum_{|\alpha|=j}\int_\Omega |D^\alpha f|^2 \end{array}$$

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

$$\begin{array}{rcl} \|p\|_{L_2(\Omega)} & \leq & C\|p\|_{\infty,X_h} \\ \|p\|_{L_\infty(\Omega)} & \leq & C\|p\|_{\infty,X_h} \end{array}$$

for all $p \in \mathcal{P}_M$, i.e. there is stable polynomial reproduction in the sense of (8.37). 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 8.43. *Let $\Omega \subset \mathbb{R}^d$ be a domain and fix a number $k \in \mathbb{N}$. 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 $h(X, \Omega) \leq h_0$ there are functions $u_1^X, \ldots, u_N^X$ on Ω such that*

$$\begin{array}{rcll} \sum_{j=1}^N u_j^X(x) p(x_j) & = & p(x) & \text{for all } p \in \mathcal{P}_k,\ x \in \Omega \\ \sum_{j=1}^N |u_j^X(x)| & \leq & c_1 & \text{for all } x \in \Omega \\ \|x - x_j\|_2 > c_2 h & \Rightarrow & u_j^X(x) = 0 & \text{for all } x \in \Omega,\ 1 \leq j \leq N \end{array} \tag{8.44}$$

then we say that the quasi–interpolation process

$$f \mapsto Q_X(f)(\cdot) := \sum_{j=1}^N u_j^X(\cdot) f(x_j) \text{ for all } f : \Omega \to \mathbb{R}$$

defined for all such X provides **stable local polynomial reproduction** *of order k.*

Then we have a rather simple local error bound:

Theorem 8.45. *Assume that $\Omega \subset \mathbb{R}^d$ is bounded and admits stable local polynomial reproduction of order k. Then there is a constant C with*

$$|f(x) - Q_X(f)(x)| \leq C h^{k+1} |f|_{k+1,\Omega^*}$$

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.

Proof: See Wendland's book [Wen05], 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 $\mathcal{P}_k$–unisolvent subsets, because they will not have bounded Lebesgue functions. We thus first aim at a proof of an inequality like

$$\|p\|_{\infty,\Omega} \le c_2 \|p\|_{\infty,X} \text{ for all } p \in \mathcal{P}_k$$

for suitable domains Ω and finite sets $X \subset \Omega$. We start with a polynomial $p \in \mathcal{P}_k$ 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

$$|p\left(x + t(z-x)\right) - p(x)| = |q(t) - q(0)| = \left|\int_0^t q'(t)dt\right| \le t\|q'\|_{\infty,[-1,1]}$$

and use Markov's inequality to get

$$|p\left(x + t(z-x)\right) - p(x)| \le t(k-1)^2\|q\|_{\infty,[-1,1]} \le t(k-1)^2\|p\|_{\infty,\Omega}$$

for all $t \in [0,1]$. But for our choice of $1 = p(x) = \|p\|_{\infty,\Omega}$ we get

$$p\left(\underbrace{x + t(z-x)}_{=:y_t}\right) = q(t) \ge 1 - 2t(k-1)^2$$

leading to $p(y_t) \ge 1/2$ for all y_t on the ray with $t = \|x - y_t\|_2 \le \frac{1}{4(k-1)^2}$. For all finite sets X with

$$\min_{z \in X} \|x - z\|_2 \le \frac{1}{4(k-1)^2}$$

this implies

$$|p|_{\infty,X} \ge \frac{1}{2} = \frac{1}{2}p(x).$$

This argument needs that the ray from x to z is contained in the domain we want to focus on.

Definition 8.46. *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 8.47. *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 $p \in \mathcal{P}_k$. □

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 8.47, this map is injective on $V := \mathcal{P}_k$, and we can proceed exactly as in the univariate case to get

Theorem 8.48. *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 (8.44) 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 (8.44). This will turn out to work for domains satisfying

Definition 8.49. *A domain $\Omega \subset \mathbb{R}^d$ 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 8.50. *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.* □

8.12 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 (8.40) 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:

$$\text{Minimize} \sum_{x_j \in X_{x,\delta}} (f(x_j) - p(x_j))^2 \Phi_\delta(x - x_j) \text{ over all } p \in \mathcal{P}_k$$

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

$$\begin{array}{rcl} X_{x,\delta} & := & \{x_k \in X \ : \ \|x - x_k\|_2 < \delta\} \\ J_{x,\delta} & := & \{j \ : \ x_j \in X_{x,\delta}\} \end{array} \tag{8.51}$$

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

$$\text{Minimize } \|\sqrt{W}(F - Pb)\|_2^2 \text{ over } b \in \mathbb{R}^Q$$

where

$$\begin{array}{rcl} F & := & (f(x_j), \ j \in J)^T \in \mathbb{R}^{|X_{x,\delta}|} \\ P & := & (p_\ell(x_j))_{j \in J, \, 1 \le \ell \le Q} \\ b & := & (b_1, \ldots, b_Q)^T \in \mathbb{R}^Q \\ W & := & (\delta_{jk} \Phi_\delta(x - x_k))_{j,k \in J} \end{array}$$

and it is a standard least–squares problem approximationg $\sqrt{W} f_X$ by $\sqrt{W} P_X b$ for $b \in \mathbb{R}^Q$. 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 $\mathcal{P}_k$–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

$$p(x) = \sum_{j \in J} a_j(x) p(x_j) \text{ for all } p \in \mathcal{P}_k \tag{8.52}$$

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

$$\sum_{j \in J} a_j^2(x) \frac{1}{\Phi_\delta(x - x_j)}. \tag{8.53}$$

Theorem 8.54. *If the set $X_{x,\delta}$ of (8.51) is $\mathcal{P}_k$–unisolvent, the moving least–squares problem has a unique solution. It coincides with the solution of the minimization of (8.53) under the constraints (8.52) 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 $b^* \in \mathbb{R}^Q$ 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 $\mathcal{P}_k$ 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 $y \in \mathbb{R}^d$ 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 (8.52) 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 (8.53) is

positive definite, and thus it attains its unique minimum on all affine subspaces like the one defined by (8.52). 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 (**??**). 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 $c \in \mathbb{R}^{|J|}$ with $P_X^T c = 0$, and by the factorization lemma 11.10 we get **Lagrange multipliers** $\lambda(x) \in \mathbb{R}^Q$ with $a(x)^T W^{-1} c = \lambda^T(x) P_X^T c$ for all $c \in \mathbb{R}^{|J|}$. 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 8.55. *If all sets $X_{x,\delta}$ for arbitrary $x \in \Omega$ and fixed δ are $\mathcal{P}_k$–unisolvent, then the solution of the moving least–squares approcimation is as smooth as the kernel Φ_δ.*

Proof: Due to global $\mathcal{P}_k$–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

$$|J| \leq (4c_1c_3 + 1)^d.$$

Altogether, we see that moving least–squares realize stable local polynomial reproduction.

8.13 Bramble–Hilbert Lemma

We now leave the stable local polynomial reproduction part and go back to (8.42). 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 $p \in \mathcal{P}_M$ 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

$$\|u\|_{L_\infty(\Omega)} \leq Ch^{M-d/2}|u|_{W_2^M(\Omega)}.$$

Clearly, an inequality like this cannot hold unless the function u is replaced by something like $u - p^*$ for some polynomial $p^* \in \mathcal{P}_M$, 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 8.56. *Let Ω be a bounded cube in $\mathbb{R}^d$ of maximal sidelength s. Then for each function u in $W_2^1(\Omega)$ there is a constant $\gamma(u)$ such that*

$$\|u - \gamma(u)\|_{L_2(\Omega)} \leq s|u|_{W_2^1(\Omega)} \tag{8.57}$$

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

$$\|v\|^2_{L_2(\Omega)} \leq s^2 \int_\Omega \left(\frac{\partial v}{\partial x_1}(y) \right)^2 dy = s^2 |v|^2_1.$$

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 (8.57) 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 8.56 is

Lemma 8.58. *("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 $p(u) \in \mathcal{P}_M$ such that

$$\|u - p(u)\|_{L_2(\Omega)} \leq C s^M |u|_{W_2^M(\Omega)}.$$

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 [BS02] 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 11.10, because the operator $Id - T_M$ with T_M being the Taylor projector of order M, vanishes on the kernel $\mathcal{P}_M$ 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

$$\|u - T_M(u)\|_{L_\infty(\Omega)} \leq C s^{M-d/2} |u|_{W_2^M(\Omega)} \tag{8.59}$$

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 (8.59). By a factorization argument, one can get a result like

$$\|u - T_M(u)\|_{L_\infty(\Omega)} \leq C |u|_{W_2^M(\Omega)}$$

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

$$\|u - T_M(u)\|^2_{L_\infty(\Omega_s)} = \|v - T_M(v)\|^2_{L_\infty(\Omega)} \leq C |v|^2_{W_2^M(\Omega)} = C s^{2M-d} |u|^2_{W_2^M(\Omega_s)}$$

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 $M > d/2$.

8.14 Globalization

From (8.59) 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

$$\begin{aligned} \|T_M u\|_{\infty,\Omega} &\leq C \|T_M u\|_{\infty,X} \\ &\leq C \left(\|T_M u - u\|_{\infty,X} + \|u\|_{\infty,X} \right) \\ &\leq C \left(\|T_M u - u\|_{\infty,\Omega} + \|u\|_{\infty,X} \right) \\ &\leq C \left(C s^{M-d/2} |u|_{W_2^M(\Omega)} + \|u\|_{\infty,X} \right) \end{aligned}$$

and proceeds via

$$\begin{aligned}\|u\|_{L_\infty(\Omega)} &\leq \|u - T_M u\|_{L_\infty(\Omega)} + \|T_M u\|_{L_\infty(\Omega)} \\ &\leq Cs^{M-d/2}|u|_{W_2^M(\Omega)} + C\left(Cs^{M-d/2}|u|_{W_2^M(\Omega)} + \|u\|_{\infty,X}\right) \\ &\leq C\left(s^{M-d/2}|u|_{W_2^M(\Omega)} + \|u\|_{\infty,X}\right)\end{aligned}$$

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 [NWW06]
  $|u|_{W_q^{|\alpha|}} \leq ch^{m-|\alpha|-d(1/p-1/q)_+}|u|_{W_p^m},\ u(X_h) = \{0\}$
  for $0 \leq |\alpha| \leq m > d/p$,
- Wendland & Rieger Num. Math. 2005 [WR05]
  $|u|_{W_q^{|\alpha|}} \leq C\left(h^{m-|\alpha|-d(1/p-1/q)_+}|u|_{W_p^m} + h^{-|\alpha|}\|u\|_{\infty,X_h}\right)$
  for $0 \leq |\alpha| \leq m > d/p$,
- Madych JAT 2006 [Mad06]
  $\|u\|_{L_q} \leq C\left(h^{m-d(1/p-1/q)_+}|u|_{W_p^m} + h^{d/max(p.q)}\|u\|_{\ell_p,X_h}\right)$
  for $0 \leq m > d/p$.

Finally, there is a more sophisticated inequality due to Christian Rieger and Barbara Zwicknagl [RZ06]. 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

$$\|u\|_{W_q^m(\Omega)} \leq C\left(\exp\left(c_1\frac{\log(c_2 h)}{\sqrt{h}}\right)\|u\|_{\mathcal{H}} + h^{-|\alpha|}\|u\|_{\ell_q(X_h)}\right) \tag{8.60}$$

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.

8.15 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 $\mathcal{N}$ for a kernel K is continuously embedded in some Sobolev space of order m, i.e.

$$\|u\|_{W_2^m(\Omega)} \leq C\|u\|_{\mathcal{N}} \text{ for all } u \in \mathcal{N}. \tag{8.61}$$

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

$$\|Q_{X_h}u\|_{W_2^m(\Omega)} \leq C\|Q_{X_h}u\|_{\mathcal{N}} \leq C\|u\|_{\mathcal{N}} \text{ for all } u \in \mathcal{N},$$

and we use Sobolev embedding from (8.61) in one of the sampling inequalities of 8.14 when applying them to the difference $u - Q_{X_h}u$. This yields in the first case

$$\begin{aligned} |u - Q_{X_h}u|_{W_q^{|\alpha|}} &\leq ch^{m-|\alpha|-d(1/2-1/q)_+}|u - Q_{X_h}u|_{W_p^m} \\ &\leq ch^{m-|\alpha|-d(1/2-1/q)_+}\|u - Q_{X_h}u\|_{\mathcal{N}} \\ &\leq Ch^{m-|\alpha|-d(1/2-1/q)_+}\|u\|_{\mathcal{N}} \end{aligned}$$

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

$$\begin{aligned} \|u - Q_{X_h}u\|_{L_q(\Omega)} &\leq Ch^{m-d(1/2-1/q)_+}|u - Q_{X_h}u|_{W_2^m} \\ &\leq Ch^{m-d(1/2-1/q)_+}\|u\|_{W_2^m} \\ &\leq Ch^{m-d(1/2-1/q)_+}\|u\|_{\mathcal{N}} \end{aligned}$$

for $0 \leq m > d/2,\ 1 \leq q \leq \infty$.

In the situation of the refined inequality (8.60), 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.

9 Construction of Kernels

For this section, we only present some additional material. The standard procedure will be like the one in the book [Wen05] 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 12.3.

9.1 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 $\mathbb{R}^d$.

9.1.1 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.

9.1.2 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 $L_2(\mathbb{R}^d)$.

9.1.3 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.

9.2 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 3.7, 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

$$N_\Omega := \{ f_a \ :\ y \to a^T y \ :\ a \in LH(\Omega) \} \tag{9.1}$$

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 (9.1) 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 3.7. 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

9.3 Translation–Invariant Kernels on $\mathbb{R}^d$

We now go back to section 3.2 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 9.2. *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 9.3. *(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

$$\begin{aligned} & \sum_{j,k=1}^{N} a_j \overline{a}_k K(x_j, x_k) \\ = \; & (2\pi)^{-d/2} \int_{\mathbb{R}^d} \sum_{j,k=1}^{N} a_j \overline{a}_k e^{i(x_j - x_k)^T \omega} c(\omega) d\omega \\ = \; & (2\pi)^{-d/2} \int_{\mathbb{R}^d} \left| \sum_{j=1}^{N} a_j e^{i x_j^T \omega} \right|^2 c(\omega) d\omega \geq 0 \end{aligned}$$

and assume that it vanishes. Then the product of the generalized trigonometric polynomial

$$p(\omega) := p_{a,X}(\omega) := \sum_{j=1}^{N} a_j e^{i x_j^T \omega} \tag{9.4}$$

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 9.5. *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

$$D^\beta p(0) = \sum_{j=1}^{N} a_j (i x_j)^\beta, \; \beta \in \mathbb{N}_0^d$$

vanish, and this means that all

$$\sum_{j=1}^{N} a_j x_j^\beta, \; \beta \in \mathbb{N}_0^d$$

are zero. If we pick a single index j, $1 \leq j \leq N$, we can find a multivariate polynomial $q_j(x)$ with the Lagrange property $q_j(x_k) = \delta_{jk}$, $1 \leq j, k \leq N$, 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 $j,\ 1 \leq j \leq N$. □

Theorem 9.6. *If c is a nonnegative even and integrable function on $\mathbb{R}^d$ which is positive on an open set, its Fourier transform is a real–valued symmetric translation–invariant positive definite kernel K_c on $\mathbb{R}^d$.* □

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 9.7. *Let K be a translation–invariant symmetric real–valued kernel on $\mathbb{R}^d$ whose Fourier transform exists, is even and nonnegative and integrable on $\mathbb{R}^d$ 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

$$K(x-y) = (1 - \|x-y\|_2)_+^4 (1 + 4\|x-y\|_2),\ x, y \in \mathbb{R}^d$$

has positive Fourier transform in $\mathbb{R}^d$ for $d \leq 3$, but this is not at all clear. We postpone such kernels for a while.

9.4 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

$$K(x - y) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} \left(1 + \|\omega\|_2^2\right)^{-m} e^{i(x-y)^T\omega} d\omega \tag{9.8}$$

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 9.9. *The reproducing kernel for Sobolev space $W_2^m(\mathbb{R}^d)$ for $m > d/2$ is given by (9.8) and turns out to have the explicit radial representation*

$$\frac{2^{1-m}}{(m-1)!} \|x - y\|_2^{m-d/2} K_{-m+d/2}(\|x - y\|_2) \tag{9.10}$$

where K_ν is the modified Bessel function of order ν.

We postpone the explicit calculation ending with the above formula, but in Figure 6 on page 5 we already presented a plot of the kernels $r^\nu K_\nu(r)$ after a rescaling to attain 1 at zero. In section 12.7.8 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.

9.5 Native Spaces of Translation–Invariant Kernels

After we have seen the special case of a kernel that directly leads to a global Sobolev space, we now go back to the more general situation of a translation–invariant kernel K_c generated from an even, 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 (3.11) in section 3.3. But we can now use Fourier transforms on it and get first

$$\begin{aligned}
(\hat{f}_{a,X})(\omega) &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} f_{a,X}(x) e^{-ix^T\omega} dx \\
&= (2\pi)^{-d/2} \sum_{j=1}^{M} a_j e^{-ix_j^T\omega} \int_{\mathbb{R}^d} K(x - x_j) e^{i(x_j - x)^T\omega} dx \\
&= \hat{K}(\omega) \sum_{j=1}^{M} a_j e^{-ix_j^T\omega} \\
&=: \hat{K}(\omega)\, p_{a,X}(\omega)
\end{aligned}$$

and then

$$\begin{aligned}
(f_{a,X}, f_{b,Y}) &= \sum_{j=1}^{M} \sum_{k=1}^{N} a_j b_k K(y_k, x_j) \\
&= (2\pi)^{-d/2} \sum_{j=1}^{M} \sum_{k=1}^{N} a_j b_k \int_{\mathbb{R}^d} \hat{K}(\omega) e^{i(y_k - x_j)^T\omega} d\omega \\
&= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{K}(\omega) \overline{\sum_{j=1}^{M} a_j e^{-ix_j^T\omega}} \sum_{k=1}^{N} b_k e^{iy_k^T\omega} d\omega \\
&= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{\hat{f}_{a,X}(\omega)\overline{\hat{f}_{b,Y}(\omega)}}{\hat{K}(\omega)} d\omega.
\end{aligned}$$

Thus we can read off the right inner product of the native space. We define

$$\mathcal{F}_K := \{f \in L_2(\mathbb{R}^d) \ : \ \int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{\hat{K}(\omega)} d\omega < \infty\}$$

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

$$\begin{aligned}(f, f_{a,X})_K &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{\hat{f}(\omega)\overline{\hat{f}_{a,X}(\omega)}}{\hat{K}(\omega)} d\omega \\ &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{f}(\omega)\overline{p_{a,X}(\omega)} d\omega \\ &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{f}(\omega) \sum_{j=1}^{M} a_j e^{ix_j^T \omega} d\omega \\ &= \sum_{j=1}^{M} a_j f(x_j)\end{aligned}$$

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 9.11. *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

$$\hat{K}(\omega) \leq C(1 + \|\omega\|_2^2)^{-m} \text{ for all } \omega \in \mathbb{R}^d. \tag{9.12}$$

Theorem 9.13. *If a translation–invariant symmetric positive definite kernel K on $\mathbb{R}^d$ satisfies (9.12) 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 (9.12) 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 $n < 2m - d/2$.

9.6 Construction of Positive Definite Radial Functions on $\mathbb{R}^d$

This subsection contains tools from [Wu95] as generalized in [SW96] 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.

9.6.1 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:

$$\begin{aligned}
\widehat{\Phi}(\omega) &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \Phi(x) e^{-ix\cdot\omega} dx \\
&= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \phi(\|x\|_2) e^{-ix\cdot\omega} dx \\
&= (2\pi)^{-d/2} \int_0^\infty \phi(r) r^{d-1} \int_{\|y\|_2=1} e^{-ir\|\omega\|_2 y\cdot\frac{\omega}{\|\omega\|_2}} dy dr.
\end{aligned}$$

This contains the function $F(r\|\omega\|_2, d)$ defined in (12.29) 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 (12.31) via a Bessel function, we get the very important equation

$$\begin{aligned} \widehat{\Phi}(\omega) &= (2\pi)^{-d/2}\sigma_{d-2}\int_0^\infty \phi(r) r^{d-1} \frac{\Gamma(\frac{d-1}{2})\Gamma(\frac{1}{2})}{(r\|\omega\|_2/2)^{(d-2)/2}} J_{(d-2)/2}(r\|\omega\|_2)dr \\ &= \|\omega\|_2^{-(d-2)/2}\int_0^\infty \phi(r) r^{d/2} J_{(d-2)/2}(r\|\omega\|_2)dr. \end{aligned} \tag{9.14}$$

that allows the Fourier transform of a radial function to be written as a univariate **Hankel transform**. Equation (9.14) 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

$$\sqrt{\frac{\pi}{2z}} J_{-1/2}(z) = \frac{\cos z}{z}. \tag{9.15}$$

This equation is not directly compatible with (12.30), because the latter does not exist for $\nu = -1/2$. But we can use the usual power series representation (12.32) of Bessel functions to get (9.15) from (12.34).

9.6.2 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 (12.38) and get

$$\begin{aligned} \widehat{\chi_1}(\omega) &= \|\omega\|_2^{-(d-2)/2}\int_0^1 r^{d/2} J_{(d-2)/2}(r\|\omega\|_2)dr \\ &= \|\omega\|_2^{-d/2} J_{d/2}(\|\omega\|_2). \end{aligned} \tag{9.16}$$

Considered as a univariate radial function, this is an entire analytic function of exponential type that we shall meet again later. Figure 29 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 9.17. *[FLW06]. 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 $\mathbb{R}^d$.* □

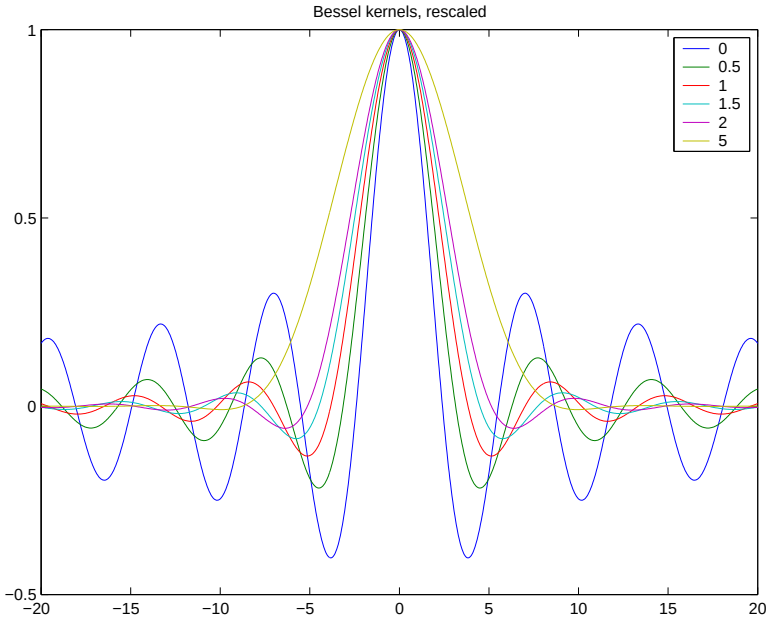

Figure 29: The Bessel kernels $r^{-\nu}J_\nu(r)$

From (12.31) 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 9.17 guarantees positive definiteness only on $\mathbb{R}^k$ for $0\le k\le d-2$. This was first observed in [FLW06].

Theorem 9.18. *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 (9.4) 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 [FLW06] 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**) $\Delta u+c^2u=0$ for $d>1$ dimensions.

A second application of the Hankel transform is the proof of (9.10) in Theorem 9.9. A more explicit and direct proof is in [Wen05] on pages 76–77, but

we cite (12.51) 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 (9.14) applied to $\phi(r) = (r^2+c^2)^{-m}$ this yields

$$\begin{aligned}\hat{\Phi}(\omega) &= \frac{\|\omega\|_2^{m-1-(d-2)/2} c^{-m+d/2}}{2^{m-1}\Gamma(m)} K_{-m+d/2}(c\|\omega\|_2) \\ &= \left(\frac{\|\omega\|_2}{c}\right)^{m-d/2} \frac{2^{1-m}}{(m-1)!} K_{-m+d/2}(c\|\omega\|_2).\end{aligned}$$

9.6.3 Change of Variables

We now introduce $t = r^2/2$ as a new variable, writing a radial basis function Φ as

$$\Phi(\cdot) = \phi(\|\cdot\|_2) = f(\|\cdot\|_2^2/2), \tag{9.19}$$

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 (9.14) 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 (12.33). If we substitute $t = s^2/2$, we find

$$\begin{aligned} \widehat{\Phi}(\omega) &= \int_0^\infty f(t) t^{\frac{d-2}{2}} H_{\frac{d-2}{2}}\left(t \cdot \frac{\|\omega\|^2}{2}\right) dt \\ &=: \left(F_{\frac{d-2}{2}} f\right)(\|\omega\|^2/2) \end{aligned} \tag{9.20}$$

with the general operator

$$(F_\nu f)(r) := \int_0^\infty f(t) t^\nu H_\nu(tr) dt. \tag{9.21}$$

Theorem 9.22. *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 $\wr(t^{-1/2})$ behaviour for $t \to \infty$ due to (12.42), 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

$$f(t)t^\nu \in L_1(\mathbb{R}_{>0}). \tag{9.23}$$

Note that both F_ν and H_ν generalize to arbitrary $\nu \in \mathbb{R}$, provided that certain restrictions on f like (9.23) 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 (9.19), 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 (9.19) 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 (12.33) of H_ν and (12.25) of the Gamma function.

9.6.4 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

$$I_\alpha(f)(r) := \int_0^\infty f(s)\frac{(s-r)_+^{\alpha-1}}{\Gamma(\alpha)}ds \tag{9.24}$$

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 (9.24). The identity (12.26) allows a direct proof of the property

$$I_\alpha \circ I_\beta = I_{\alpha+\beta} \tag{9.25}$$

for all $\alpha,\ \beta > 0$ by application of Fubini's theorem. Differentiation and integration by parts imply

$$\begin{aligned} I_{-1}^n \circ I_\alpha &= I_{\alpha-n} \quad 0 < \alpha < n \\ I_{n+\alpha} \circ I_{-1}^n &= I_\alpha \quad \alpha > 0, n > 0. \end{aligned}$$

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 (9.25) holds for all $\alpha, \beta \in \mathbb{R}$ 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 9.26. *The operators I_α on $\mathcal{S}_{rad}$ form an abelian group under composition which is isomorphic to $\mathbb{R}$ 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.

$$\begin{array}{rcll} I_\alpha(f(t+x))(r) & = & I_\alpha(f(t))(r+x) & \alpha \in \mathbb{R},\ x \geq 0 \\ I_\alpha(f(tx))(r) & = & x^{-\alpha} I_\alpha(f(t))(rx) & \alpha \in \mathbb{R},\ x \geq 0 \\ I_\alpha(e^{-st})(r) & = & s^{-\alpha} e^{-sr} & \alpha \in \mathbb{R},\ s > 0 \\ I_\alpha(t^{-\beta}\Gamma(\beta))(r) & = & r^{-(\beta-\alpha)}\Gamma(\beta-\alpha) & \beta > 0,\ \alpha < \beta \\ I_\alpha((x+t)^{-\beta}\Gamma(\beta))(r) & = & (x+r)^{-(\beta-\alpha)}\Gamma(\beta-\alpha) & \beta > 0,\ \alpha < \beta,\ x > 0 \\ I_\alpha\left(\dfrac{(s-t)_+^{\beta-1}}{\Gamma(\beta)}\right)(r) & = & \dfrac{(s-r)_+^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)} & \beta > 0,\ \alpha+\beta > 0 \end{array}$$

We shall make specific use of the "semi-integration" operator and its inverse, the "semi-differentiation", as given by

$$\begin{array}{rcl} I_{1/2}(f)(r) & = & \displaystyle\int_r^\infty \frac{f(s)}{\sqrt{\pi(s-r)}} ds \\ I_{-1/2}(f)(r) & = & \displaystyle -\int_r^\infty \frac{f'(s)}{\sqrt{\pi(s-r)}} ds \\ & = & I_{1/2} \circ I_{-1}(f)(r), \end{array} \tag{9.27}$$

that are inverses of each other.

A very simple representation of the operators I_α is possible via the **Laplace transform**

$$L(\varphi)(r) := \int_0^\infty \varphi(s) e^{-rs} ds \tag{9.28}$$

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).

9.6.5 Basic Transitions

The main advantage of $\mathcal{S}_{rad}$ and the definition (9.21) of the radial Fourier transform using (9.20) 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 (12.35) to get

$$\begin{aligned} -\frac{d}{dr}F_\nu(f)(r) &= (I_{-1}\circ F_\nu)(f)(r) \\ &= -\frac{d}{dr}\int_0^\infty f(t)t^\nu H_\nu(rt)dt \\ &= -\int_0^\infty f(t)t^\nu \frac{d}{dr}H_\nu(rt)dt \\ &= \int_0^\infty f(t)t^{\nu+1}H_{\nu+1}(rt)dt \\ &= F_{\nu+1}(f)(r). \end{aligned} \tag{9.29}$$

Going back to $\nu = (d-2)/2$, we see that the $(d+2)$-variate Fourier transform of a radial function after the substitution (9.19) is nothing else than the negative univariate derivative of the d-variate Fourier transform after (9.19). 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 (12.36) to get another identity on tempered functions:

$$\begin{aligned} F_\nu(-f')(r) &= \int_0^\infty -f'(s)s^\nu H_\nu(sr)ds \\ &= \int_0^\infty f(s)s^{\nu-1}H_{\nu-1}(sr)dsdt \\ &= = F_{\nu-1}(f)(r). \end{aligned} \tag{9.30}$$

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 (9.27). We shall treat this problem in general, comparing two arbitrary instances F_ν and F_μ.

9.6.6 Identities for Transforms, First Version

We can easily evaluate the action of the Fourier operator on the Laplace transform as

$$\begin{aligned}
F_\nu(L(\varphi))(r) &= \int_0^\infty s^\nu H_\nu(sr) \int_0^\infty \varphi(t) e^{-st} dt ds \\
&= \int_0^\infty \varphi(t) \int_0^\infty s^\nu H_\nu(sr) e^{-st} ds dt \\
&= \int_0^\infty \varphi(t) t^{-\nu-1} \int_0^\infty x^\nu H_\nu(xr/t) e^{-x} dx dt \\
&= \int_0^\infty \varphi(t) t^{-\nu-1} e^{-r/t} dt \\
&= \int_0^\infty \varphi(1/s) s^{\nu-1} e^{-sr} ds \\
&= L\left(\varphi(1/\cdot)(\cdot)^{\nu-1}\right).
\end{aligned}$$

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.

9.6.7 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 9.31. *For $\nu > \mu > -1$ and all $s, r \geq 0$ we have*

$$F_\mu \left(\frac{s^{-\nu}(s-\cdot)_+^{\nu-\mu-1}}{\Gamma(\nu-\mu)} \right)(r) \;=\; H_\nu(rs).$$

Proof: We directly calculate the assertion and use (12.37) from page 252. 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 9.32. *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 (12.45) or [AS70] p. 487, 11.4.41) after substitutions of the type $t = s^2/2$. In detail, we have

$$\begin{aligned}
& \left(F_\mu H_\nu\left(\tfrac{u^2}{2}\,\cdot\right)\right)\left(\tfrac{r^2}{2}\right) \\
= & \int_0^\infty t^\mu H_\mu\left(\tfrac{r^2}{2}\,t\right) H_\nu\left(\tfrac{u^2}{2}\,t\right) dt \\
= & \int_0^\infty \left(\tfrac{s^2}{2}\right)^\mu \cdot s \cdot H_\mu\left(\tfrac{r^2}{2}\,\cdot\tfrac{s^2}{2}\right) H_\nu\left(\tfrac{u^2}{2}\,\cdot\tfrac{s^2}{2}\right) ds \\
= & \int_0^\infty 2^{-\mu} s^{2\mu+1} \left(\tfrac{rs}{2}\right)^{-\mu} \left(\tfrac{us}{2}\right)^{-\nu} J_\nu(us) ds \\
= & \ 2^\nu r^{-\mu} r^{-\nu} \int_0^\infty s^{\mu-\nu+1} J_\mu(rs) J_\nu(us) ds \\
= & \ \frac{2^\nu r^{-\mu} u^{-\nu} 2^{\mu-\nu+1} r^\mu (u^2-r^2)_+^{\nu-\mu-1}}{u^\nu \Gamma(\nu-\mu)} \\
= & \ \frac{1}{\Gamma(\nu-\mu)} \left(\tfrac{u^2}{2}\right)^{-\nu} \left(\tfrac{u^2}{2} - \tfrac{r^2}{2}\right)_+^{\nu-\mu-1} .
\end{aligned}$$

□

The above result can be used to derive the d–variate Fourier transform of the kernel

$$K(x,y) := H_\nu\left(\frac{c^2\|x-y\|_2^2}{4}\right) = \left(\frac{c\|x-y\|_2}{4}\right)^{-\nu} J_\nu\left(\frac{c\|x-y\|_2}{2}\right) \qquad (9.33)$$

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

$$\left(F_{(d-2)/2}H_\nu\left(\frac{c^2}{2}\cdot\right)\right)\left(\frac{\|\omega\|_2^2}{2}\right) = \frac{\left(\frac{c^2}{2}\right)^{-\nu}\left(\frac{c^2}{2}-\frac{\|\omega\|_2^2}{2}\right)_+^{\nu-d/2}}{\Gamma(\nu-d/2+1)} \tag{9.34}$$

proving

Theorem 9.35. *If $2\nu+2>d$ holds, the scaled Bessel kernel (9.33) is positive definite on $\mathbb{R}^d$ and has the compactly supported Fourier transform (9.34) due to Theorems 9.22 and 9.7.*

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 9.36. *Let $\nu > \mu > -1$. Then for all tempered radial test functions $f \in \mathcal{S}_{rad}$ we have*

$$F_\mu \circ F_\nu = I_{\nu-\mu} \tag{9.37}$$

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 9.32 to obtain

$$\begin{aligned}
&\int_0^\infty H_\mu(tr)t^\mu \int_0^\infty H_\nu(st)s^\nu f(s)dsdt\\
=\ &\int_0^\infty s^\nu f(s)\int_0^\infty t^\mu H_\mu(tr)H_\nu(ts)dt\,ds\\
=\ &\int_0^\infty s^\nu f(s)\cdot F_\mu(H_\nu(s\cdot))(r)ds\\
=\ &\int_0^\infty f(s)\frac{(s-r)_+^{\nu-\mu-1}}{\Gamma(\nu-\mu)}\,ds = (I_{\nu-\mu}f)(r).
\end{aligned}$$

□

By the above theorems it is easy to see that

$$I_\alpha H_\nu = H_{\nu-\alpha}$$

for all $\alpha < \nu+1$, generalizing (12.35).

9.6.8 Wendland's Functions

Due to a result of Askey [Ask73] 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 [Wen05]

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 (9.30) 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

$$\mu \geq \lfloor d/2 \rfloor + 1 + k. \tag{9.38}$$

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 (9.38). Let us do a straightforward evaluation. This yields

$$\begin{aligned} I_k A_\mu(r) &= \int_0^\infty (1 - \sqrt{2s})_+^\mu \frac{(s-r)_+^{k-1}}{(k-1)!} \\ &= \int_{\sqrt{2r}}^1 t(1-t)^\mu \frac{(t^2/2 - r)_+^{k-1}}{(k-1)!} \\ &= \int_x^1 t(1-t)^\mu \frac{(t^2 - x^2)_+^{k-1}}{(k-1)! 2^{k-1}} \end{aligned} \tag{9.39}$$

for $0 \leq r \leq 1/2$ or $0 \leq x = \sqrt{2r} \leq 1$. 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 $d \leq 3$ 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 [Wen95] 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 [Gas75], 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 [Wen95] then imply the asymptotic behaviour

$$F_\nu I_k A_\mu(r^2/2) = F_{\nu+k}A_\mu(r^2/2) \geq cr^{-d-2k-1}$$

for the necessary values of μ from (9.38).

9.7 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 5.4, but here we fix the space $\mathcal{P}$ to be the space $\mathcal{P}_m^d$ of d–variate polynomials of order at most m. Fur-

thermore, we use the notion of Fourier transforms of functionals as provided in section 12.6.

Assumption 9.40. *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

$$\widehat{\Phi}(\omega) = \wr(\|\omega\|^{-d-\beta_0}) \tag{9.41}$$

with some real value β_0 for ω near zero, and it must have the behavior

$$\widehat{\Phi} \in L_1 \text{ near infinity.} \tag{9.42}$$

Then define $m := \max(0, \lfloor \beta_0 \rfloor) \geq 0$ to get the restriction

$$\beta_0 < 2m \tag{9.43}$$

that will often occur later. Finally, let the usual bilinear form on L be representable by

$$(\lambda_{a,X}, \lambda_{b,Y})_\Phi = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{\Phi}(\omega) \sum_{j=1}^{M} \sum_{k=1}^{N} a_j \overline{b_k} e^{i(x_j - y_k)\cdot\omega} d\omega, \tag{9.44}$$

where the functionals $\lambda_{a,X} \in L$ satisfy the **moment conditions** *(5.4) in the form*

$$\lambda_{a,X}(\mathcal{P}_m^d) = \{0\}, \tag{9.45}$$

and thus we may use the notation $(\mathcal{P}_m^d)_{\mathbb{R}^d}^\perp$ for L.

Lemma 9.46. *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 (9.45), we can use Example 12.22 to get our result. □

Theorem 9.47. *Under the above assumptions the function $\Phi(x-y)$ is conditionally positive definite of order $\geq m$ on $\mathbb{R}^d$.*

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 (9.44) is of order $\wr(\|\omega\|^{2m-d-\beta_0})$ near zero, and the integral is well-defined due to (9.43) and (9.42). Nonnegativity of $\widehat{\Phi}$ proves that the bilinear form is positive semidefinite. The rest is as in the proofs of Theorems 9.6 and 12.8. □

The reader should be aware that we did not assume $\widehat{\Phi}$ to be the usual Fourier transform. We thus cannot use equations (12.7) or (12.12), but we have the general identity

$$\sum_{j=1}^{M}\sum_{k=1}^{N} a_j b_k \Phi(x_j - y_k) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{\Phi}(\omega) \sum_{j=1}^{M}\sum_{k=1}^{N} a_j \overline{b_k} e^{i(x_j - y_k)\cdot\omega} d\omega.$$

that is identical to (9.44) and is valid for all functionals in L due to Assumption 9.40. It will nicely serve as a substitute for (12.12), but note that it does not allow single point-evaluation functionals in the left-hand side.

9.8 Examples

We now present special cases of (9.44) for radial kernels

$$K(x,y) = \Phi(x-y) = \phi(\|x-y\|_2),\ x,y \in \mathbb{R}^d$$

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

$$\phi(r) := (c^2 + r^2)^{\beta/2},\ r \geq 0,\ c > 0,\ \beta \in \mathbb{R} \setminus 2\mathbb{N}_0$$

we get the function

$$\hat{\phi}(s) = \frac{2^{1+\beta/2}}{\Gamma(-\beta/2)} \left(\frac{s}{c}\right)^{-\frac{\beta+d}{2}} K_{\frac{\beta+d}{2}}(cs),\ s \in \mathbb{R}$$

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 (9.44), 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

$$\phi(r) := (-1)^{\lceil \beta/2 \rceil} r^\beta, \ r \geq 0, \ \beta \in \mathbb{R}_{>0} \setminus 2\mathbb{N}$$

we get the positive function

$$\hat{\phi}(s) = \frac{2^{\beta+d/2}\Gamma((\beta+d)/2)}{(-1)^{\lceil \beta/2 \rceil}\Gamma(-\beta/2)} s^{-\beta-d}, \ s \in \mathbb{R}$$

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

$$\phi(r) := (-1)^{k+1} r^{2k} \log r, \ r \geq 0, \ k \in \mathbb{N}$$

we get

$$\hat{\phi}(s) = 2^{2k-1+d/2} s^{-2k-d}, \ s \in \mathbb{R}$$

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**.

9.9 Connection to $L_2(\mathbb{R}^d)$

We now go back to Definition ?? of the native space via (??) and Corollary ??as

$$\mathcal{N}_{K,\Xi} := \mathcal{P}_m^d + \overline{G} = \mathcal{P}_m^d + \overline{F}.$$

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 5.4 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 9.40 makes sure that the mapping

$$\mathcal{L} \ : \ \lambda \mapsto \widehat{\lambda}\sqrt{\widehat{\Phi}}, \ \ L = (\mathcal{P}_m^d)_{\mathbb{R}^d}^{\perp} \to L_2(\mathbb{R}^d)$$

is well-defined. Indeed, the function $\mathcal{L}(\lambda)$ is in L_2 near infinity due to (9.42), and it is continuous around zero due to (9.43), since $\widehat{\lambda}$ has a zero of order at least m at the origin.

With the results of the previous section, (9.44) takes the form

$$(\lambda_{a,X}, \lambda_{b,Y})_{\Phi} = (2\pi)^{-d/2}(\mathcal{L}\lambda_{a,X}, \mathcal{L}\lambda_{b,Y})_{L_2(\mathbb{R}^d)}. \tag{9.48}$$

Theorem 9.49. *Let Assumption 9.40 be satisfied, and let m be minimal with respect to (9.43). 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 (9.44) 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 12.5. 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

$$u(x) = \sum_{j \in \mathbb{Z}^d} u(jh) \operatorname{Sinc}_d \left(\frac{x - jh}{h} \right), \quad x \in \mathbb{R}^d$$

where

$$\operatorname{Sinc}_d(x_1, \ldots, x_d) = \prod_{j=1}^{d} \frac{\sin \pi x_j}{\pi x_j},$$

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 $\lambda_u \in \operatorname{clos}(L)$. 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 $p \in \mathcal{P}_m^d$ 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 (9.41) on $\widehat{\Phi}$ and the minimality of m in (9.43) 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 $\operatorname{clos}(L)$ by defining the function

$$f_{\lambda_u}(x) := (\lambda_u, \delta_{x,\Xi})_\Phi, \ x \in \mathbb{R}^d$$

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 (9.44) in the form

$$(\mu, \lambda)_{\Phi_\delta} := (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{\Phi}_\delta(\omega) \widehat{\lambda}(\omega) \overline{\widehat{\mu}(\omega)} d\omega$$

as a definition of an inner product, but we do not need Φ_δ explicitly.

Now we approximate a given $f \in L_2(\mathbb{R}^d)$ 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

$$\begin{aligned}\|\widehat{\lambda}(\sqrt{\widehat{\Phi}_\delta} - \sqrt{\widehat{\Phi}})\|_2^2 &= \|\widehat{\lambda} \cdot \sqrt{\widehat{\Phi}_\delta}(1 - \sqrt{\widehat{\Phi}/\widehat{\Phi}_\delta})\|_2^2 \\ &\leq \sum_k \int_{I_k} |\widehat{\lambda}(\omega)|^2 \widehat{\Phi}_\delta(\omega) d\omega.\end{aligned}$$

The full integral

$$\int_{\mathbb{R}^d} |\widehat{\lambda}(\omega)|^2 \widehat{\Phi}_\delta(\omega) d\omega = \|\widehat{\lambda} \cdot \sqrt{\widehat{\Phi}_\delta}\|_2^2$$

can be bounded independent of δ, because it approximates $\|f\|_2^2$. Thus we are able to pick δ small enough to guarantee

$$\sum_k \int_{I_k} |\widehat{\lambda}(\omega)|^2 \widehat{\Phi}_\delta(\omega) d\omega \leq \varepsilon/2$$

yielding an overall bound $\|f - \sqrt{\widehat{\Phi}}\widehat{\lambda}\|_2 \leq \varepsilon$. □

9.10 Characterization of Native Spaces

We now can re–express the native space $\mathcal{N}_{K,\Xi} := \mathcal{P}_m^d + \overline{G} = \mathcal{P}_m^d + \overline{F}$. via Fourier transforms.

Theorem 9.50. *The native space $\mathcal{N}_{K,\Xi} := \mathcal{P}_m^d + \overline{G}$ for a conditionally positive definite function of order m on $\mathbb{R}^d$ satisfying Assumption 9.40 coincides with the space of all functions f on $\mathbb{R}^d$ that can be written as*

$$f_h(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{h}(\omega) \sqrt{\widehat{\Phi}(\omega)} \left(e^{ix\cdot\omega} - \sum_{j=1}^{Q} p_j(x) e^{i\xi_j \cdot \omega} \right) d\omega \tag{9.51}$$

plus polynomials from $\mathcal{P}_m^d$ and where $\hat{h} \in L_2(\mathbb{R}^d)$. The above functions are spanning the space $\overline{G}$. The bilinear form on $\overline{G}$ can be rewritten as

$$(f_g, f_h)_\Phi = (2\pi)^{-d/2} (g, h)_{L_2(\mathbb{R}^d)}. \tag{9.52}$$

Proof: We first focus on (9.51). Starting with an arbitrary $h \in L_2(\mathbb{R}^d)$ and a fixed $\mathcal{P}_m^d$-unisolvent set $\Xi \subset \mathbb{R}^d$, we mimic the technique of Riesz maps to define a function

$$f_h(x) := (\widehat{h}, \mathcal{L}\delta_{(x)})_{L_2(\mathbb{R}^d)}. \tag{9.53}$$

This is (9.51). Since

$$\lambda f_h = (\widehat{h}, \mathcal{L}\lambda)_{L_2(\mathbb{R}^d)}$$

follows easily from (9.53) for all $\lambda \in L$, we can transform this equation further into

$$\begin{aligned} \lambda f_h &= (\widehat{h}, \mathcal{L}\lambda)_{L_2(\mathbb{R}^d)} \\ &= (\mathcal{L}^{-1}\widehat{h}, \lambda)_\Phi. \end{aligned}$$

By the previous section, (9.48) with Theorem 9.49 yields that $\mathcal{L}^{-1}$ maps $L_2(\mathbb{R}^d)$ 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 **??**. Thus the above identity can be extended to

$$\begin{aligned} \lambda f_h &= (\widehat{h}, \mathcal{L}\lambda)_{L_2(\mathbb{R}^d)} \\ &= (\mathcal{L}^{-1}\widehat{h}, \lambda)_\Phi \\ &= (\mathcal{R}\mathcal{L}^{-1}\widehat{h}, \mathcal{R}\lambda)_\Phi \text{ for all } \lambda \in L \end{aligned}$$

proving

$$f_h = \mathcal{R}\mathcal{L}^{-1}\widehat{h} \in \overline{G}.$$

□

By (9.52) we also get

Corollary 9.54. *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 (9.51) 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.

9.11 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 (9.51). 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

$$\|\omega\|^{2|\alpha|}\hat{\Phi}(\omega) \leq \wr(\|\omega\|^{2|\alpha|-d-\beta_0}).$$

Near infinity, we have not yet made any assumptions about the behavior of $\hat{\Phi}$. For simplicity, we mimic (9.41) as

$$\hat{\Phi}(\omega) \leq \wr(\|\omega\|^{-d-\beta_\infty}) \text{ near } \infty. \tag{9.55}$$

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...

10 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.

10.1 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 $\Omega \subset \mathbb{R}^d$. This recovery should use data $g(x_j), 1 \le j \le M$ for a finite set $X = \{x_1, \ldots, x_M\} \subset \Omega \subset \mathbb{R}^d$.

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

$$(u_X^*)^T(x) := (1, -u_1^*(x), \ldots, -u_M^*(x))^T \in \mathbb{R}^{M+1}$$

with the Lagrange basis of (**??**) and get the special form

$$\begin{aligned} P_X^2(x) &= K(x,x) - 2\sum_{j=1}^N u_j^*(x)K(x_j,x) + \sum_{j,k=1}^N u_j^*(x)u_k^*(x)K(x_j,x_k) \\ &= (u_X^*)^T(x)A_{x,X}(u_X^*)(x) \\ &\ge \sigma(A_{x,X})\left(1 + \sum_{j=1}^M |u_j^*(x)|^2\right) \\ &\ge \sigma(A_{x,X}) \end{aligned} \tag{10.1}$$

of the power function (??), 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 10.2. *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 (??) 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 (10.1) 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

$$P_{X_h}^2(x) \leq F(h_{X,\Omega}) \text{ for all } x \in \Omega \tag{10.3}$$

where F is a monotonic function of the fill distance $h_{X,\Omega}$ defined in (7.2). On the other hand, the lower bounds for σ which we shall prove in this chapter, will be of the form

$$\sigma(A_X) \geq G(q_X) \text{ for all } X = \{x_1, \ldots, x_M\} \subset \Omega \tag{10.4}$$

with the **separation distance**

$$q := q_X := \min_{1 \le i \ne j \le M} \|x_i - x_j\|_2. \tag{10.5}$$

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

$$F(t\sqrt{d}) \ge G(t) \tag{10.6}$$

for all $t \ge 0$. This allows to check the quality of the bounds (10.3) and (10.4), since the lowest possible bounds F and the largest possible bounds G must necessarily satisfy (10.6) and are optimal, if they turn (10.6) into an equality. This opens the race for optimal bounds of the form (10.3) and (10.4), 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. (10.6) is extended to

$$F(t\sqrt{d}) \ge G(t) \ge C \cdot F(c \cdot t) \tag{10.7}$$

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.

10.2 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 (??) and (??) 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 (10.5).

We generalize the technique of Narcowich and Ward [NW91a] [NW91b] 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 9.40 allows the formula

$$\sum_{j=1}^{M}\sum_{k=1}^{M} \alpha_j\alpha_k\Phi(x_j - x_k) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{\Phi}(\omega) \left| \sum_{j=1}^{M} \alpha_j e^{ix_j\cdot\omega} \right|^2 d\omega \tag{10.8}$$

for all $\mathcal{P}_m^d$–nondegenerate sets $X = \{x_1, \ldots, x_M\}$ and all vectors $\alpha \in \mathbb{R}^M$ such that $\lambda_{X,M,\alpha}$ is a functional that annihilates $\mathcal{P}_m^d$. This is just another way of writing (9.44).

The left–hand side of (10.8) 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 $\mathbb{R}^d \setminus \{0\}$ of $\widehat{\Phi}$ that satisfies

$$\widehat{\Phi}(\omega) \geq \widehat{\Psi}(\omega) \qquad \text{on } \mathbb{R}^d \setminus \{0\} \tag{10.9}$$

and that itself leads to a similar quadratic form

$$\sum_{j=1}^{M}\sum_{k=1}^{M} \alpha_j \alpha_k \widehat{\Psi}(x_j - x_k) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{\Psi}(\omega) \left| \sum_{j=1}^{M} \alpha_j e^{ix_j \cdot \omega} \right|^2 d\omega \tag{10.10}$$

for another basis function $\widehat{\Psi}$ and a weaker constraint on $\alpha \in \mathbb{R}^M$ (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 (10.10). Then clearly for all $\alpha \in \mathbb{R}^M$ 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 10.11. *Let Φ be a conditionally positive definite function on $\mathbb{R}^d$ that satisfies Assumption 9.40. Furthermore, let $X = \{x_1, \ldots, x_M\} \subset \mathbb{R}^d$ be any set of Lagrange data locations having separation distance (10.5). With the function*

$$\phi_0(r) := \inf_{\|\omega\|_\infty \leq 2r} \widehat{\Phi}(\omega), \tag{10.12}$$

the smallest eigenvalue σ of the quadratic form associated to the matrix

$$A_{X,\Phi} = (\Phi(x_j - x_k))_{1 \leq j,k \leq M},$$

restricted as usual to the subspace of $\mathbb{R}^M$ that contains the coefficient vectors α of functionals $\lambda_{X,M,\alpha} \in \mathcal{P}_\Omega^\perp$ has the lower bound

$$\sigma \geq \frac{1}{2} \frac{\phi_0(K)}{\Gamma(d/2+1)} \left(\frac{K}{\sqrt{2}} \right)^d \tag{10.13}$$

for any $K > 0$ satisfying

$$K \geq \frac{4}{q}\left(2\pi\Gamma^2\left(d/2+1\right)\right)^{\frac{1}{d+1}} \tag{10.14}$$

or, a fortiori,

$$K \geq \frac{9.005\ d}{q}. \tag{10.15}$$

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 $\mathbb{R}^d$ 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

$$\mathrm{supp}\left(\widehat{\Psi}_K\right) = \left\{x \in \mathbb{R}^d : \|x\|_2 \leq 2K\right\} =: B_{2K}(0).$$

We now use the formula (12.27) 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 (10.9) on all of $\mathbb{R}^d$.

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 (9.16). 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 (12.40) 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 (10.10) 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

$$\max_{1\leq j\leq M}\sum_{\substack{k=1\\k\neq j}}\Psi_K(x_j-x_k) \leq \frac{1}{2}\Psi_K(0). \tag{10.16}$$

This is done by a tricky summation argument of Narcowich and Ward [NW91b] using (12.39) which proves (10.16) for K satisfying (10.14). 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 (12.27) 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 (12.39) as follows:

$$\begin{aligned} \Psi_K(x_j - x_k) &= \phi_0(K)\Gamma\left(\frac{d}{2}+1\right) 2^{d/2} \|x_j - x_k\|^{-d} J_{d/2}^2(K \cdot \|x_j - x_k\|) \\ &= \phi_0(K)\Gamma\left(\frac{d}{2}+1\right) 2^{d/2} K^{-1} \|x_j - x_k\|^{-d-1} \\ &\quad \cdot (K \cdot \|x_j - x_k\|_2) J_{d/2}^2(K \cdot \|x_j - x_k\|) \\ &\leq \phi_0(K)\Gamma\left(\frac{d}{2}+1\right) 2^{d/2} K^{-1} ((n-1)q)^{-d-1} \frac{2^{d+2}}{\pi} \\ &= \Psi_K(0) \left(\frac{4}{K(n-1)q}\right)^{d+1} \pi^{-1} \Gamma^2\left(\frac{d}{2}+1\right). \end{aligned}$$

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

$$\begin{aligned}
\sum_{k\neq j}\Psi_K(x_j-x_k) &= \sum_{n=2}^{\infty}\sum_{x_k\in E_n}\Psi(x_j-x_k)\\
&\leq \Psi_K(0)\left(\frac{4}{Kq}\right)^{d+1}\pi^{-1}\Gamma^2\left(\frac{d}{2}+1\right)\sum_{n=2}^{\infty}3n^{d-1}(n-1)^{-d-1}\\
&\leq \Psi_K(0)\left(\frac{4}{Kq}\right)^{d+1}\pi^{-1}\Gamma^2\left(\frac{d}{2}+1\right)6\sum_{n=2}^{\infty}(n-1)^{-2}\\
&\leq \Psi_K(0)\left(\frac{4}{Kq}\right)^{d+1}\pi^{-1}\Gamma^2\left(\frac{d}{2}+1\right)\pi^2\\
&= \Psi_K(0)\left(\frac{4}{Kq}\right)^{d+1}\pi\Gamma^2\left(\frac{d}{2}+1\right)\\
&\leq \tfrac{1}{2}\Psi_K(0)
\end{aligned}$$

if we choose K according to (10.16).

It remains to show that (10.15) implies (10.14). We use a variation of Stirling's formula in the form

$$\Gamma(1+x)\leq\sqrt{2\pi x}x^xe^{-x}e^{1/12x},\qquad x>0$$

to get

$$\begin{aligned}
2\pi\Gamma^2\left(d/2+1\right) &\leq 2\pi^2d^{d+1}(2e)^{-d}e^{1/3d},\\
\left(2\pi\Gamma^2\left(d/2+1\right)\right)^{\frac{1}{d+1}} &\leq \frac{d}{2e}\left(4e\pi^2\right)^{\frac{1}{d+1}}e^{\frac{1}{3d(d+1)}}\\
&\leq d\frac{\pi}{\sqrt{e}}\cdot e^{1/6}\leq d\cdot 2.2511
\end{aligned}$$

such that

$$K\geq\frac{9.005}{qd}$$

is satisfactory for all cases. □

We now want to look at the specific cases for applications. From (10.13) and (10.14) we see that

$$\sigma=\sigma(q)=\geq\wr\left(q^{-d}\phi_0(cd/q)\right)$$

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 3.

$\phi(r)$	Lower Bound in $\wr$ 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 3: Lower Bounds of Smallest Eigenvalue Based on Lagrange Data with Separation Distance q

10.3 Stability in Function Space

This text is from a recent preprint with Stefano deMarchi, and needs some brushing–up.

10.3.1 Lebesgue Constants

Given a positive definite kernel $\Phi : \Omega \times \Omega \to \mathbb{R}$, the recovery of functions from function values $f(x_j)$ on the set $X = \{x_1, ..., x_N\} \subset \Omega \subseteq \mathbb{R}^d$ of N different data sites can be done via interpolants of the form

$$s_{f,X} := \sum_{j=1}^{N} \alpha_j \Phi(\cdot, x_j) \ . \tag{10.16}$$

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 (10.16) takes the usual Lagrangian form

$$s_{f,X} = \sum_{j=1}^{N} f(x_j) u_j. \tag{10.16}$$

As in the (univariate) polynomial case, based on the representation (10.16) 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 (10.16). 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 [Wen05]. 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 [FI96, DMSW05].
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 $\mathbb{R}^d$ [Wen05].
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

$$0 < c(1 + \|\omega\|_2^2)^{-\tau} \leq \hat{\Phi}(\omega) \leq C(1 + \|\omega\|_2^2)^{-\tau} \tag{10.16}$$

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*.

10.3.2 Results for Limited Smoothness

Under the assumption (10.16) the space V_X will be a subspace of Sobolev space $W_2^\tau(\Omega)$. Our central result then is

Theorem 10.17. *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

$$\|u_j\|_{L_\infty(\Omega)} \le C.$$

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

$$f_{|X} \mapsto s_{f,X}, \;\; \mathbb{R}^N \to L_2(\Omega)$$

if the above norms are chosen.

Theorem 10.18. *Under the above assumptions,*

$$\|s_{f,X}\|_{2(\Omega)} \le C \left(\frac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2} h_{X,\Omega}^{d/2}\|f\|_{2,X},$$

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

$$\|u_j\|_{L_2(\Omega)} \le C \left(\frac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2} h_{X,\Omega}^{d/2}$$

and for quasi-uniform data locations they behave like

$$\|u_j\|_{L_2(\Omega)} \le C h_{X,\Omega}^{d/2}.$$

10.3.3 L_∞ Bounds

Our most important tool for the proof of Theorem 10.17 is the **sampling inequality** (cf. [WR05, Th. 2.6])

$$\|u\|_{L_\infty(\Omega)} \le C \left(h_{X,\Omega}^{\tau-d/2}\|u\|_{W_2^\tau(\Omega)} + \|u\|_{\infty,X}\right), \;\; \forall u \in W_2^\tau(\Omega), \tag{10.18}$$

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 [SW02], but we repeat the basic notation here. Let Φ satisfy (10.16). Then [SW02] has

$$\|s\|^2_{W_2^\tau(\Omega)} \le Cq_X^{-2\tau+d}\|s\|^2_{2,X} \le CNq_X^{-2\tau+d}\|s\|^2_{\infty,X} \text{ for all } s \in V_X$$

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 10.17 follow. □.

10.3.4 L_2 Bounds

For the L_2 case covered by Theorem 10.18, we take the sampling inequality

$$\|f\|_{L_2(\Omega)} \le C\left(h^\tau_{X,\Omega}\|f\|_{W_2^\tau(\Omega)} + \|f\|_{\ell_2(X)}h^{d/2}_{X,\Omega}\right), \quad \forall f \in W_2^\tau(\Omega) \tag{10.18}$$

of [Mad06, Thm. 3.5]. We can apply the sampling inequality as

$$\begin{aligned}
\|s_{f,X}\|_{L_2(\Omega)} &\le C\left(h^\tau_{X,\Omega}\|s_{f,X}\|_{W_2^\tau(\Omega)} + \|s_{f,X}\|_{\ell_2(X)}h^{d/2}_{X,\Omega}\right)\\
&\le C\left(h^\tau_{X,\Omega}\|s_{f,X}\|_{W_2^\tau(\Omega)} + \|f\|_{\ell_2(X)}h^{d/2}_{X,\Omega}\right),\\
&\le C\left(\tfrac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2}\|f\|_{\ell_2(X)}h^{d/2}_{X,\Omega},\\
\|u_j\|_{L_2(\Omega)} &\le C\left(h^\tau_{X,\Omega}\|u_j\|_{W_2^\tau(\Omega)} + \|u_j\|_{\ell_2(X)}h^{d/2}_{X,\Omega}\right)\\
&\le C\left(h^{\tau-d/2}_{X,\Omega}\|u_j\|_{W_2^\tau(\Omega)} + 1\right)h^{d/2}_{X,\Omega}\\
&\le C\left(\left(\tfrac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2} + 1\right)h^{d/2}_{X,\Omega}
\end{aligned}$$

This proves Theorem 10.18. □

11 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.

11.1 Normed Linear Spaces

For completeness, we recall some basics from normed linear spaces over a field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.

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 $\mathcal{M} = \mathbb{K}$, i.e. for linear **functionals** $\lambda : \mathcal{N} \to \mathbb{K}$. If they are *continuous*, they have an operator norm

$$\|\lambda\|_{\mathcal{N}^*} := \|\lambda\|_{\mathcal{N},\mathbb{K}} := \sup_{x\in\mathcal{N}\setminus\{0\}} \frac{|\lambda(x)|}{\|x\|_{\mathcal{N}}} \leq C.$$

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.

11.2 Pre–Hilbert Spaces

Definition 11.1. *A set $\mathcal{H}$ and a mapping $(\cdot,\cdot)_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \to \mathbb{K}$ form a* **pre-Hilbert space** *or a* **Euclidean** *space over $\mathbb{K}$, if the following holds:*

1. *$\mathcal{H}$ is a vector space over $\mathbb{K}$.*

2. *$(\cdot,\cdot)_{\mathcal{H}}$ is a Hermitian positive definite inner product, linear in the first and antilinear in the second argument.*

Then

$$\|x\|_{\mathcal{H}}^2 := (x,x)_{\mathcal{H}},\ x \in \mathcal{H} \tag{11.2}$$

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**

$$|(u,v)_{\mathcal{H}}| \leq |u|_{\mathcal{H}}|v|_{\mathcal{H}}$$

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 (11.2). To prove the Cauchy-Schwarz inequality for $\mathbb{K} = \mathbb{R}$ 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 $b^2 \leq ac$ 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". In the complex case, the argument is similar. For real t, we get

$$\varphi(t) := |u + tv|_{\mathcal{H}}^2 = |u|_{\mathcal{H}}^2 + 2t \text{ Re } ((u, v)_{\mathcal{H}}) + t^2 |v|_{\mathcal{H}}^2$$

and see that

$$| \text{ Re } ((u, v)_{\mathcal{H}})|^2 \leq |u|_{\mathcal{H}}^2 |v|_{\mathcal{H}}^2$$

holds, Taking purely imaginary t leads to the same for the imaginary part, proving the complex case, too.

Now we add some simple facts about pre-Hilbert spaces:

1. For two nonzero elements x, y of $\mathcal{H}$ over $\mathbb{R}$ 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

$$\|x + y\|_{\mathcal{H}}^2 = (x + y, x + y)_{\mathcal{H}} = \|x\|_{\mathcal{H}}^2 + (x, y)_{\mathcal{H}} + (y, x)_{\mathcal{H}} + \|y\|_{\mathcal{H}}^2.$$

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, provided that we work over $\mathbb{R}$.

11.3 Sequence Spaces

For illustration, we can look at **sequence spaces**. Let I be a finite or countably infinite set, and take the space

$$S_0 := \text{ span } \{\{\xi_i\}_{i \in I} \ : \ \xi_i \neq 0 \text{ for only finitely many } i \in I\}. \tag{11.3}$$

Then take a sequence $\{\lambda_i\}_{i\in I}$ of positive numbers, and define the inner product

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

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

$$S_{\lambda,I} := \text{ span}\,\{\{\xi_i\}_{i\in I} \ : \ \sum_{i\in I} \lambda_i |\xi_i|^2 < \infty\} \tag{11.4}$$

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

$$\begin{aligned}
|\mu(\xi)|^2 &= \left| \sum_{i\in I} \xi_i \mu_i \right|^2 \\
&= \left| \sum_{i\in I} \sqrt{\lambda_i} \xi_i \frac{1}{\sqrt{\lambda_i}} \mu_i \right|^2 \\
&\le \left(\sum_{i\in I} \lambda_i |\xi_i|^2 \right) \left(\sum_{i\in I} \frac{1}{\lambda_i} |\mu_i|^2 \right) \\
&= \|\xi\|_{\lambda,I}^2 \|\mu\|_{1/\lambda,I}^2.
\end{aligned}$$

The linear map

$$R \ : \ S_{1/\lambda,I} \to S_{\lambda,I}, \ \ \mu \mapsto \{\overline{\mu_i}/\lambda_i\}_i$$

has the properties

$$\mu(\xi) = \sum_i \mu_i \xi_i = \sum_i \lambda_i \xi_i \frac{\mu_i}{\lambda_i} = (\xi, R(\mu))_{\lambda,I}$$

and

$$(R(\rho), R(\mu))_{\lambda,I} = \sum_i \lambda_i \frac{\overline{\rho_i}\mu_i}{\lambda_i^2} = (\mu, \rho)_{1/\lambda,I}.$$

We shall generalize this later, but note that R is an antilinear isometry.

11.4 Best Approximations

Definition 11.5. *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 11.6. *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*

$$(u - u^*, v)_{\mathcal{H}} = 0 \text{ for all } v \in \mathcal{M} \tag{11.7}$$

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 us consider the case $\mathbb{K} = \mathbb{R}$ first. 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 (11.7) and write any other element $v \in \mathcal{M}$ as $v = u^* + 1 \cdot (v - u^*)$. Then (11.7) implies that the quadratic function u_{u^*-v} is minimal at $\alpha = 0$, proving $u_{u^*-v}(1) = \|u - v\|_{\mathcal{H}} \geq u_{u^*-v}(0) = \|u - u^*\|_{\mathcal{H}}$. 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 (11.7). This finishes the real case.

For the complex case, we have to discuss

$$u_v(\alpha) := \|u - u^* + \alpha v\|_{\mathcal{H}}^2 = \|u - u^*\|_{\mathcal{H}}^2 + 2 \text{ Re } (\alpha(u - u^*, v)_{\mathcal{H}}) + |\alpha|^2 \|v\|_{\mathcal{H}}^2$$

for all complex α. If u^* is a best approximation with $(u - u^*, v)_{\mathcal{H}} \neq 0$, we can take

$$\alpha = t \frac{\overline{(u - u^*, v)_{\mathcal{H}}}}{|(u - u^*, v)_{\mathcal{H}}|}$$

with some real t and do the same argument as above to prove (11.7). The other conclusions work like in the real case. □

Corollary 11.8. *The first statement of Theorem 11.6 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 11.9. *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*

$$\|u^*\|_{\mathcal{H}} = \inf_{\substack{v \in \mathcal{H} \\ \lambda_j(v) = \lambda_j(u) \\ 1 \leq j \leq M}} \|v\|_{\mathcal{H}},$$

iff the variational identity

$$(v, u^*)_{\mathcal{H}} = 0 \text{ for all } v \in \mathcal{H} \text{ with } \lambda_j(v) = 0,\ 1 \leq j \leq M.$$

holds, or iff there are scalars $\alpha_1, \ldots, \alpha_M$ such that

$$(v, u^*)_{\mathcal{H}} = \sum_{j=1}^{M} \alpha_j \lambda_j(v) \text{ for all } v \in \mathcal{H}.$$

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 11.6 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 11.10. *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 $\widetilde{B} : X/\ker A \to Z$ 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. □

11.5 Hilbert Spaces

So far, Theorem 11.6 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 11.11. *A pre-Hilbert space $\mathcal{H}$ over $\mathbb{K}$ with inner product $(\cdot,\cdot)_{\mathcal{H}}$ is a* **Hilbert space**, *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 11.12. *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* $\Pi_{\mathcal{M}}^2 = \Pi_{\mathcal{M}}$. *If* Id *is the identity mapping, then* $Id - \Pi_{\mathcal{M}}$ *is another projector, mapping* $\mathcal{H}$ *onto the* **orthogonal complement**

$$\mathcal{M}^\perp := \{\ u \in \mathcal{H}\ :\ (u,v)_{\mathcal{H}} = 0 \textit{ for all } v \in \mathcal{M}\ \}.$$

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 11.6, and we postpone the general case for a moment. The orthogonality statement follows in general from Theorem 11.6, 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 $\alpha, \beta \in \mathbb{R}$ and all $u, v \in \mathcal{H}$, using the variational identity in Theorem 11.6. To prove linearity of the projectors, we use uniqueness of the best approximation, as follows from Theorem 11.6. 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

$$E_{\mathcal{M}}(u) \le \|u - w_n\|_{\mathcal{H}} \le \|u - v_n\|_{\mathcal{H}},$$

such that the sequence $\{\|u - w_n\|_{\mathcal{H}}\}_n$ converges to $E_{\mathcal{M}}(u)$, too. We now fix indices $m \ge n$ 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 11.13. *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 11.12. 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 11.14. *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. □

11.6 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 $\lambda : \mathcal{N} \to \mathbb{R}$ 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 11.15. *(**Riesz representation theorem**)*
Any continuous linear functional λ on a Hilbert space $\mathcal{H}$ can be written as

$$\lambda = (\cdot, g_\lambda)_{\mathcal{H}} \tag{11.16}$$

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

$$\begin{aligned} 0 &= (\lambda(u)h_\lambda - \lambda(h_\lambda)u, h_\lambda)_{\mathcal{H}}, \\ \lambda(u)(h_\lambda, h_\lambda)_{\mathcal{H}} &= \lambda(h_\lambda)(u, h_\lambda)_{\mathcal{H}}, \\ \lambda(u) &= (u, \overline{\lambda(h_\lambda)}h_\lambda)_{\mathcal{H}} \end{aligned}$$

and we can define $g_\lambda := \overline{\lambda(h_\lambda)}h_\lambda$ to get (11.16).

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 $g_\lambda := \overline{\lambda(h_\lambda)}h_\lambda$, we get $\|\lambda\|_{\mathcal{H}^*} = \|g_\lambda\|_{\mathcal{H}}$. Uniqueness of g_λ satisfying (11.16) is easy to prove, because for any other $\tilde{g}_\lambda$ with (11.16) 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 11.17. *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. Another description is*

$$(f, R(\lambda))_{\mathcal{H}} = \lambda(f) \text{ for all } \lambda \in \mathcal{H}^*,\ f \in \mathcal{H}.$$

Theorem 11.18. *The Riesz map is an antilinear 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*

$$(R(\lambda), R(\mu))_{\mathcal{H}} = (\mu, \lambda)_{\mathcal{H}^*} \text{ for all } \lambda, \mu \in \mathcal{H}^*.$$

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 antilinear due to

$$\begin{aligned}(f, R(a\lambda + b\mu))_{\mathcal{H}} &= (a\lambda + b\mu)(f) \\ &= a\lambda(f) + b\mu(f) \\ &= a(f, R(\lambda))_{\mathcal{H}} + b(f, R(\mu))_{\mathcal{H}} \\ &= (f, \overline{a} \cdot R(\lambda) + \overline{b} \cdot R(\mu))_{\mathcal{H}} \text{ for all } f \in \mathcal{H},\ a, b \in \mathbb{R}\end{aligned}$$

because this implies that $R(a\lambda + b\mu) - (\overline{a} \cdot R(\lambda) + \overline{b} \cdot R(\mu))$ 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

$$\lambda(R(\mu)) = (R(\mu), R(\lambda))_{\mathcal{H}} = \overline{\mu(R(\lambda))} \text{ for all } \lambda, \mu \in \mathcal{H}^*.$$

We define a "new" bilinear form

$$(\lambda, \mu)_* := \lambda(R(\mu)) = (R(\mu), R(\lambda))_{\mathcal{H}} = \overline{\mu(R(\lambda))} \text{ for all } \lambda, \mu \in \mathcal{H}^*$$

on $\mathcal{H}^*$ which clearly is positive definite and sequilinear, 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 antilinear 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

$$\lambda_f(g) := (g, f)_{\mathcal{H}} \text{ for all } g \in \mathcal{H}.$$

This functional clearly is continuous due to

$$|\lambda_f(g)| = |(g, f)_{\mathcal{H}}| \leq \|f\|_{\mathcal{H}} \|g\|_{\mathcal{H}} \text{ for all } g \in \mathcal{H}.$$

by Cauchy–Schwarz, and we compare f now with $R(\lambda_f)$ to get

$$\begin{aligned}(g, f - R(\lambda_f))_{\mathcal{H}} &= (g, f)_{\mathcal{H}} - (g, R(\lambda_f))_{\mathcal{H}} \\ &= (g, f)_{\mathcal{H}} - \lambda_f(g) \\ &= (g, f)_{\mathcal{H}} - (g, f)_{\mathcal{H}} = 0\end{aligned}$$

for all $g \in \mathcal{H}$, proving $f = R(\lambda_f)$. □

11.7 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

$$\delta_x \in \mathcal{H}^* \ : \ \mathcal{H} \to \mathbb{R}, \ f \mapsto f(x)$$

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**

$$K \ : \ \Omega \times \Omega \to \mathbb{R}$$

which is Hermitian due to

$$\begin{array}{rcl} K(x,y) & := & R(\delta_x)(y) \\ & = & \delta_y(R(\delta_x)) \\ & = & (\delta_y, \delta_x)_{\mathcal{H}^*} \\ & = & \overline{(\delta_x, \delta_y)_{\mathcal{H}^*}} \\ & = & \overline{K(y,x)}. \end{array}$$

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.

11.8 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 11.19. *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 $n, m \geq K(\epsilon)$ we have

$$\|u_n - u_m\|_{\mathcal{H}} \leq \epsilon.$$

Now take $m \geq K(\epsilon)$ and the fixed Cauchy sequence $\overline{\{u_m\}_n} = J(u_m)$. Then

$$\|J(u_m) - \overline{\{u_n\}_n}\|_{\mathcal{J}} = \lim_{n\to\infty} \|u_m - u_n\|_{\mathcal{H}} \leq \epsilon$$

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

$$\|\overline{\{u_n^{(m)}\}_n} - J(v_m)\|_{\mathcal{J}} \leq 1/m.$$

Due to

$$\begin{aligned} \|v_n - v_m\|_{\mathcal{H}} &= \|J(v_n) - J(v_m)\|_{\mathcal{J}} \\ &\leq \|J(v_n) - \overline{\{u_n^{(n)}\}_n}\|_{\mathcal{J}} + \\ &\quad + \|\overline{\{u_n^{(n)}\}_n} - \overline{\{u_n^{(m)}\}_n}\|_{\mathcal{J}} + \|\overline{\{u_n^{(m)}\}_n} - J(v_m)\|_{\mathcal{J}} \\ &\to 0 \end{aligned}$$

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

$$B(\overline{\{u_n\}_n}) := \lim_{n\to\infty} A(u_n). \tag{11.20}$$

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 (11.20) 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

$$\begin{array}{rcl} J & : & \mathcal{H} \to \mathcal{J} \\ \tilde{J} & : & \mathcal{H} \to \tilde{\mathcal{J}} \end{array}$$

This leads to two continuous maps

$$\begin{array}{rcl} B & : & \mathcal{J} \to \tilde{\mathcal{J}}, \quad B \circ J = \tilde{J} \\ \tilde{B} & : & \tilde{\mathcal{J}} \to \mathcal{J}, \quad \tilde{B} \circ \tilde{J} = J \end{array} \tag{11.21}$$

and we conclude

$$\begin{array}{rcll} \tilde{B} \circ B & = & Id & \text{on } J(\mathcal{H}) \\ B \circ \tilde{B} & = & Id & \text{on } \tilde{J}(\mathcal{H}) \end{array}$$

and this extends continuously to the completion, thus

$$\begin{array}{rcll} \tilde{B} \circ B & = & Id & \text{on } \mathcal{J} \\ B \circ \tilde{B} & = & Id & \text{on } \tilde{\mathcal{J}}. \end{array}$$

But then we have isomorphisms between $\mathcal{J}$ and $\tilde{\mathcal{J}}$ which must be isometric due to (11.21). 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 $\tilde{\mathcal{J}}$. □

11.9 Applications

As an application, consider a pre–Hilbert space $\mathcal{H}$ of functions on some set Ω having a reproducing kernel $K : \Omega \times \Omega \to \mathbb{R}$ 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

$$\tilde{\lambda}_x \ : \ \mathcal{J} \to \mathbb{R}$$

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 11.3. 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 (11.4), 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 (11.4) 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.

12 Required Results from Real Analysis

Here, we provide some material that is often not contained in standard courses on Analysis or on Numerical Methods. But we assume readers to be familiar with multiindex notation and partial derivatives.

12.1 Multivariate Taylor Formula

Theorem 12.1. *Let x and y be two points in $\mathbb{R}^d$, and assume that a d–variate real–valued function f is m–times continuously differentiable on the line segment $[x,y]$ connecting x and y. With the Taylor polynomial*

$$T_x(f)(y) := \sum_{|\alpha|=0}^{m-1} D^\alpha f(x) \frac{(y-x)^\alpha}{\alpha!}$$

we then have

$$|f(y) - T_x(f)(y)| \leq \|y - x\|_\infty^m \sum_{|\alpha|=m} \frac{1}{\alpha!} |D^\alpha f(\xi(x, y, \alpha))|$$

with certain points $\xi(x, y, \alpha)$ on the line segment between x and y.

Proof: We consider the univariate function

$$g(t) := f(x + t(y - x)),\ 0 \leq t \leq 1$$

on $[0, 1]$ and write down its standard univariate Taylor representation as

$$\begin{aligned} f(y) &= g(1) \\ &= \sum_{j=0}^{m-1} g^{(j)}(0) \frac{1^j}{j!} + \int_0^1 g^{(m)}(s) \frac{(1-s)^{m-1}}{(m-1)!} ds. \end{aligned}$$

We now prove that the derivatives of g are

$$g^{(j)}(t) = \sum_{|\alpha|=j} \binom{j}{\alpha} D^\alpha f(x + t(y - x))(y - x)^\alpha.$$

This is clearly true for $j = 0$, and we proceed by induction via

$$\begin{aligned} g^{(j+1)}(t) &= \frac{d}{dt} \sum_{|\alpha|=j} \binom{j}{\alpha} D^\alpha f(x + t(y - x))(y - x)^\alpha \\ &= \sum_{|\alpha|=j} \binom{j}{\alpha} \left(\sum_{k=1}^d D^{\alpha+e_k} f(x + t(y - x))(y_k - x_k) \right) (y - x)^\alpha \\ &= \sum_{|\alpha|=j} \binom{j}{\alpha} \sum_{k=1}^d D^{\alpha+e_k} f(x + t(y - x))(y - x)^{\alpha+e_k}. \end{aligned}$$

All multiindices $\alpha + e_k$ now have $|\alpha + e_k| = j + 1$, but we want to rearrange them into multiindices β with $|\beta| = j + 1$. The number of possibilities such a β can be written as $\beta = \alpha + e_k$ is

$$\begin{aligned} \sum_{\substack{\beta = \alpha + e_k \\ |\alpha| = j \\ 1 \leq k \leq d}} \binom{j}{\alpha} &= \sum_{k=1, \beta_k > 0}^d \binom{j}{\beta - e_k} \\ &= \binom{j+1}{\beta} \sum_{k=1}^d \frac{\beta_k}{j+1} = \binom{j+1}{\beta} \end{aligned}$$

finishing the induction. Now Taylor's formula yields

$$\begin{aligned} & f(y) \\ = & \sum_{j=0}^{m-1} \sum_{|\alpha|=j} \binom{j}{\alpha} D^\alpha f(x)(y-x)^\alpha \frac{1^j}{j!} \\ & + \int_0^1 \frac{(1-s)^{m-1}}{(m-1)!} \sum_{|\alpha|=m} \binom{m}{\alpha} D^\alpha f(x+s(y-x))(y-x)^\alpha ds \\ = & \sum_{|\alpha|=0}^{m-1} D^\alpha f(x) \frac{(y-x)^\alpha}{\alpha!} \\ & + m \int_0^1 (1-s)^{m-1} \sum_{|\alpha|=m} D^\alpha f(x+s(y-x)) \frac{(y-x)^\alpha}{\alpha!} ds \end{aligned}$$

and the residual has the bound

$$\begin{aligned} & |f(y) - T_x(f)(y)| \\ \leq & m \int_0^1 (1-s)^{m-1} \left| \sum_{|\alpha|=m} D^\alpha f(x+s(y-x)) \frac{(y-x)^\alpha}{\alpha!} \right| ds \\ \leq & m \int_0^1 (1-s)^{m-1} \sum_{|\alpha|=m} |D^\alpha f(x+s(y-x))| \frac{\|y-x\|_\infty^m}{\alpha!} ds \\ \leq & \|y-x\|_\infty^m \sup_{0 \leq s \leq 1} \sum_{|\alpha|=m} \frac{1}{\alpha!} |D^\alpha f(x+s(y-x))| \\ = & \|y-x\|_\infty^m \sum_{|\alpha|=m} \frac{1}{\alpha!} |D^\alpha f(\xi(x,y,\alpha))| \end{aligned}$$

with certain points $\xi(x, y, \alpha)$ on the line segment between x and y. □

12.2 Lebesgue Integration

12.2.1 L_2 spaces

Lemma 12.2. *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 (12.13) in some more

detail. They are in $\mathcal{S}$ for all positive values of ϵ, and Lemma 12.14 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 (12.16) by an infinitely differentiable function with compact support, e.g.

$$\varphi(\epsilon, x) = \left\{ \begin{array}{ll} c(\epsilon)\exp(-1/(\epsilon^2 - \|x\|_2^2)) & \|x\|_2 < \epsilon \\ 0 & \|x\|_2 \geq \epsilon \end{array} \right\} \tag{12.3}$$

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 12.14 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

$$\int_{\mathbb{R}^d} |M_\epsilon(f)(x)|^2 dx \leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(y)|^2\varphi(\epsilon, z)dydz = \int_{\mathbb{R}^d} |f(y)|^2 dy$$

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

$$\int_{\mathbb{R}^d} |f(x) - M_\epsilon(f)(x)|^2 dx \leq \int_{\|z\|_2<\epsilon} \varphi(\epsilon, z) \int_{\mathbb{R}^d} |f(y + z) - f(y)|^2 dy dz,$$

and we use the continuity of the shift operator as proven in Lemma 12.2 to make this as small as we want by picking a suitably small ϵ. This shows

$$\lim_{\epsilon \to 0} \|f - M_\epsilon(f)\|_{L_2(\mathbb{R}^d)} = 0$$

and proves

Lemma 12.4. *The space $\mathcal{S}$ of test functions is dense in $L_2(\mathbb{R}^d)$.* □

Lemma 12.5. *The space $C_0^\infty(\mathbb{R}^d)$ of compactly supported infinitely differentiable functions is dense in $L_2(\mathbb{R}^d)$.*

Proof: We can use a standard density lemma to go over from an $f \in {}_2(\mathbb{R}^d)$ 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. □

12.3 Fourier Transforms on $\mathbb{R}^d$

This section concerns an important tool for analysis of kernels on $\mathbb{R}^d$. There are two major possibilities to pick a space $\mathcal{S}$ of test functions on $\mathbb{R}^d$ to start with, and we take the **tempered test functions** forming **Schwartz** space $\mathcal{S}$ that are verbally defined as complex-valued functions on $\mathbb{R}^d$ whose partial derivatives exist for all orders and decay faster than any polynomial towards infinity.

Definition 12.6. *For a test function $u \in \mathcal{S}$, the* **Fourier transform** *is*

$$\widehat{u}(\omega) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} u(x) e^{-ix\cdot\omega} dx, \tag{12.7}$$

where ω varies in $\mathbb{R}^d$ 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 $u \in L_1(\mathbb{R}^d)$, 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 $\mathbb{R}^d$. 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 $\mathbb{R}^d$. We simply use substitution in

$$\begin{aligned}
\widehat{u_\gamma}(i\omega) &= (2\pi)^{-d/2} \textstyle\int_{\mathbb{R}^d} e^{-\gamma\|x\|_2^2} e^{x\cdot\omega} dx \\
&= (2\pi)^{-d/2} e^{\|\omega\|_2^2/4\gamma} \textstyle\int_{\mathbb{R}^d} e^{-\|\sqrt{\gamma}x - \omega/2\sqrt{\gamma}\|_2^2} dx \\
&= (2\pi\gamma)^{-d/2} e^{\|\omega\|_2^2/4\gamma} \textstyle\int_{\mathbb{R}^d} e^{-\|y\|_2^2} dy
\end{aligned}$$

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

$$\begin{aligned}
c^2 &= \textstyle\int_{\mathbb{R}^2} e^{-\|y\|_2^2} dy \\
&= \textstyle\int_0^{2\pi} \int_0^\infty e^{-r^2} r \, dr \, d\varphi \\
&= 2\pi \textstyle\int_0^\infty e^{-r^2} r \, dr \\
&= -\pi \textstyle\int_0^\infty (-2r) e^{-r^2} \, dr \\
&= \pi.
\end{aligned}$$

This proves the first assertion of

Theorem 12.8. *The Gaussian*

$$u_\gamma(x) = \exp(-\gamma\|x\|_2^2)$$

has Fourier transform

$$\widehat{u_\gamma}(\omega) = (2\gamma)^{-d/2} e^{-\|\omega\|_2^2/4\gamma} \tag{12.9}$$

and is unconditionally positive definite on $\mathbb{R}^d$.

To understand the second assertion, we add

Definition 12.10. *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 (12.9):

$$\begin{aligned} \exp(-\beta\|\omega\|_2^2) &= (4\pi\beta)^{-d/2} \textstyle\int_{\mathbb{R}^d} e^{-\|x\|_2^2/4\beta} e^{-ix\cdot\omega} dx \\ &= (2\pi)^{-d/2} \textstyle\int_{\mathbb{R}^d} (2\beta)^{-d/2} e^{-\|x\|_2^2/4\beta} e^{+ix\cdot\omega} dx. \end{aligned}$$

Then take any set $X = \{x_1, \ldots, x_M\} \subset \mathbb{R}^d$ of M distinct points and any vector $\alpha \in \mathbb{R}^M$ to form

$$\begin{aligned} \alpha^T A_{X,u_\gamma} \alpha &= \sum_{j,k=1}^{M} \alpha_j \alpha_k \exp(-\gamma\|x_j - x_k\|_2^2) \\ &= \sum_{j,k=1}^{M} \alpha_j \alpha_k (4\pi\gamma)^{-d/2} \int_{\mathbb{R}^d} e^{-\|x\|_2^2/4\gamma} e^{-ix\cdot(x_j - x_k)} dx \\ &= (4\pi\gamma)^{-d/2} \int_{\mathbb{R}^d} e^{-\|x\|_2^2/4\gamma} \sum_{j,k=1}^{M} \alpha_j \alpha_k e^{-ix\cdot(x_j - x_k)} dx \\ &= (4\pi\gamma)^{-d/2} \int_{\mathbb{R}^d} e^{-\|x\|_2^2/4\gamma} \left| \sum_{j=1}^{M} \alpha_j e^{-ix\cdot x_j} \right|^2 dx \ge 0. \end{aligned}$$

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 $x \in \mathbb{R}^d$ 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 $y \in \mathbb{R}^d$ 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 12.11. *The Fourier transform is bijective on $\mathcal{S}$, and its inverse is the transform*

$$\check{u}(x) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} u(\omega) e^{ix\cdot\omega} d\omega. \tag{12.12}$$

Proof: The multivariate derivative D^α of $\widehat{u}$ can be taken under the integral sign, because u is in $\mathcal{S}$. Then

$$(D^\alpha \widehat{u})(\omega) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} u(x)(-ix)^\alpha e^{-ix\cdot\omega} dx,$$

and we multiply this by ω^β and use integration by parts

$$\begin{aligned}
\omega^\beta (D^\alpha \widehat{u})(\omega) &= (2\pi)^{-d/2} \textstyle\int_{\mathbb{R}^d} u(x)(-ix)^\alpha (i)^\beta (-i\omega)^\beta e^{-ix\cdot\omega} dx \\
&= (2\pi)^{-d/2} \textstyle\int_{\mathbb{R}^d} u(x)(-ix)^\alpha (i)^\beta \frac{d^\beta}{dx^\beta} e^{-ix\cdot\omega} dx \\
&= (2\pi)^{-d/2} (-1)^{|\alpha|+|\beta|} i^{\alpha+\beta} \textstyle\int_{\mathbb{R}^d} e^{-ix\cdot\omega} \frac{d^\beta}{dx^\beta} (u(x)x^\alpha) dx
\end{aligned}$$

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

$$u(x) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{u}(\omega) e^{ix\cdot\omega} d\omega$$

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:

$$\begin{aligned}
(2\pi)^{-d/2} \textstyle\int_{\mathbb{R}^d} \widehat{u}(\omega) e^{ix\cdot\omega} d\omega &= (2\pi)^{-d} \textstyle\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} u(y) e^{i(x-y)\cdot\omega} dy d\omega \\
&= \lim_{\epsilon \searrow 0} (2\pi)^{-d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} u(y) e^{i(x-y)\cdot\omega - \epsilon\|\omega\|_2^2} dy d\omega \\
&= \lim_{\epsilon \searrow 0} (2\pi)^{-d} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} e^{i(x-y)\cdot\omega - \epsilon\|\omega\|_2^2} d\omega \right) u(y) dy \\
&= \lim_{\epsilon \searrow 0} \int_{\mathbb{R}^d} \varphi(\epsilon, x-y) u(y) dy
\end{aligned}$$

with

$$\varphi(\epsilon, z) := (2\pi)^{-d} \int_{\mathbb{R}^d} e^{iz\cdot\omega - \epsilon\|\omega\|_2^2} d\omega. \tag{12.13}$$

The proof is completed by application of the following result that is useful in many contexts: □

Lemma 12.14. *The family of functions $\varphi(\epsilon, z)$ of (12.13) approximates the point evaluation functional in the sense*

$$u(x) = \lim_{\epsilon \searrow 0} \int_{\mathbb{R}^d} \varphi(\epsilon, x-y) u(y) dy \tag{12.15}$$

for all functions u that are in $L_1(\mathbb{R}^d)$ 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

$$\varphi(\epsilon, x) = (4\pi\epsilon)^{-d/2} e^{-\|x\|_2^2/4\epsilon} \tag{12.16}$$

and

$$\int_{\mathbb{R}^d} \varphi(\epsilon, x) dx = (4\pi\epsilon)^{-d/2} \int_{\mathbb{R}^d} e^{-\|x\|_2^2/4\epsilon} dx = 1.$$

To prove (12.15), we start with some given $\delta > 0$ and first find some ball $B_\rho(x)$ of radius $\rho(\delta)$ around x such that $|u(x) - u(y)| \leq \delta/2$ holds uniformly

for all $y \in B_\rho(x)$. Then we split the integral in

$$\begin{aligned} |u(x) - \textstyle\int_{\mathbb{R}^d} \varphi(\epsilon, x-y)u(y)dy| &= |\textstyle\int_{\mathbb{R}^d} \varphi(\epsilon, x-y)(u(x)-u(y))dy| \\ &\leq \textstyle\int_{\|y-x\|_2 \leq \rho} \varphi(\epsilon, x-y)|u(x)-u(y)|dy \\ &\quad + \textstyle\int_{\|y-x\| > \rho} \varphi(\epsilon, x-y)|u(x)-u(y)|dy \\ &\leq \delta/2 + (4\pi\epsilon)^{-d/2} e^{-\rho^2/4\epsilon} 2\|u\|_1 \\ &\leq \delta \end{aligned}$$

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

$$(f,g)_{L_2(\mathbb{R}^d)} := \int_{\mathbb{R}^d} f(x)\overline{g(x)}dx \tag{12.17}$$

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**

$$(\widehat{w}, \widehat{u})_{L_2(\mathbb{R}^d)} = (w,u)_{L_2(\mathbb{R}^d)} \tag{12.18}$$

for the Fourier transform on $\mathcal{S}$, proving that the Fourier transform is isometric on $\mathcal{S}$ as a subspace of $L_2(\mathbb{R}^d)$.

12.4 Fourier Transform in $L_2(\mathbb{R}^d)$

The test functions from $\mathcal{S}$ are dense in $L_2(\mathbb{R}^d)$ (see Lemma 12.4 for details), and thus we have

Theorem 12.19. *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 (12.18) 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 (12.3), replacing the Gaussian in the representation (12.16) 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 $v \in L_2(\mathbb{R}^d)$.

12.5 Poisson Summation Formula

This comes in several forms:

$$\begin{aligned}
(2\pi)^{-d/2} \sum_{k\in\mathbb{Z}^d} \hat{u}(k) &= \sum_{j\in\mathbb{Z}^d} u(2\pi j) \\
(2\pi)^{-d/2} \sum_{k\in\mathbb{Z}^d} \hat{u}(k) e^{ik^T x} &= \sum_{j\in\mathbb{Z}^d} u(x + 2\pi j) \\
(2\pi)^{-d/2} \sum_{k\in\mathbb{Z}^d} u(k) e^{-ik^T \omega} &= \sum_{j\in\mathbb{Z}^d} \hat{u}(\omega + 2\pi j) \\
(2\pi)^{-d/2} \sum_{k\in\mathbb{Z}^d} u(hk) e^{-ihk^T \omega} &= h^{-d} \sum_{j\in\mathbb{Z}^d} \hat{u}\left(\omega + \frac{2\pi j}{h}\right)
\end{aligned}$$

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

$$f(x) = \sum_{k\in\mathbb{Z}^d} c_k e^{ik^T x}, \;\; c_k = (2\pi)^{-d} \int_{[-\pi,\pi]^d} f(x) e^{-ik^T x} dx,$$

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 $\frac{2\pi}{h}$.

12.6 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 12.20. *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 12.21

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 (12.17) 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 12.22

A very important class of functionals for our purposes consists of the space $\mathcal{P}_\Omega^\perp = L$ of functionals of the form

$$\lambda_{a,X} := \sum_{x_j \in X} a_j f(x_j) \tag{12.23}$$

for finite sets $X \subset \Omega$ and $a \in \mathbb{R}^{|X|}$ that vanish on $\mathcal{P}_m^d$. Their action on a test function v is

$$\begin{aligned} \lambda_{a,X}(v) &= \sum_{j=1}^{M} a_j v(x_j) \\ &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{v}(\omega) \sum_{j=1}^{M} a_j e^{ix_j\cdot\omega} d\omega \\ &= \widehat{\lambda}_{a,X}(\widehat{v}) \end{aligned}$$

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 $\mathcal{P}_m^d$. 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 $\mathcal{P}_m^d$. We shall use both of these facts later.

Example 12.24

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

$$\begin{aligned} \left(-i\tfrac{d}{dx}\right)^\alpha v(x) &= \left(-i\tfrac{d}{dx}\right)^\alpha (2\pi)^{-d/2} \textstyle\int_{\mathbb{R}^d} \widehat{v}(\omega) e^{+ix\cdot\omega} d\omega \\ &= (2\pi)^{-d/2} \textstyle\int_{\mathbb{R}^d} \widehat{v}(\omega) (-i \cdot i\omega)^\alpha e^{+ix\cdot\omega} d\omega \\ &= (2\pi)^{-d/2} \textstyle\int_{\mathbb{R}^d} \widehat{v}(\omega) \omega^\alpha e^{+ix\cdot\omega} d\omega, \end{aligned}$$

and the associated functional is

$$v \mapsto \left(-i\frac{d}{dx}\right)^\alpha v(x)$$

at $x = 0$.

12.7 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 12.3. This section, so far, is in raw and unsorted form, because all required formulae are just collected here.

12.7.1 Gamma Function

The **Gamma function** is defined by

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \tag{12.25}$$

and has the properties

$$\begin{aligned} \Gamma(z+1) &= z\Gamma(z), \quad z \notin -\mathbb{N} \\ \Gamma(k+1) &= k!, \quad k \in \mathbb{N} \\ \Gamma(1/2) &= \sqrt{\pi}. \end{aligned}$$

The equation

$$\int_0^1 u^{x-1}(1-u)^{y-1} du = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{12.26}$$

for any $x, y > 0$ will be useful.

12.7.2 Volumes and Surface Integrals

The volume of the d-dimensional ball

$$B_r(0) := \{\, x \in \mathbb{R}^d \;:\; \|x\|_2 \le r \,\}$$

of radius r is

$$\text{vol } B_r(0) = \frac{r^d \pi^{d/2}}{\Gamma(1 + d/2)}. \tag{12.27}$$

The surface area σ_{d-1} of the $d-1$-dimensional sphere in $\mathbb{R}^d$ for $d \ge 1$ is

$$\sigma_{d-1} = \text{ vol}\,(S^{d-1}) = 2\pi^{d/2}/\Gamma(d/2). \tag{12.28}$$

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.

12.7.3 Bessel Functions

For Bessel functions, the standard source of information is [Wat95].

We consider the function $F(r\|\omega\|_2, d)$ defined by the integral

$$F(t,d) := \int_{\|y\|_2=1} e^{-ity\cdot z} dy \tag{12.29}$$

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**

$$J_\nu(t) = \frac{(t/2)^\nu}{\Gamma(\frac{2\nu+1}{2})\Gamma(\frac{1}{2})} \int_{-1}^1 e^{its}(1-s^2)^{\frac{2\nu-1}{2}} ds \tag{12.30}$$

which is well-defined for $\text{Re}\,(\nu) > -\frac{1}{2}$. We end up with $\nu = \frac{d-2}{2}$ and get

$$\begin{aligned} F(t,d) &= \sigma_{d-2} \frac{\Gamma(\frac{d-1}{2})\Gamma(\frac{1}{2})}{(t/2)^{(d-2)/2}} J_{(d-2)/2}(t) \\ &= 2\pi^{d/2} (t/2)^{-(d-2)/2} J_{(d-2)/2}(t). \end{aligned} \tag{12.31}$$

Direct integration shows that this is also valid for $d = 2$ or $\nu = 0$, using $\sigma_0 = 2$.

12.7.4 Power Series of Bessel Functions

The Bessel function of (12.30) has the power series representation

$$J_\nu(t) = \left(\frac{t}{2}\right)^\nu \sum_{j=0}^\infty \frac{\left(-\frac{t^2}{4}\right)^j}{j!\Gamma(\nu+j+1)} \tag{12.32}$$

that is valid for all $t \in \mathbb{C} \setminus \{0\}$ and all $\nu \in \mathbb{C}$. The integral representation (12.30) is first proven to be identical to the power series representation (12.32) on its domain of definition. Since the power series is convergent everywhere, the general definition of J_ν can then be done by (12.32). 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 (12.26) 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 (12.30) to yield the power series representation.

Looking at (12.32), we can define a function H_ν by

$$\left(\frac{z}{2}\right)^{-\nu} J_\nu(z) =: H_\nu(z^2/4) = \sum_{k=0}^{\infty}\frac{(-z^2/4)^k}{k!\Gamma(k+\nu+1)} \tag{12.33}$$

for $\nu \in \mathbb{C}$. This function often occurs in the text.

In a very special situation the power series representation (12.32) implies

$$\begin{aligned}J_{-1/2}(t) &= \left(\frac{t}{2}\right)^{-1/2}\sum_{j=0}^{\infty}\frac{\left(-\frac{t^2}{4}\right)^j}{j!\Gamma(j+1/2)} \\ &= \left(\frac{t}{2}\right)^{-1/2}\sum_{j=0}^{\infty}\frac{(-1)^j t^{2j}}{2^{2j}j!((j-1)/2)((j-3)/2)\ldots(1/2)\sqrt{\pi}} \\ &= \left(\frac{t}{2}\right)^{-1/2}\sum_{j=0}^{\infty}\frac{(-1)^j t^{2j}}{(2j)!\sqrt{\pi}} \\ &= \left(\frac{t}{2}\right)^{-1/2}\frac{1}{\sqrt{\pi}}\cos(t) \\ &= \sqrt{\frac{2}{\pi}}\frac{\cos(t)}{\sqrt{t}},\end{aligned} \tag{12.34}$$

and the other Bessel functions with half-integer order are similarly obtainable as linear combinations of elementary functions.

12.7.5 Relations Between Bessel Functions

By differentiation of the H_ν function from (12.33) we get

$$\begin{aligned} -\frac{d}{dt}H_\nu(rt) &= -\tfrac{d}{dt}\sum_{k=0}^{\infty}\frac{(-rt)^k}{k!\Gamma(\nu+k+1)} \\ &= r\sum_{k=1}^{\infty}\frac{(-rt)^{k-1}}{(k-1)!\Gamma(\nu+k+1)} \\ &= r\sum_{k=0}^{\infty}\frac{(-rt)^k}{k!\Gamma(\nu+k+2)} \\ &= rH_{\nu+1}(rt). \end{aligned} \tag{12.35}$$

and

$$\begin{aligned} \frac{d}{dt}t^\nu H_\nu(rt) &= \tfrac{d}{dt}\sum_{k=0}^{\infty}\frac{(-rt)^k t^\nu}{k!\Gamma(\nu+k+1)} \\ &= r\sum_{k=0}^{\infty}\frac{(-r)^k(\nu+k)t^{\nu+k-1}}{k!\Gamma(\nu+k+1)} \\ &= \sum_{k=0}^{\infty}\frac{(-rt)^k t^{\nu-1}}{k!\Gamma(\nu-1+k+1)} \\ &= t^{\nu-1}H_{\nu-1}(rt). \end{aligned} \tag{12.36}$$

We further need a special identity for Bessel functions:

$$J_{\mu+\nu+1}(t) = \frac{t^{\nu+1}}{2^\nu\Gamma(\nu+1)}\int_0^1 J_\mu(ts)s^{\mu+1}(1-s^2)^\nu ds,\ t>0, \nu>-1, \mu>-\frac{1}{2}. \tag{12.37}$$

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

$$J_{d/2}(t) = t \int_0^1 J_{(d-2)/2}(ts) s^{d/2} ds. \tag{12.38}$$

12.7.6 Bounds on Bessel Functions

We continue with two properties of Bessel functions from [NW91b]:

$$J_{d/2}^2(z) \leq \frac{2^{d+2}}{\pi z}, \quad z > 0 \tag{12.39}$$

$$\lim_{z\to 0} z^{-d} J_{d/2}^2(z) = \frac{1}{2^d \Gamma^2(1+d/2)}. \tag{12.40}$$

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 (12.39) is much more difficult and must (for now) be left to the cited literature. Similarly, there is a weaker, but more general bound

$$|J_\nu(x)| \leq 1 \tag{12.41}$$

for all $x \in \mathbb{R}$ and $\nu \geq 0$ ([AS70], 9.1.60, p. 362). Both of the above bounds should combine into the general inequality

$$|J_\nu(|x|)| \leq \wr(|x|^{-1/2}),\ x \to \infty \tag{12.42}$$

in view of [AS70], 9.2.1, p. 364. These things will be added later.

12.7.7 Integrals Involving Bessel Functions

From [AS70] 11.4.16, p. 486 we take the moment equations

$$\int_0^\infty t^\mu J_\nu(t)dt = 2^\mu \frac{\Gamma((\nu+\mu+1)/2)}{\Gamma((\nu-\mu+1)/2)} \tag{12.43}$$

which are valid for $\text{Re}\,(\nu+\mu) > -1$, $\text{Re}\,(\mu) < 1/2$. We now use these to derive similar equations for the H_ν functions by

$$\begin{aligned}
\int_0^\infty s^\rho H_\nu(s)ds &= \int_0^\infty (z^2/4)^\rho H_\nu(z^2/4)(z/2)dz \\
&= \int_0^\infty (z^2/4)^\rho (z/2)^{-\nu} J_\nu(z)(z/2)dz \\
&= 2^{\nu-1-2\rho} \int_0^\infty z^{2\rho-\nu+1} J_\nu(z)dz \\
&= \frac{\Gamma(\rho+1)}{\Gamma(\nu-\rho)}
\end{aligned} \tag{12.44}$$

for $\rho > -1$ and $\nu > 2\rho + \frac{1}{2}$.

Another citation from [AS70] 11.4.41, p. 487 is the **Weber-Schafheitlin integral**

$$\begin{aligned}
&\int_0^\infty t^{\mu-\nu+1} J_\mu(at) J_\nu(bt)dt \\
&\qquad = \left\{ \begin{array}{ll} 0 & 0 < b < a \\ \dfrac{2^{\mu-\nu+1} a^\mu (b^2-a^2)^{\nu-\mu-1}}{b^\nu \Gamma(\nu-\mu)} & 0 < a < b \end{array} \right\}
\end{aligned} \tag{12.45}$$

for $\text{Re}\,\nu > \text{Re}\,\mu > -1$ and $a \neq b > 0$.

12.7.8 Bessel Functions of Third Kind

The Bessel function K_ν of third kind (alias Mcdonald function) is defined as

$$K_\nu(z) := \int_0^\infty e^{z\cosh t} \cosh(\nu t) \tag{12.46}$$

for $z \neq 0,\ |\arg z| < \pi/2$ and all $\nu \in \mathbb{C}$. From this it follows that

$$K_\nu = K_{-\nu} \tag{12.47}$$

holds and that K_ν is positive for real parameters $\nu,\ z$. For the special case $\operatorname{Re} \nu > -1/2$ there is an integral representation

$$K_\nu(z) = \frac{\pi^{1/2}(z/2)^\nu}{\Gamma(\nu + 1/2)} \int_1^\infty e^{-zt}(t^2 - 1)^{\nu - 1/2} dt. \tag{12.48}$$

Its asymptotics near zero is

$$K_\nu(z) = \frac{(z/2)^{-\nu}}{\Gamma(\nu)} + \mathcal{O}(1) \tag{12.49}$$

for $\nu > 0$ and real, while it behaves like

$$K_\nu(z) = \frac{\sqrt{\pi}}{\sqrt{2z}} e^{-z}(1 + \wr(z^{-1})), \tag{12.50}$$

near infinity for $|\nu| \geq 1/2$. The asymptotics of K_0 near zero are like

$$K_0(r) = \frac{1}{e} - \log(r/2) + \mathcal{O}(1) \text{ for } r \to 0.$$

Due to [AS70], 11.4.44, p.488 it is related to the J_ν Bessel functions via the identity

$$\int_0^\infty \frac{t^{\nu+1} J_\nu(at)}{(t^2 + z^2)^{\mu+1}} dt = \frac{a^\mu z^{\nu-\mu}}{2^\mu \Gamma(\mu + 1)} K_{\nu-\mu}(az) \tag{12.51}$$

for $a, z > 0,\ -1 < \nu < 2\mu + 3/2$. It satisfies the differential equations

$$\begin{aligned} K'_\nu(z) &= K_{\nu-1}(z) - \frac{\nu}{z} K_\nu(z) \\ \frac{d}{dz}(z^\nu K_\nu(z)) &= -z^\nu K_{\nu-1}(z). \end{aligned} \tag{12.52}$$

The second equation, combined with (12.47), 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**.

List of Tables

List of Figures

References

[ABW06] 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.

[ADR94] 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.

[Aro50] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

[AS70] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1970.

[Ask73] R. Askey. Radial characteristic functions. Technical Report 1262, Univ. of Wisconsin, 1973. MRC Technical Sum. Report.

[BCR84] C. Berg, J. P. R. Christensen, and P. Ressel. *Harmonic Analysis on Semigroups*, volume 100 of *Graduate Texts in Mathematics*. Springer, 1984. ISBN 3-540-90925-7.

[BK05] M. Botsch and L. Kobbelt. Real-time shape editing using radial basis functions. In *Computer Graphics Forum*, volume 24, pages 611 – 621, 2005.

[BKO+96] 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.

[BS00] T.M. Bisgaard and Z. Sasvari. *Characteristic Functions and Moment Sequences, Vol. I.* Nova Science, 2000.

[BS02] S.C. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods, Second edition.* Springer, 2002.

[Buh88] M.D. Buhmann. Convergence of univariate quasi–interpolation using multiquadrics. *IMA Journal of Numerical Analysis*, 8:365–383, 1988.

[Buh90] M.D. Buhmann. Multivariate cardinal interpolation with radial–basis functions. *Constructive Approximation*, 6:225–255, 1990.

[Buh03] M.D. Buhmann. *Radial Basis Functions, Theory and Implementations*. Cambridge University Press, 2003.

[CBC+01] 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.

[CC05a] 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.

[CC05b] 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.

[CFB97] 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.

[CST00] N. Cristianini and J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, Cambridge, 2000.

[DDFI05] 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.

[DF02] T.A. Driscoll and B. Fornberg. Interpolation in the limit of increasingly flat radial basis functions. *Comput. Math. Appl.*, 43:413–422, 2002.

[DMS10] St. De Marchi and R. Schaback. Stability of kernel-based interpolation. *Adv. in Comp. Math.*, 32:155–161, 2010.

[DMSW05] Stefano De Marchi, R. Schaback, and H. Wendland. Near-optimal data-independent point locations for radial basis function interpolation. *Adv. Comput. Math.*, 23(3):317–330, 2005.

[Fas06] G.E. Fasshauer. RBF collocation methods and pseudospectral methods. Preprint, 2006.

[Fas07] G. F. Fasshauer. *Meshfree Approximation Methods with MATLAB*, volume 6 of *Interdisciplinary Mathematical Sciences*. World Scientific Publishers, Singapore, 2007.

[FI96] M.S. Floater and A. Iske. Thinning and approximation of large sets of scattered data. In F. Fontanella, K. Jetter, and P.-J. Laurent, editors, *Advanced Topics in Multivariate Approximation*, pages 87–96. World Scientific Publishing, Singapore, 1996.

[FI98] M.S. Floater and A. Iske. Thinning algorithms for scattered data interpolation. *BIT*, 38:4:705–720, 1998.

[FLW06] B. Fornberg, E. Larsson, and G. Wright. A new class of oscillatory radial basis functions. *Comput. Math. Appl.*, 51(8):1209–1222, 2006.

[Fus08a] E.J. Fuselier. Improved stability estimates and a characterization of the native space for matrix-valued RBFs. to appear in Adv. in Comp. Math., 2008.

[Fus08b] E.J. Fuselier. Sobolev-type approximation rates for divergence-free and curl-free RBF interpolants. *Mathematics of Computation*, 77:1407–1423, 2008.

[FW04] 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.

[Gas75] George Gasper. Positive integrals of Bessel functions. *SIAM J. Math. Anal.*, 6(5):868–881, 1975.

[GvL96] G. Golub and C. van Loan. *Matrix computations*. The Johns Hopkins University Press, 1996. Third edition.

[GW07] P. Giesl and H. Wendland. Meshless collocation: Error estimates with application to dynamical systems. *SIAM Journal on Numerical Analysis*, 45:1723–1741, 2007.

[Han92] P.C. Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM Review*, 34:561–580, 1992.

[Han94] P.C. Hansen. Regularization tools: a MATLAB package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms*, 6:1–35, 1994.

[Han00] 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.

[HO93] 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.

[HS10] Y.C. Hon and R. Schaback. Solving the 3d laplace equation by meshless collocation via harmonic kernels. Preprint, 2010.

[HW03] 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.

[HW05] 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.

[JZ05] 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.

[LF03] 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.

[LF05] 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.

[Li04] J. Li. A radial basis meshless method for solving inverse boundary value problems. *Comm. Numer. Methods Engrg.*, 20(1):51–61, 2004.

[Li05] 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.

[Low05a] S. Lowitzsch. Error estimates for matrix-valued radial basis function interpolation. *J. Approx. Theory*, 137(2):238–249, 2005.

[Low05b] S. Lowitzsch. Matrix-valued radial basis functions: stability estimates and applications. *Adv. Comput. Math.*, 23(3):299–315, 2005.

[LYY05] Y.J. Lee, G.J. Yoon, and J. Yoon. Convergence property of increasingly flat radial basis function interpolation to polynomial interpolation. preprint, 2005.

[Mad06] W. R. Madych. An estimate for multivariate interpolation. II. *J. Approx. Theory*, 142(2):116–128, 2006.

[Mes62] H. Meschkowski. *Hilbertsche Räume mit Kernfunktion*. Springer, Berlin, 1962.

[Mey96] A. Meyenburg. Interpolation mit Translaten einer geraden 2π–periodischen stetigen Funktion. Master's thesis, Universität Göttingen, 1996. Staatsexamensarbeit NAM–E–390.

[MN92] W.R. Madych and S.A. Nelson. Bounds on multivariate polynomials and exponential error estimates for multiquadric interpolation. *Journal of Approximation Theory*, 70:94–114, 1992.

[MS96] V. Maz'ya and G. Schmidt. On approximate approximations using Gaussian kernels. *IMA Journal of Numerical Analysis*, 16:13–29, 1996.

[MS09] St. Müller and R. Schaback. A Newton basis for kernel spaces. *Journal of Approximation Theory*, 161:645–655, 2009. doi:10.1016/j.jat.2008.10.014.

[Nar95] F.J. Narcowich. Generalized Hermite interpolation by positive definite kernels on a Riemannian manifold. *Journal of Mathematical Analysis and Applications*, 190:165–193, 1995.

[Nas06] 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.

[NFN00] 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.

[NW91a] 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.

[NW91b] F.J. Narcowich and J.D. Ward. Norms of inverses for matrices associated with scattered data. In P.J. Laurent, A. LeMéhauté, and L.L. Schumaker, editors, *Curves and Surfaces*, pages 341–348. Academic Press, Boston, 1991.

[NW94] F.J. Narcowich and J.D. Ward. Generalized Hermite interpolation via matrix–valued conditionally positive definite functions. *Mathematics of Computation*, 63(208):661–687, 1994.

[NWW06] F.J. Narcowich, J.D. Ward, and H. Wendland. Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. *Constructive Approximation*, 24:175–186, 2006.

[NWW07] F.J. Narcowich, J.D. Ward, and G.B. Wright. Divergence-free RBFs on surfaces. *J. Fourier Anal. Appl.*, 13:643–663, 2007.

[OBS03] 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.

[Opf06] R. Opfer. Multiscale kernels. *Adv. in Comput. Math.*, 25(4):357–380, 2006.

[RTSD03] 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.

[RZ06] C. Rieger and B. Zwicknagl. Sampling inequalities for infinitely smooth functions, with applications to interpolation and machine learning. preprint Gï¿½ttingen/Leipzig, 2006.

[Sai05] 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.

[Sas94] Z. Sasvari. *Positive Definite and Definitizable Functions*. Berlin: Akademie Verlag, 1994.

[Sch95] R. Schaback. Error estimates and condition numbers for radial basis function interpolation. *Advances in Computational Mathematics*, 3:251–264, 1995.

[Sch97a] 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.

[Sch97b] R. Schaback. Reconstruction of multivariate functions from scattered data. Manuscript, available via http://www.num.math.uni-goettingen.de/schaback/research/group.html, 1997.

[Sch05] 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.

[Sch06a] R. Schaback. Limit problems for interpolation by analytic radial basis functions. Preprint, to appear in J. of Computational and Applied Mathematics, 2006.

[Sch06b] R. Schaback. Recovery of functions from weak data using unsymmetric meshless kernel-based methods. Preprint, 2006.

[Sch07] R. Schaback. Unsymmetric meshless methods for operator equations. Preprint Göttingen, 2007.

[Sch09a] R. Schaback. Solving the Laplace equation by meshless collocation using harmonic kernels. *Adv. in Comp. Math.*, 31:457–470, 2009. DOI 10.1007/s10444-008-9078-3.

[Sch09b] D. Schmid. *Scattered Data Approximation on the Rotation Group and Generalizations*. PhD thesis, Technische Universität München, 2009.

[SS02] B. Schölkopf and A.J. Smola. *Learning with Kernels*. MIT Press, Cambridge, 2002.

[STC04] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[SW96] R. Schaback and Z. Wu. Operators on radial basis functions. *J. Comp. Appl. Math.*, 73:257–270, 1996.

[SW02] R. Schaback and H. Wendland. Inverse and saturation theorems for radial basis function interpolation. *Math. Comp.*, 71(238):669–681 (electronic), 2002.

[SW06] R. Schaback and H. Wendland. Kernel techniques: from machine learning to meshless methods. *Acta Numerica*, 15:543–639, 2006.

[TWN04] 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.

[Wat33] G. N. Watson. Notes on generating functions of polynomials: (2), Hermite polynomials. *J. London Math. Soc.*, 8:194–199, 1933.

[Wat95] 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.

[Wen95] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4:389–396, 1995.

[Wen05] H. Wendland. *Scattered Data Approximation*. Cambridge University Press, 2005.

[WHW05] T. Wei, Y.C. Hon, and Y.B. Wang. Reconstruction of numerical derivatives from scattered noisy data. *Inverse Problems*, 21:657 – 672, 2005.

[WK05] 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.

[WR05] H. Wendland and C. Rieger. Approximate interpolation with applications to selecting smoothing parameters. *Numer. Math.*, 101:643–662, 2005.

[Wu92] Z. Wu. Hermite–Birkhoff interpolation of scattered data by radial basis functions. *Approximation Theory and its Applications*, 8/2:1–10, 1992.

[Wu95] Z. Wu. Characterization of positive definite radial functions. In T. Lyche M. Daehlen and L.L. Schumaker, editors, *Mathematical Methods for Curves and Surfaces*, pages 573–578. Vanderbilt University Press, Nashville, TN, 1995.

[Yoo01] J. Yoon. Spectral approximation orders of radial basis function interpolation on the Sobolev space. *SIAM J. Math. Anal.*, 33(4):946–958 (electronic), 2001.

[ZS10] B. Zwicknagl and R. Schaback. Interpolation and approximation in Taylor spaces. in preparation, 2010.

Index