

Chapter 5

Error Bounds and the Variational Approach

In order to estimate the approximation properties of the functions studied thus far we will introduce the variational approach to scattered data interpolation. This approach was used first for radial basis function interpolation by Madych and Nelson [417], and later adopted by many others (see, e.g., [393, 394], [518], [545], [628, 629], [658]). We will see that for every strictly positive definite radial function there is an associated Hilbert space in which the radial basis function interpolant provides the best approximation to a given function. This optimality of interpolants in Hilbert space is the subject of the theory of optimal recovery described in the late 1950s by Golomb and Weinberger in their paper [264]. The following discussion follows mostly the presentation in Wendland's book [634].

5.1 Reproducing Kernel Hilbert Spaces

We begin with

Definition 5.1.1 *Let \mathcal{H} be a real Hilbert space of functions $f : \Omega \rightarrow \mathbb{R}$. A function $K : \Omega \times \Omega \rightarrow \mathbb{R}$ is called reproducing kernel for \mathcal{H} if*

1. $K(\mathbf{x}, \cdot) \in \mathcal{H}$ for all $\mathbf{x} \in \Omega$,
2. $f(\mathbf{x}) = \langle f, K(\cdot, \mathbf{x}) \rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$ and all $\mathbf{x} \in \Omega$.

It is known that the reproducing kernel of a Hilbert space is unique, and that existence of a reproducing kernel is equivalent to the fact that the point evaluation functionals $\delta_{\mathbf{x}}$ are bounded linear functionals, i.e., there exists a positive constant $M = M_{\mathbf{x}}$ such that

$$|\delta_{\mathbf{x}}f| = |f(\mathbf{x})| \leq M\|f\|_{\mathcal{H}}$$

for all $f \in \mathcal{H}$. This latter fact is due to the Riesz Representation Theorem.

Other properties of reproducing kernels are

Theorem 5.1.2 *Suppose \mathcal{H} is a Hilbert space of functions $f : \Omega \rightarrow \mathbb{R}$ with reproducing kernel K and \mathcal{H}^* its dual space, i.e., the space of linear functionals on \mathcal{H} . Then we have*

1. $K(\mathbf{x}, \mathbf{y}) = \langle K(\mathbf{x}, \cdot), K(\cdot, \mathbf{y}) \rangle_{\mathcal{H}}$ for $\mathbf{x}, \mathbf{y} \in \Omega$.
2. $K(\mathbf{x}, \mathbf{y}) = K(\mathbf{y}, \mathbf{x})$ for $\mathbf{x}, \mathbf{y} \in \Omega$.
3. Convergence in Hilbert space norm implies pointwise convergence.

Proof: Since by (1) of Definition 5.1.1 $K(\mathbf{x}, \cdot) \in \mathcal{H}$ for every $\mathbf{x} \in \Omega$, the reproducing property (2) of the definition gives us

$$K(\mathbf{x}, \mathbf{y}) = \langle K(\mathbf{x}, \cdot), K(\cdot, \mathbf{y}) \rangle_{\mathcal{H}}$$

for all $\mathbf{x}, \mathbf{y} \in \Omega$. (2) follows from (1) by the symmetry of the Hilbert space inner product. For (3) we use the reproducing property of K along with the Cauchy-Schwarz inequality:

$$|f(\mathbf{x}) - f_n(\mathbf{x})| = |\langle f - f_n, K(\cdot, \mathbf{x}) \rangle_{\mathcal{H}}| \leq \|f - f_n\|_{\mathcal{H}} \|K(\cdot, \mathbf{x})\|_{\mathcal{H}}.$$

□

Moreover, the reproducing kernel K is known to be positive definite. In the following we use a slight generalization of the notion of a positive definite function to a positive definite kernel. Essentially, we replace $\Phi(\mathbf{x}_j - \mathbf{x}_k)$ in Definition 1.2.5 by $K(\mathbf{x}_j, \mathbf{x}_k)$.

Theorem 5.1.3 *Suppose \mathcal{H} is a reproducing kernel Hilbert function space with reproducing kernel $K : \Omega \times \Omega \rightarrow \mathbb{R}$. Then K is positive definite. Moreover, K is strictly positive definite if and only if the point evaluation functionals are linearly independent in \mathcal{H}^* .*

Proof: Since the kernel is real-valued we can restrict ourselves to a quadratic form with real coefficients. For distinct points $\mathbf{x}_1, \dots, \mathbf{x}_N$ and nonzero $\mathbf{c} \in \mathbb{R}^N$ we have

$$\begin{aligned} \sum_{j=1}^N \sum_{k=1}^N c_j c_k K(\mathbf{x}_j, \mathbf{x}_k) &= \sum_{j=1}^N \sum_{k=1}^N c_j c_k \langle K(\mathbf{x}_j, \cdot), K(\cdot, \mathbf{x}_k) \rangle_{\mathcal{H}} \\ &= \left\langle \sum_{j=1}^N c_j K(\mathbf{x}_j, \cdot), \sum_{k=1}^N c_k K(\cdot, \mathbf{x}_k) \right\rangle_{\mathcal{H}} \\ &= \left\| \sum_{j=1}^N c_j K(\mathbf{x}_j, \cdot) \right\|_{\mathcal{H}}^2 \geq 0. \end{aligned}$$

To establish the second claim we assume K is not strictly positive definite and show that the point evaluation functionals must be linearly dependent. If K is not strictly positive definite then there exist distinct points $\mathbf{x}_1, \dots, \mathbf{x}_N$ and nonzero coefficients such that

$$\sum_{j=1}^N \sum_{k=1}^N c_j c_k K(\mathbf{x}_j, \mathbf{x}_k) = 0.$$

The first part of the proof therefore implies

$$\sum_{j=1}^N c_j K(\mathbf{x}_j, \cdot) = \sum_{j=1}^N c_j K(\cdot, \mathbf{x}_j) = 0.$$

Taking the Hilbert space inner product with an arbitrary $f \in \mathcal{H}$ and using the reproducing property of K we get that

$$\begin{aligned}
0 &= \langle f, \sum_{j=1}^N c_j K(\cdot, \mathbf{x}_j) \rangle_{\mathcal{H}} \\
&= \sum_{j=1}^N c_j \langle f, K(\cdot, \mathbf{x}_j) \rangle_{\mathcal{H}} \\
&= \sum_{j=1}^N c_j f(\mathbf{x}_j) \\
&= \sum_{j=1}^N c_j \delta_{\mathbf{x}_j}(f).
\end{aligned}$$

This, however, shows the linear dependence of the point evaluation functionals $\delta_{\mathbf{x}}(f) = f(\mathbf{x})$. An analogous argument can be used to show the converse. \square

This theorem provides a connection between strictly positive definite functions and reproducing kernels. Our interest, however, lies in the other direction. Since we are starting with strictly positive definite functions, we need to show how to construct an associated reproducing kernel Hilbert space.

5.2 Native Spaces for Strictly Positive Definite Functions

First, we note that Definition 5.1.1 tells us that \mathcal{H} contains all functions of the form

$$f = \sum_{j=1}^N c_j K(\mathbf{x}_j, \cdot)$$

provided $\mathbf{x}_j \in \Omega$. In Theorem 5.1.2 we showed that

$$\begin{aligned}
\|f\|_{\mathcal{H}}^2 &= \langle f, f \rangle_{\mathcal{H}} = \langle \sum_{j=1}^N c_j K(\mathbf{x}_j, \cdot), \sum_{k=1}^N c_k K(\cdot, \mathbf{x}_k) \rangle_{\mathcal{H}} \\
&= \sum_{j=1}^N \sum_{k=1}^N c_j c_k \langle K(\mathbf{x}_j, \cdot), K(\cdot, \mathbf{x}_k) \rangle_{\mathcal{H}} \\
&= \sum_{j=1}^N \sum_{k=1}^N c_j c_k K(\mathbf{x}_j, \mathbf{x}_k).
\end{aligned}$$

Therefore, we *define* the space

$$H_K(\Omega) = \text{span}\{K(\cdot, \mathbf{y}) : \mathbf{y} \in \Omega\}$$

with an associated bilinear form

$$\left\langle \sum_{j=1}^N c_j K(\cdot, \mathbf{x}_j), \sum_{k=1}^N d_k K(\cdot, \mathbf{y}_k) \right\rangle_K = \sum_{j=1}^N \sum_{k=1}^N c_j d_k K(\mathbf{x}_j, \mathbf{y}_k).$$

Theorem 5.2.1 *If $K : \Omega \times \Omega \rightarrow \mathbb{R}$ is a symmetric strictly positive definite kernel, then the bilinear form $\langle \cdot, \cdot \rangle_K$ defines an inner product on $H_K(\Omega)$. Furthermore, $H_K(\Omega)$ is a pre-Hilbert space with reproducing kernel K .*

Proof: $\langle \cdot, \cdot \rangle_K$ is obviously bilinear and symmetric. We just need to show that $\langle f, f \rangle_K > 0$ for nonzero $f \in H_K(\Omega)$. Any such f can be written in the form

$$f = \sum_{j=1}^N c_j K(\cdot, \mathbf{x}_j), \quad \mathbf{x}_j \in \Omega.$$

Then

$$\langle f, f \rangle_K = \sum_{j=1}^N \sum_{k=1}^N c_j c_k K(\mathbf{x}_j, \mathbf{x}_k) > 0$$

since K is strictly positive definite. The reproducing property follows from

$$\langle f, K(\cdot, \mathbf{x}) \rangle_K = \sum_{j=1}^N c_j K(\mathbf{x}_j, \mathbf{x}) = f(\mathbf{x}).$$

□

The *native space* $\mathcal{N}_K(\Omega)$ of K is now defined to be the completion of $H_K(\Omega)$ with respect to the K -norm $\|\cdot\|_K$ so that $\|f\|_K = \|f\|_{\mathcal{N}_K(\Omega)}$ for all $f \in H_K(\Omega)$. The technical details concerned with this construction are discussed in [634].

In the special case when we are dealing with strictly positive definite (translation invariant) functions $\Phi(\mathbf{x} - \mathbf{y}) = K(\mathbf{x}, \mathbf{y})$ and when $\Omega = \mathbb{R}^s$ we get a characterization of native spaces in terms of Fourier transforms.

Theorem 5.2.2 *Suppose $\Phi \in C(\mathbb{R}^s) \cap L_1(\mathbb{R}^s)$ is a real-valued strictly positive definite function. Define*

$$\mathcal{G} = \left\{ f \in L_2(\mathbb{R}^s) \cap C(\mathbb{R}^s) : \frac{\hat{f}}{\sqrt{\hat{\Phi}}} \in L_2(\mathbb{R}^s) \right\}$$

and equip this space with the bilinear form

$$\langle f, g \rangle_{\mathcal{G}} = \frac{1}{\sqrt{(2\pi)^s}} \left\langle \frac{\hat{f}}{\sqrt{\hat{\Phi}}}, \frac{\hat{g}}{\sqrt{\hat{\Phi}}} \right\rangle_{L_2(\mathbb{R}^s)} = \frac{1}{\sqrt{(2\pi)^s}} \int_{\mathbb{R}^s} \frac{\hat{f}(\boldsymbol{\omega}) \overline{\hat{g}(\boldsymbol{\omega})}}{\hat{\Phi}(\boldsymbol{\omega})} d\boldsymbol{\omega}.$$

Then \mathcal{G} is a real Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{G}}$ and reproducing kernel $\Phi(\cdot - \cdot)$. Hence, \mathcal{G} is the native space of Φ on \mathbb{R}^s , i.e., $\mathcal{G} = \mathcal{N}_{\Phi}(\mathbb{R}^s)$ and both inner product coincide. In particular, every $f \in \mathcal{N}_{\Phi}(\mathbb{R}^s)$ can be recovered from its Fourier transform $\hat{f} \in L_1(\mathbb{R}^s) \cap L_2(\mathbb{R}^s)$.

Remarks:

1. This theorem shows that the native spaces can be viewed as a generalization of the standard Sobolev spaces. Indeed, for $m > s/2$ the Sobolev space W_2^m can be defined as

$$W_2^m(\mathbb{R}^s) = \left\{ f \in L_2(\mathbb{R}^s) \cap C(\mathbb{R}^s) : \hat{f}(\cdot)(1 + \|\cdot\|_2^2)^{m/2} \in L_2(\mathbb{R}^s) \right\}.$$

Therefore, any strictly positive definite function Φ whose Fourier transform decays only algebraically has a Sobolev space as its native space. In particular, the compactly supported Wendland functions $\Phi_{s,k} = \varphi_{s,k}(\|\cdot\|_2)$ of Chapter 4 can be shown to have native spaces $\mathcal{N}_{\Phi_{s,k}}(\mathbb{R}^s) = W_2^{s/2+k+1/2}(\mathbb{R}^s)$ (where the restriction $s \geq 3$ is required for the special case $k = 0$).

2. Native spaces for strictly conditionally positive definite functions can also be constructed. However, since this is more technical, we limit the discussion here to strictly positive definite functions, and refer the interested reader to the book by Wendland [634] or the papers [554, 555] by Schaback.
3. The native spaces of the powers and thin plate (or surface) splines of Examples 2 and 3 of Sections 3.3 and 3.4 can be shown to be the so-called Beppo-Levi spaces of order k

$$BL_k(\mathbb{R}^s) = \{f \in C(\mathbb{R}^s) : D^\alpha f \in L_2(\mathbb{R}^s) \text{ for all } |\alpha| = k, \alpha \in \mathbb{N}^s\},$$

where D^α denotes a *generalized derivative* of order α (defined in the same spirit as the generalized Fourier transform). In fact, the intersection of all Beppo-Levi spaces $BL_k(\mathbb{R}^s)$ of order $k \leq m$ yields the Sobolev space $W_2^m(\mathbb{R}^s)$. For more details see [634]. These spaces were already studied in the early papers by Duchon [168, 169, 170, 171].

4. The native spaces for Gaussians and (inverse) multiquadrics are rather small. For example, according to Theorem 5.2.2, for Gaussians the Fourier transform of $f \in \mathcal{N}_\Phi(\Omega)$ must decay faster than the Fourier transform of the Gaussian (which is itself a Gaussian). It is known that, however, even though the native space of Gaussians is small, it does contain the so-called *band-limited functions*, i.e., functions whose Fourier transform is compactly supported. These functions play an important role in *sampling theory* where Shannon's famous Sampling Theorem [575] states that any band-limited function can be completely recovered from its discrete samples provided the function is sampled at a sampling rate at least twice its bandwidth. The content of this theorem was already known to Whitaker [640] in 1915.

Theorem 5.2.3 *Suppose $f \in C(\mathbb{R}^s) \cap L_1(\mathbb{R}^s)$ such that its Fourier transform vanishes outside the cube $Q = [-\frac{1}{2}, \frac{1}{2}]^s$. Then f can be uniquely reconstructed from its values on \mathbb{Z}^s , i.e.,*

$$f(\mathbf{x}) = \sum_{\boldsymbol{\xi} \in \mathbb{Z}^s} f(\boldsymbol{\xi}) \text{sinc}(\mathbf{x} - \boldsymbol{\xi}), \quad \mathbf{x} \in \mathbb{R}^s.$$

Here the sinc function is defined for any $\mathbf{x} = (x_1, \dots, x_s) \in \mathbb{R}^s$ as $\text{sinc } \mathbf{x} = \prod_{i=1}^s \frac{\sin(\pi x_i)}{\pi x_i}$. For more details on Shannon's Sampling Theorem see, e.g., Chapter 29 in the book [132] by Cheney and Light or the paper [610] by Unser.

5.3 The Power Function and Error Estimates for Functions in $\mathcal{N}_\Phi(\Omega)$

Our goal in this section is to provide error estimates for scattered data interpolation with strictly (conditionally) positive definite functions. In their final form these estimates will need to involve some kind of measure of the data distribution. The measure that is usually used is the so-called *fill distance*

$$h = h_{\mathcal{X},\Omega} = \sup_{\mathbf{x} \in \Omega} \min_{\mathbf{x}_j \in \mathcal{X}} \|\mathbf{x} - \mathbf{x}_j\|_2$$

which indicates how well the data fill out the domain Ω . The fill distance denotes the radius of the largest possible empty ball that can be placed among the data locations. We will be interested in whether the error

$$\|f - \mathcal{P}_h f\|_\infty$$

tends to zero as $h \rightarrow 0$, and if so, how fast. Here $\{\mathcal{P}_h\}_h$ presents a sequence of interpolation (or, more generally, projection) operators that vary with the fill distance h . For example, \mathcal{P}_h could denote interpolation to data given at $(2^n + 1)^s$, $n = 1, 2, \dots$, equally spaced points in the unit cube in \mathbb{R}^s (with $h = 2^{-n}$). Of course, the definition of the fill distance allows for scattered data as well.

Since we want to use the machinery of reproducing kernel Hilbert spaces we will concentrate on error estimates for functions $f \in \mathcal{N}_\Phi$. In the next section we will also mention some more general estimates.

The term that is often used to measure the speed of convergence to zero is *approximation order*. We say that the approximation operator \mathcal{P}_h has L_p -approximation order k if

$$\|f - \mathcal{P}_h f\|_p = \mathcal{O}(h^k) \quad \text{for } h \rightarrow 0.$$

Moreover, if we can also show that $\|f - \mathcal{P}_h f\|_p \neq o(h^k)$, then \mathcal{P}_h has *exact* L_p -approximation order k . We will concentrate mostly on the case $p = \infty$, but approximation order in other norms can also be studied.

In order to keep the following discussion as transparent as possible we will restrict ourselves to strictly positive definite functions. With (considerably) more technical details the following can also be formulated for strictly conditionally positive definite functions (see [634] for details).

The key idea for the following discussion is to express the interpolant in Lagrange form, i.e., using *cardinal basis functions*. This idea is due to Schaback and Wu [658]. In the previous chapters we have established that, for any strictly positive definite function Φ , the linear system

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

with $A_{ij} = \Phi(\mathbf{x}_i - \mathbf{x}_j)$, $i, j = 1, \dots, N$, $\mathbf{c} = [c_1, \dots, c_N]^T$, and $\mathbf{y} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)]^T$ has a unique solution. In the following we will consider the more general situation where Φ is a strictly positive definite kernel, i.e., the entries of A are given by $A_{ij} = \Phi(\mathbf{x}_i, \mathbf{x}_j)$.

In order to obtain the cardinal basis functions u_j^* , $j = 1, \dots, N$, with the property $u_j^*(\mathbf{x}_i) = \delta_{ij}$ we consider the linear system

$$A\mathbf{u}^*(\mathbf{x}) = \mathbf{b}(\mathbf{x}),$$

where the matrix A is as above (and therefore invertible), $\mathbf{u}^* = [u_1^*, \dots, u_N^*]^T$, and $\mathbf{b} = [\Phi(\cdot, \mathbf{x}_1), \dots, \Phi(\cdot, \mathbf{x}_N)]^T$. Thus,

Theorem 5.3.1 *Suppose Φ is a strictly positive definite kernel on \mathbb{R}^s . Then, for any distinct points $\mathbf{x}_1, \dots, \mathbf{x}_N$, there exist functions $u_j^* \in \text{span}\{\Phi(\cdot, \mathbf{x}_j), j = 1, \dots, N\}$ such that $u_j^*(\mathbf{x}_i) = \delta_{ij}$.*

Therefore, we can write $\mathcal{P}f$ in the cardinal form

$$\mathcal{P}f(\mathbf{x}) = \sum_{j=1}^N f(\mathbf{x}_j)u_j^*(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^s.$$

Another important ingredient in our estimates is the so-called *power function*. To this end, we consider a domain $\Omega \subseteq \mathbb{R}^s$. Then for any strictly positive definite kernel $\Phi \in C(\Omega \times \Omega)$, any set of distinct points $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$, and any vector $\mathbf{u} \in \mathbb{R}^N$, we define the quadratic form

$$\begin{aligned} Q(\mathbf{u}) &= \Phi(\mathbf{x}, \mathbf{x}) - 2 \sum_{j=1}^N u_j \Phi(\mathbf{x}, \mathbf{x}_j) + \sum_{i=1}^N \sum_{j=1}^N u_i u_j \Phi(\mathbf{x}_i, \mathbf{x}_j) \\ &= \langle \Phi(\cdot, \mathbf{x}), \Phi(\cdot, \mathbf{x}) \rangle_{\mathcal{N}_\Phi(\Omega)} - 2 \sum_{j=1}^N u_j \langle \Phi(\cdot, \mathbf{x}), \Phi(\cdot, \mathbf{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &\quad + \sum_{i=1}^N \sum_{j=1}^N u_i u_j \langle \Phi(\cdot, \mathbf{x}_i), \Phi(\cdot, \mathbf{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \langle \Phi(\cdot, \mathbf{x}) - \sum_{j=1}^N u_j \Phi(\cdot, \mathbf{x}_j), \Phi(\cdot, \mathbf{x}) - \sum_{j=1}^N u_j \Phi(\cdot, \mathbf{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \left\| \Phi(\cdot, \mathbf{x}) - \sum_{j=1}^N u_j \Phi(\cdot, \mathbf{x}_j) \right\|_{\mathcal{N}_\Phi(\Omega)}^2. \end{aligned} \tag{5.1}$$

Here we have used the definition of the native space norm from the previous section.

Then

Definition 5.3.2 *Suppose $\Omega \subseteq \mathbb{R}^s$ and $\Phi \in C(\Omega \times \Omega)$ is strictly positive definite on \mathbb{R}^s . For any distinct points $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$ the power function is defined by*

$$[P_{\Phi, \mathcal{X}}(\mathbf{x})]^2 = Q(\mathbf{u}^*(\mathbf{x})),$$

where \mathbf{u}^* is the vector of cardinal functions from Theorem 5.3.1.

Remarks:

1. The name *power function* was chosen by Schaback based on its connection to the power function of a statistical decision function [622].

2. In the paper [658] by Wu and Schaback the power function was referred to as *kriging function*. This terminology comes from geostatistics (see, e.g., [474]).

Now we can give a first generic error estimate.

Theorem 5.3.3 *Let $\Omega \subseteq \mathbb{R}^s$, $\Phi \in C(\Omega \times \Omega)$ be strictly positive definite on \mathbb{R}^s , and suppose that the points $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ are distinct. Denote the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on \mathcal{X} by $\mathcal{P}f$. Then for every $\mathbf{x} \in \Omega$*

$$|f(\mathbf{x}) - \mathcal{P}f(\mathbf{x})| \leq P_{\Phi, \mathcal{X}}(\mathbf{x}) \|f\|_{\mathcal{N}_\Phi(\Omega)}.$$

Proof: We express the interpolant in its cardinal form and apply the reproducing property of Φ . This gives us

$$\begin{aligned} \mathcal{P}f(\mathbf{x}) &= \sum_{j=1}^N f(\mathbf{x}_j) u_j^*(\mathbf{x}) \\ &= \sum_{j=1}^N u_j^*(\mathbf{x}) \langle f, \Phi(\cdot, \mathbf{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \langle f, \sum_{j=1}^N u_j^*(\mathbf{x}) \Phi(\cdot, \mathbf{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)}. \end{aligned}$$

For f the reproducing property of Φ yields

$$f(\mathbf{x}) = \langle f, \Phi(\cdot, \mathbf{x}) \rangle_{\mathcal{N}_\Phi(\Omega)}.$$

Now we combine these two formulas and apply the Cauchy-Schwarz inequality

$$\begin{aligned} |f(\mathbf{x}) - \mathcal{P}f(\mathbf{x})| &= \left| \langle f, \Phi(\cdot, \mathbf{x}) - \sum_{j=1}^N u_j^*(\mathbf{x}) \Phi(\cdot, \mathbf{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \right| \\ &\leq \|f\|_{\mathcal{N}_\Phi(\Omega)} \left\| \Phi(\cdot, \mathbf{x}) - \sum_{j=1}^N u_j^*(\mathbf{x}) \Phi(\cdot, \mathbf{x}_j) \right\|_{\mathcal{N}_\Phi(\Omega)} \\ &= \|f\|_{\mathcal{N}_\Phi(\Omega)} P_{\Phi, \mathcal{X}}(\mathbf{x}), \end{aligned}$$

where we have applied (5.1) and the definition of the power function. \square

Remark: One of the main benefits of Theorem 5.3.3 is that we are now able to estimate the interpolation error by considering two independent phenomena:

- the smoothness of the data (measured in terms of the native space norm of f – which is independent of the data locations),
- and the contribution due to the use of the basic function Φ and the distribution of the data (measured in terms of the power function – independent of the actual data values).

This is analogous to the standard error estimate for polynomial interpolation cited in most numerical analysis texts.

The next step is to refine this error bound by expressing the influence of the data locations in terms of the fill distance. And then, of course, the bound needs to be specialized to various choices of basic functions Φ .

The strategy to obtaining most error bounds in numerical analysis is to take advantage of the polynomial precision of a method (at least locally), and then to apply a Taylor expansion. With this in mind we observe

Theorem 5.3.4 *Let $\Omega \subseteq \mathbb{R}^s$, and suppose $\Phi \in C(\Omega \times \Omega)$ is strictly positive definite on \mathbb{R}^s . Let $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ be a set of distinct points in Ω , and define the quadratic form Q as in (5.1). The minimum of Q is given by the vector $\mathbf{u}^*(\mathbf{x})$ from Theorem 5.3.1, i.e.,*

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

Proof: Using the linear system notation employed earlier, we note that

$$Q(\mathbf{u}) = \Phi(\mathbf{x}, \mathbf{x}) - 2\mathbf{u}^T \mathbf{b}(\mathbf{x}) + \mathbf{u}^T A\mathbf{u}.$$

The minimum of this quadratic form is given by the solution of the linear system

$$A\mathbf{u} = \mathbf{b}(\mathbf{x}).$$

This, however, yields the cardinal functions $\mathbf{u} = \mathbf{u}^*(\mathbf{x})$. □

Remark: The arguments used in the previous proof suggest two alternative representations of the power function. Using the matrix-vector notation, the power function is given as

$$P_{\Phi, \mathcal{X}}(\mathbf{x}) = \sqrt{Q(\mathbf{u}^*(\mathbf{x}))} = \sqrt{\Phi(\mathbf{x}, \mathbf{x}) - 2(\mathbf{u}^*(\mathbf{x}))^T \mathbf{b}(\mathbf{x}) + (\mathbf{u}^*(\mathbf{x}))^T A\mathbf{u}^*(\mathbf{x})}.$$

However, by the definition of the cardinal functions $A\mathbf{u}^*(\mathbf{x}) = \mathbf{b}(\mathbf{x})$, and therefore we have the two new variants

$$\begin{aligned} P_{\Phi, \mathcal{X}}(\mathbf{x}) &= \sqrt{\Phi(\mathbf{x}, \mathbf{x}) - (\mathbf{u}^*(\mathbf{x}))^T \mathbf{b}(\mathbf{x})} \\ &= \sqrt{\Phi(\mathbf{x}, \mathbf{x}) - (\mathbf{u}^*(\mathbf{x}))^T A\mathbf{u}^*(\mathbf{x})}. \end{aligned}$$

In the proof below we will use a special coefficient vector $\tilde{\mathbf{u}}$ which provides the polynomial precision desired for the proof of the refined error estimate. Its existence is guaranteed by the following theorem on local polynomial reproduction proved by Wendland in [634]. This theorem requires the notion of a domain which satisfies an interior cone condition.

Definition 5.3.5 *A region $\Omega \subseteq \mathbb{R}^s$ satisfies an interior cone condition if there exists an angle $\theta \in (0, \pi/2)$ and a radius $r > 0$ such that for every $\mathbf{x} \in \Omega$ there exists a unit vector $\boldsymbol{\xi}(\mathbf{x})$ such that the cone*

$$C = \{\mathbf{x} + \lambda \mathbf{y} : \mathbf{y} \in \mathbb{R}^s, \|\mathbf{y}\|_2 = 1, \mathbf{y}^T \boldsymbol{\xi}(\mathbf{x}) \geq \cos \theta, \lambda \in [0, r]\}$$

is contained in Ω .

Remark: A consequence of the interior cone condition is the fact that a domain that satisfies this condition contains balls of a controllable radius. In particular, this will be important when bounding the remainder of the Taylor expansions below. For more details see [634].

Existence of an approximation scheme with local polynomial precision is guaranteed by

Theorem 5.3.6 *Suppose $\Omega \subseteq \mathbb{R}^s$ is bounded and satisfies an interior cone condition, and let ℓ be a non-negative integer. Then there exist positive constants h_0 , c_1 , and c_2 such that for all $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$ with $h_{\mathcal{X},\Omega} \leq h_0$ and every $\mathbf{x} \in \Omega$ there exist numbers $\tilde{u}_1(\mathbf{x}), \dots, \tilde{u}_N(\mathbf{x})$ with*

1. $\sum_{j=1}^N \tilde{u}_j(\mathbf{x}) p(\mathbf{x}_j) = p(\mathbf{x})$ for all $p \in \Pi_\ell^s$,
2. $\sum_{j=1}^N |\tilde{u}_j(\mathbf{x})| \leq c_1$,
3. $\tilde{u}_j(\mathbf{x}) = 0$ if $\|\mathbf{x} - \mathbf{x}_j\|_2 \geq c_2 h_{\mathcal{X},\Omega}$.

Remark: Property (1) yields the polynomial precision, and property (3) shows that the scheme is local. The bound in property (2) is essential for controlling the growth of error estimates and the quantity on the left-hand side of (2) is known as the *Lebesgue constant* at \mathbf{x} .

The error estimate can now be formulated in terms of the fill distance.

Theorem 5.3.7 *Suppose $\Omega \subseteq \mathbb{R}^s$ is bounded and satisfies an interior cone condition. Suppose $\Phi \in C^{2k}(\Omega \times \Omega)$ is symmetric and strictly positive definite. Denote the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on the set \mathcal{X} by $\mathcal{P}f$. Then there exist positive constant h_0 and C (independent of \mathbf{x} , f and Φ) such that*

$$|f(\mathbf{x}) - \mathcal{P}f(\mathbf{x})| \leq C C_\Phi(\mathbf{x})^{1/2} h_{\mathcal{X},\Omega}^k \|f\|_{\mathcal{N}_\Phi(\Omega)},$$

provided $h_{\mathcal{X},\Omega} \leq h_0$. Here

$$C_\Phi(\mathbf{x}) = \max_{\mathbf{w}, \mathbf{z} \in \Omega \cap B(\mathbf{x}, c_2 h_{\mathcal{X},\Omega})} |\Phi(\mathbf{w}, \mathbf{z})|.$$

Proof: By Theorem 5.3.3 we know

$$|f(\mathbf{x}) - \mathcal{P}f(\mathbf{x})| \leq P_{\Phi, \mathcal{X}}(\mathbf{x}) \|f\|_{\mathcal{N}_\Phi(\Omega)}.$$

Therefore, we now bound the power function in terms of the fill distance. We know that the power function is defined by

$$[P_{\Phi, \mathcal{X}}(\mathbf{x})]^2 = Q(\mathbf{u}^*(\mathbf{x})).$$

Moreover, we know by Theorem 5.3.4 that the quadratic form $Q(\mathbf{u})$ is minimized by $\mathbf{u} = \mathbf{u}^*(\mathbf{x})$. Therefore, any other coefficient vector \mathbf{u} will yield an upper bound on the power function. We take $\mathbf{u} = \tilde{\mathbf{u}}(\mathbf{x})$ from Theorem 5.3.6 with polynomial precision of degree $\ell \geq 2k - 1$.

We will make repeated use of the multivariate Taylor expansion

$$\Phi(\mathbf{w}, \mathbf{z}) = \sum_{|\boldsymbol{\beta}| < 2k} \frac{D_2^\beta \Phi(\mathbf{w}, \mathbf{w})}{\beta!} (\mathbf{z} - \mathbf{w})^\beta + R(\mathbf{w}, \mathbf{z})$$

with remainder

$$R(\mathbf{w}, \mathbf{z}) = \sum_{|\boldsymbol{\beta}| = 2k} \frac{D_2^\beta \Phi(\mathbf{w}, \boldsymbol{\xi}_{\mathbf{w}, \mathbf{z}})}{\beta!} (\mathbf{z} - \mathbf{w})^\beta,$$

where $\boldsymbol{\xi}_{\mathbf{w}, \mathbf{z}}$ lies somewhere on the line segment connecting \mathbf{w} and \mathbf{z} .

Following the argumentation above we have

$$[P_{\Phi, \mathcal{X}}(\mathbf{x})]^2 \leq Q(\mathbf{u}) = \Phi(\mathbf{x}, \mathbf{x}) - 2 \sum_j u_j \Phi(\mathbf{x}, \mathbf{x}_j) + \sum_i \sum_j u_i u_j \Phi(\mathbf{x}_i, \mathbf{x}_j),$$

where the sums are over those indices j with $u_j \neq 0$. Now we apply the Taylor expansion to both $\Phi(\mathbf{x}, \mathbf{x}_j)$ and $\Phi(\mathbf{x}_i, \mathbf{x}_j)$. This yields

$$\begin{aligned} Q(\mathbf{u}) &= \Phi(\mathbf{x}, \mathbf{x}) - 2 \sum_j u_j \left[\sum_{|\boldsymbol{\beta}| < 2k} \frac{D_2^\beta \Phi(\mathbf{x}, \mathbf{x})}{\beta!} (\mathbf{x}_j - \mathbf{x})^\beta + R(\mathbf{x}, \mathbf{x}_j) \right] \\ &\quad + \sum_i \sum_j u_i u_j \left[\sum_{|\boldsymbol{\beta}| < 2k} \frac{D_2^\beta \Phi(\mathbf{x}_i, \mathbf{x}_i)}{\beta!} (\mathbf{x}_j - \mathbf{x}_i)^\beta + R(\mathbf{x}_i, \mathbf{x}_j) \right]. \end{aligned}$$

Next, the polynomial precision property of the coefficient vector \mathbf{u} simplifies this expression to

$$\begin{aligned} Q(\mathbf{u}) &= \Phi(\mathbf{x}, \mathbf{x}) - 2\Phi(\mathbf{x}, \mathbf{x}) - 2 \sum_j u_j R(\mathbf{x}, \mathbf{x}_j) \\ &\quad + \sum_i u_i \underbrace{\sum_{|\boldsymbol{\beta}| < 2k} \frac{D_2^\beta \Phi(\mathbf{x}_i, \mathbf{x}_i)}{\beta!} (\mathbf{x} - \mathbf{x}_i)^\beta}_{=\Phi(\mathbf{x}_i, \mathbf{x}) - R(\mathbf{x}_i, \mathbf{x})} + \sum_i \sum_j u_i u_j R(\mathbf{x}_i, \mathbf{x}_j). \end{aligned}$$

Rearranging the terms and another application of the Taylor expansion results in

$$\begin{aligned} Q(\mathbf{u}) &= -\Phi(\mathbf{x}, \mathbf{x}) - \sum_j u_j \left[2R(\mathbf{x}, \mathbf{x}_j) - \sum_i u_i R(\mathbf{x}_i, \mathbf{x}_j) \right] \\ &\quad + \sum_i u_i [\Phi(\mathbf{x}_i, \mathbf{x}) - R(\mathbf{x}_i, \mathbf{x})]. \end{aligned}$$

Theorem 5.3.6 allows us to bound $\sum_j |u_j| \leq c_1$. Moreover, since $\|\mathbf{x} - \mathbf{x}_j\|_2 \leq c_2 h_{\mathcal{X}, \Omega}$ and $\|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq 2c_2 h_{\mathcal{X}, \Omega}$ the remainder terms can be bounded as stated. \square

Remarks:

1. For infinitely smooth strictly positive definite functions such as the Gaussians and the inverse multiquadrics we see that the approximation order k is arbitrarily high.
2. For strictly positive definite functions with limited smoothness such as the Wendland functions $\varphi_{s,k}$ the approximation order is limited by the smoothness of the basic function.
3. The estimate in Theorem 5.3.7 is still generic, since it does not account for the particular basic function Φ being used for the interpolation.
4. We point out that the factor C_Φ may still depend on $h_{\mathcal{X},\Omega}$. For most basic functions it will be possible to use C_Φ to “squeeze out” additional powers of h . This is the reason for splitting the constant in front of the h -power into a generic C and a C_Φ .

The statement of Theorem 5.3.7 can be generalized for strictly conditionally positive definite functions and also to cover the error for approximating the derivatives of f by derivatives of $\mathcal{P}f$. We state this general theorem without comment.

Theorem 5.3.8 *Suppose $\Omega \subseteq \mathbb{R}^s$ is open and bounded and satisfies an interior cone condition. Suppose $\Phi \in C^{2k}(\Omega \times \Omega)$ is symmetric and strictly conditionally positive definite of order m on \mathbb{R}^s . Denote the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on the $(m-1)$ -unisolvent set \mathcal{X} by $\mathcal{P}f$. Fix $\alpha \in \mathbb{N}_0^s$ with $|\alpha| \leq k$. Then there exist positive constant h_0 and C (independent of \mathbf{x} , f and Φ) such that*

$$|D^\alpha f(\mathbf{x}) - D^\alpha \mathcal{P}f(\mathbf{x})| \leq CC_\Phi(\mathbf{x})^{1/2} h_{\mathcal{X},\Omega}^{k-|\alpha|} |f|_{\mathcal{N}_\Phi(\Omega)},$$

provided $h_{\mathcal{X},\Omega} \leq h_0$. Here

$$C_\Phi(\mathbf{x}) = \max_{\substack{\beta, \gamma \in \mathbb{N}_0^s \\ |\beta| + |\gamma| = 2k}} \max_{\mathbf{w}, \mathbf{z} \in \Omega \cap B(\mathbf{x}, c_2 h_{\mathcal{X},\Omega})} |D_1^\beta D_2^\gamma \Phi(\mathbf{w}, \mathbf{z})|.$$

5.4 More on Error Estimates

The additional refinement of the error estimate of Theorem 5.3.8 for specific functions Φ is rather technical (for details see, e.g., the book by Wendland [634]). We only list the final bounds for various functions Φ .

Application of Theorem 5.3.8 to infinitely smooth functions such as Gaussians or (inverse) multiquadrics immediately yields arbitrarily high algebraic convergence rates, i.e., for every $\ell \in \mathbb{N}$ and $|\alpha| \leq \ell$ we have

$$|D^\alpha f(\mathbf{x}) - D^\alpha \mathcal{P}f(\mathbf{x})| \leq C_\ell h^{\ell-|\alpha|} |f|_{\mathcal{N}_\Phi(\Omega)}.$$

whenever $f \in \mathcal{N}_\Phi(\Omega)$. Considerable amount of work has gone into investigating the dependence of the constant C_ℓ on ℓ . Using different proof techniques it is possible to

show that for Gaussians $\Phi(\mathbf{x}) = e^{-\alpha\|\mathbf{x}\|^2}$, $\alpha > 0$, we get for some positive constant c that

$$\|f - \mathcal{P}f\|_{L_\infty(\Omega)} \leq e^{\frac{-c|\log h_{\mathcal{X},\Omega}|}{h_{\mathcal{X},\Omega}}} \|f\|_{\mathcal{N}_\Phi(\Omega)} \quad (5.2)$$

provided $h_{\mathcal{X},\Omega}$ is sufficiently small and $f \in \mathcal{N}_\Phi(\Omega)$. The corresponding result for (inverse) multiquadrics $\Phi(\mathbf{x}) = (\|\mathbf{x}\|^2 + \alpha^2)^\beta$, $\alpha > 0$, $\beta < 0$, or $\beta > 0$ and $\beta \notin \mathbb{N}$, is

$$\|f - \mathcal{P}f\|_{L_\infty(\Omega)} \leq e^{\frac{-c}{h_{\mathcal{X},\Omega}}} |f|_{\mathcal{N}_\Phi(\Omega)} \quad (5.3)$$

For functions with finite smoothness (such as powers, thin plate splines, and Wendland's compactly supported functions) it is possible to bound the constant $C_\Phi(\mathbf{x})$ and thereby to improve the order predicted by Theorem 5.3.8 by some additional powers of h . This results in the following error estimates.

For the powers $\Phi(\mathbf{x}) = (-1)^{\lceil\beta/2\rceil} \|\mathbf{x}\|^\beta$, $\beta > 0$, $\beta \notin 2\mathbb{N}$, we get

$$|D^\alpha f(\mathbf{x}) - D^\alpha \mathcal{P}f(\mathbf{x})| \leq Ch^{\frac{\beta}{2} - |\alpha|} |f|_{\mathcal{N}_\Phi(\Omega)}. \quad (5.4)$$

provided $|\alpha| \leq \frac{\lceil\beta\rceil - 1}{2}$ and $f \in \mathcal{N}_\Phi(\Omega)$.

For thin plate splines $\Phi(\mathbf{x}) = (-1)^{k+1} \|\mathbf{x}\|^{2k} \log \|\mathbf{x}\|$, we get

$$|D^\alpha f(\mathbf{x}) - D^\alpha \mathcal{P}f(\mathbf{x})| \leq Ch^{k - |\alpha|} |f|_{\mathcal{N}_\Phi(\Omega)}. \quad (5.5)$$

provided $|\alpha| \leq k - 1$ and $f \in \mathcal{N}_\Phi(\Omega)$.

For Wendland's compactly supported functions $\Phi_{s,k}(\mathbf{x}) = \varphi_{s,k}(\|\mathbf{x}\|)$ this first refinement leads to

$$|D^\alpha f(\mathbf{x}) - D^\alpha \mathcal{P}f(\mathbf{x})| \leq Ch^{k + \frac{1}{2} - |\alpha|} \|f\|_{\mathcal{N}_\Phi(\Omega)}. \quad (5.6)$$

provided $|\alpha| \leq k$ and $f \in \mathcal{N}_\Phi(\Omega)$.

Remark: The convergence result for the compactly supported functions assumes that the support radius is kept fixed, and that only the domain Ω is filled out by adding more points to \mathcal{X} , and thus decreasing the fill distance $h_{\mathcal{X},\Omega}$. However, this means that for small fill distances (with fixed support radius) the system matrices of the interpolation problem become more and more dense – and thus the advantage of the compact support is lost. This point of view is referred to in the literature as the *non-stationary approach*. We are guaranteed convergence, but at the cost of increased computational complexity. Another possibility is presented by the *stationary approach*, for which we scale the support radius proportional to the fill distance. In this case the sparsity of the interpolation matrix remains fixed, however, convergence is lost. We will revisit this phenomenon later.

The powers and thin plate splines can be interpreted as a generalization of univariate natural splines. Therefore, one can see that the approximation order estimates obtained via the native space approach are not optimal. For example, for interpolation with thin plate splines $\Phi(\mathbf{x}) = \|\mathbf{x}\|^2 \log \|\mathbf{x}\|$ one would expect order $\mathcal{O}(h^2)$, but the above estimate yields only $\mathcal{O}(h)$.

One can improve the estimates for functions with finite smoothness (i.e., powers, thin plate splines, Wendland's functions) by either (or both) of the following two ideas:

- by requiring the data function f to be even smoother than what the native space prescribes, i.e., by building certain boundary conditions into the native space;
- by using weaker norms to measure the error.

The idea to localize the data by adding boundary conditions was introduced in a paper by Light and Wayne [394]. This “trick” allows us to double the approximation order. The second idea can already be found in the early work by Duchon [170]. After applying both improvements the final approximation order estimate for interpolation with the compactly supported functions $\Phi_{s,k}$ is (see [629])

$$\|f - \mathcal{P}f\|_{L_2(\Omega)} \leq Ch^{2k+1+s} \|f\|_{W_2^{2k+1+s}(\mathbf{R}^s)}, \quad (5.7)$$

where f is assumed to lie in the subspace $W_2^{2k+1+s}(\mathbf{R}^s)$ of $\mathcal{N}_\Phi(\mathbf{R}^s)$. For powers and thin plate splines one obtains L_2 -error estimates of order $\mathcal{O}(h^{\beta+s})$ and $\mathcal{O}(h^{2k+s})$, respectively. These estimates are optimal, i.e., exact approximation orders, as shown by Bejancu [48].

Work on improved error bounds is also due to others such as Bejancu, Johnson, Powell, Ron, Schaback, and Yoon. In particular, recent work by Yoon provides L_p error estimates for so-called *shifted surface splines* for functions f in standard Sobolev spaces. These functions include all of the (inverse) multiquadrics, powers and thin plate splines. They are of the form

$$\Phi(\mathbf{x}) = \begin{cases} (-1)^{\lceil \beta-s/2 \rceil} (\|\mathbf{x}\|^2 + \alpha^2)^{\beta-s/2}, & s \text{ odd,} \\ (-1)^{\beta-s/2+1} (\|\mathbf{x}\|^2 + \alpha^2)^{\beta-s/2} \log(\|\mathbf{x}\|^2 + \alpha^2)^{1/2}, & s \text{ even,} \end{cases}$$

where $\beta \in \mathbf{N}$, $\beta > s/2$.

Yoon [668] has the following theorem that is formulated in the stationary setting.

Theorem 5.4.1 *Let Φ be a shifted surface spline with parameter α proportional to the fill distance $h_{\mathcal{X},\Omega}$. Then there exists a positive constant C (independent of \mathcal{X}) such that for every $f \in W_2^m(\Omega) \cap W_\infty^m(\Omega)$ we have*

$$\|f - \mathcal{P}f\|_{L_p(\Omega)} \leq Ch^{\gamma_p} |f|_{W_2^m(\mathbf{R}^s)}, \quad 1 \leq p \leq \infty,$$

with

$$\gamma_p = \min\{m, m - s/2 + s/p\}.$$

Furthermore, if $f \in W_2^k(\Omega) \cap W_\infty^k(\Omega)$ with $\max\{0, s/2 - s/p\} < k < m$, then

$$\|f - \mathcal{P}f\|_{L_p(\Omega)} = o(h^{\gamma_p - m + k}).$$

Remarks:

1. Using the localization idea mentioned above Yoon’s estimates can be “doubled” to $\mathcal{O}(h^{m+\gamma_p})$.

2. Yoon’s estimates address the question of how well the infinitely smooth (inverse) multiquadrics approximate functions that are less smooth than those in their native space. For example, Theorem 5.4.1 states that approximation to functions in $W_2^2(\Omega)$, $\Omega \subseteq \mathbb{R}^3$, by multiquadrics $\Phi(\mathbf{x}) = \sqrt{\|\mathbf{x}\|^2 + \alpha^2}$ is of the order $\mathcal{O}(h^2)$. However, it needs to be emphasized that this refers to stationary approximation, i.e., α is scaled proportional to the fill distance, whereas the spectral order given in (5.3) corresponds to the non-stationary case with fixed α . Similar numerical evidence was also provided much earlier by Schaback [545].
3. Moreover, the second part of Yoon’s result is a step toward *exact* approximation orders.
4. In order to obtain the estimates for the infinitely smooth functions, Yoon localizes the data function f by preconditioning it via convolution with a Hörmander smoothing kernel.

5.5 The Connection to Optimal Recovery

In the paper [264] by Michael Golomb and Hans Weinberger the following general problem is studied: Given the values $f_1 = \lambda_1(f), \dots, f_N = \lambda_N(f) \in \mathbb{R}$, where $\{\lambda_1, \dots, \lambda_N\}$ is a linearly independent set of linear functionals (called *information functionals* yielding the *information* about f), how does one “best” approximate the value $\lambda(f)$ where λ is a given linear functional and f is unknown? The value $\lambda(f)$ is also referred to as a *feature* of f . Moreover, what is the total range of values for $\lambda(f)$?

Remarks:

1. This is a very general problem formulation that allows not only for interpolation of function values, but also for other types of data (such as values of derivatives, integrals of f , moments of f , etc.), as well as other types of approximation.
2. Optimal recovery was also studied in detail by Micchelli, Rivlin and Winograd [457, 458, 459, 460].

In a Hilbert space setting the solution to this “optimal recovery problem” is shown to be the *minimum-norm interpolant*. More precisely, given $f_1 = \lambda_1(f), \dots, f_N = \lambda_N(f) \in \mathbb{R}$ with $\{\lambda_1, \dots, \lambda_N\} \subseteq \mathcal{H}^*$, the minimum-norm interpolant is that function $s^* \in \mathcal{H}$ that satisfies

$$\lambda_j(s^*) = f_j, \quad j = 1, \dots, N,$$

and for which

$$\|s^*\|_{\mathcal{H}} = \min_{\substack{s \in \mathcal{H} \\ \lambda_j(s) = f_j, j=1, \dots, N}} \|s\|_{\mathcal{H}}.$$

It turns out that the radial basis function interpolant satisfies these criteria if \mathcal{H} is taken as the associated native space $\mathcal{N}_{\Phi}(\Omega)$. The proofs of the two “optimality theorems” below require the following two lemmas.

Lemma 5.5.1 Assume Φ is a symmetric strictly positive definite kernel on \mathbb{R}^s and let $\mathcal{P}f$ be the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$. Then

$$\langle \mathcal{P}f, \mathcal{P}f - s \rangle_{\mathcal{N}_\Phi(\Omega)} = 0$$

for all interpolants $s \in \mathcal{N}_\Phi(\mathcal{X})$, i.e., with $s(\mathbf{x}_j) = f(\mathbf{x}_j)$, $j = 1, \dots, N$.

Proof: The interpolant $\mathcal{P}f$ is of the form

$$\mathcal{P}f = \sum_{j=1}^N c_j \Phi(\cdot, \mathbf{x}_j).$$

Using this representation, the symmetry of the kernel Φ and its reproducing property we have

$$\begin{aligned} \langle \mathcal{P}f, \mathcal{P}f - s \rangle_{\mathcal{N}_\Phi(\Omega)} &= \left\langle \sum_{j=1}^N c_j \Phi(\cdot, \mathbf{x}_j), \mathcal{P}f - s \right\rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \sum_{j=1}^N c_j \langle \Phi(\cdot, \mathbf{x}_j), \mathcal{P}f - s \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \sum_{j=1}^N c_j \langle \mathcal{P}f - s, \Phi(\cdot, \mathbf{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \sum_{j=1}^N c_j (\mathcal{P}f - s)(\mathbf{x}_j) \\ &= 0 \end{aligned}$$

since both $\mathcal{P}f$ and s interpolate f on \mathcal{X} . □

For the next result, remember the definition of the space $H_\Phi(\mathcal{X})$ as the linear span

$$H_\Phi(\mathcal{X}) = \left\{ s = \sum_{j=1}^N c_j \Phi(\cdot, \mathbf{x}_j), \mathbf{x}_j \in \mathcal{X} \right\}$$

given at the beginning of this chapter.

Lemma 5.5.2 Assume Φ is a strictly positive definite kernel on \mathbb{R}^s and let $\mathcal{P}f$ be the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$. Then

$$\langle f - \mathcal{P}f, s \rangle_{\mathcal{N}_\Phi(\Omega)} = 0$$

for all $s \in H_\Phi(\mathcal{X})$.

Proof: Any $s \in H_\Phi(\mathcal{X})$ can be written in the form

$$s = \sum_{j=1}^N c_j \Phi(\cdot, \mathbf{x}_j).$$

Using this representation of s as well as the reproducing property of Φ we have

$$\begin{aligned}
\langle f - \mathcal{P}f, s \rangle_{\mathcal{N}_\Phi(\Omega)} &= \langle f - \mathcal{P}f, \sum_{j=1}^N c_j \Phi(\cdot, \mathbf{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&= \sum_{j=1}^N c_j \langle f - \mathcal{P}f, \Phi(\cdot, \mathbf{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&= \sum_{j=1}^N c_j (f - \mathcal{P}f)(\mathbf{x}_j).
\end{aligned}$$

This last expression, however, is zero since $\mathcal{P}f$ interpolates f on \mathcal{X} , i.e., $(f - \mathcal{P}f)(\mathbf{x}_j) = 0$, $j = 1, \dots, N$. \square

The following ‘‘energy splitting’’ theorem is an immediate consequence of Lemma 5.5.2. It says that the native space energy of f can be split into the energy of the interpolant $\mathcal{P}f$ and that of the residual $f - \mathcal{P}f$.

Corollary 5.5.3 *The orthogonality property of Lemma 5.5.2 implies the energy split*

$$\|f\|_{\mathcal{N}_\Phi(\Omega)}^2 = \|f - \mathcal{P}f\|_{\mathcal{N}_\Phi(\Omega)}^2 + \|\mathcal{P}f\|_{\mathcal{N}_\Phi(\Omega)}^2.$$

Proof: The statement follows from

$$\begin{aligned}
\|f\|_{\mathcal{N}_\Phi(\Omega)}^2 &= \|f - \mathcal{P}f + \mathcal{P}f\|_{\mathcal{N}_\Phi(\Omega)}^2 \\
&= \langle (f - \mathcal{P}f) + \mathcal{P}f, (f - \mathcal{P}f) + \mathcal{P}f \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&= \|f - \mathcal{P}f\|_{\mathcal{N}_\Phi(\Omega)}^2 + 2\langle f - \mathcal{P}f, \mathcal{P}f \rangle_{\mathcal{N}_\Phi(\Omega)} + \|\mathcal{P}f\|_{\mathcal{N}_\Phi(\Omega)}^2
\end{aligned}$$

and the fact that $\langle f - \mathcal{P}f, \mathcal{P}f \rangle_{\mathcal{N}_\Phi(\Omega)} = 0$ by Lemma 5.5.2. \square

Remark: The above energy split is the fundamental idea behind a number of Krylov-type iterative algorithms for approximately solving the interpolation problem when very large data sets are involved (see, e.g., the papers [212] and [213] by Faul and Powell or [562] by Schaback and Wendland).

The following theorem shows the first optimality property of strictly conditionally positive definite kernels. It is taken from [634].

Theorem 5.5.4 *Suppose $\Phi \in C(\Omega \times \Omega)$ is a strictly conditionally positive definite kernel with respect to the finite-dimensional space $P \subseteq C(\Omega)$ and that \mathcal{X} is P -unisolvent. If the values f_1, \dots, f_N are given, then the interpolant $\mathcal{P}f$ is the minimum-norm interpolant to $\{f_j\}_{j=1}^N$, i.e.,*

$$|\mathcal{P}f|_{\mathcal{N}_\Phi(\Omega)} = \min_{\substack{s \in \mathcal{N}_\Phi(\Omega) \\ s(\mathbf{x}_j) = f_j, j=1, \dots, N}} |s|_{\mathcal{N}_\Phi(\Omega)}.$$

Proof: We consider only the strictly positive definite case. Consider an arbitrary interpolant $s \in \mathcal{N}_\Phi(\Omega)$ to f_1, \dots, f_N . Since $\mathcal{P}f \in \mathcal{N}_\Phi(\Omega)$ we can apply Lemma 5.5.1 and get

$$\langle \mathcal{P}f, \mathcal{P}f - s \rangle_{\mathcal{N}_\Phi(\Omega)} = 0.$$

Now

$$\begin{aligned}
|\mathcal{P}f|_{\mathcal{N}_\Phi(\Omega)}^2 &= \langle \mathcal{P}f, \mathcal{P}f - s + s \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&= \langle \mathcal{P}f, \mathcal{P}f - s \rangle_{\mathcal{N}_\Phi(\Omega)} + \langle \mathcal{P}f, s \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&= \langle \mathcal{P}f, s \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&\leq |\mathcal{P}f|_{\mathcal{N}_\Phi(\Omega)} |s|_{\mathcal{N}_\Phi(\Omega)}
\end{aligned}$$

so that the statement follows. \square

Remarks:

1. The space P mentioned in Theorem 5.5.4 is usually taken as a space of multivariate polynomials.
2. For thin plate splines $\phi(r) = r^2 \log r$, $r = \|\mathbf{x}\|_2$ with $\mathbf{x} = (x, y) \in \mathbb{R}^2$, the corresponding semi-norm in the Beppo-Levi space $BL_2(\mathbb{R}^2)$ is

$$|f|_{BL_2(\mathbb{R}^2)}^2 = \int_{\mathbb{R}^2} \left| \frac{\partial^2 f}{\partial x^2}(\mathbf{x}) \right|^2 + 2 \left| \frac{\partial^2 f}{\partial x \partial y}(\mathbf{x}) \right|^2 + \left| \frac{\partial^2 f}{\partial y^2}(\mathbf{x}) \right|^2 d\mathbf{x},$$

which is the bending energy of a thin plate, and thus explains the name of these functions.

Another nice property of the radial basis function interpolant is that it is at the same time the best Hilbert-space approximation to the given data, and thus not just any projection of f but the *orthogonal projection*. More precisely,

Theorem 5.5.5 *Let*

$$H_\Phi(\mathcal{X}) = \left\{ s = \sum_{j=1}^N c_j \Phi(\cdot, \mathbf{x}_j) + p \mid p \in P \text{ and } \sum_{j=1}^N c_j q(\mathbf{x}_j) = 0 \text{ for all } q \in P \text{ and } \mathbf{x}_j \in \mathcal{X} \right\},$$

where $\Phi \in C(\Omega \times \Omega)$ is a strictly conditionally positive definite kernel with respect to the finite-dimensional space $P \subseteq C(\Omega)$ and \mathcal{X} is P -unisolvent. If only the values $f_1 = f(\mathbf{x}_1), \dots, f_N = f(\mathbf{x}_N)$ are given, then the interpolant $\mathcal{P}f$ is the best approximation to f from $H_\Phi(\mathcal{X})$ in $\mathcal{N}_\Phi(\Omega)$, i.e.,

$$|f - \mathcal{P}f|_{\mathcal{N}_\Phi(\Omega)} \leq |f - s|_{\mathcal{N}_\Phi(\Omega)}$$

for all $s \in H_\Phi(\mathcal{X})$.

Proof: We consider only the strictly positive definite case. As explained in Section 5.2, the native space $\mathcal{N}_\Phi(\Omega)$ is the completion of $H_\Phi(\Omega)$ with respect to the $\|\cdot\|_\Phi$ -norm so that $\|f\|_\Phi = \|f\|_{\mathcal{N}_\Phi(\Omega)}$ for all $f \in H_\Phi(\Omega)$. Also, $\mathcal{X} \subseteq \Omega$. Therefore, we can express best approximation from $H_\Phi(\mathcal{X})$ by

$$\langle f - s^*, s \rangle_{\mathcal{N}_\Phi(\Omega)} = 0 \quad \text{for all } s \in H_\Phi(\mathcal{X}).$$

However, Lemma 5.5.2 shows that $s^* = \mathcal{P}f$ satisfies this relation. \square

Remarks:

1. The connection between radial basis function interpolation and the optimal recovery theory by Golomb and Weinberger was pointed out by various people (e.g., Schaback [545, 551], or Light and Wayne [394]).
2. These optimality properties of radial basis function interpolants play an important role in applications such as in the design of support vector machines in artificial intelligence or the numerical solutions of partial differential equations.
3. The optimality results above imply that one could also start with some Hilbert space \mathcal{H} with norm $\|\cdot\|_{\mathcal{H}}$ and ask to find the minimum norm interpolant (i.e., Hilbert space best approximation) to some given data. In this way any given space defines a set of *optimal basis functions*, generated by the reproducing kernel of \mathcal{H} . This is how Duchon approached the subject in his papers [168, 169, 170, 171]. More recently, Kybic, Blu and Unser [356, 357] take this point of view and explain very nicely from a sampling theory point of view how the thin plate splines can be interpreted a fundamental solutions of the differential operator defining the semi-norm in the Beppo-Levi space $BL_2(\mathbb{R}^2)$, and thus radial basis functions can be viewed as *Green's functions*.

A third optimality result is in the context of quasi-interpolation, i.e.,

Theorem 5.5.6 *Suppose $\Phi \in C(\Omega \times \Omega)$ is a strictly conditionally positive definite kernel with respect to the finite-dimensional space $P \subseteq C(\Omega)$. Suppose \mathcal{X} is P -unisolvent and $\mathbf{x} \in \Omega$ is fixed. Let $u_j^*(\mathbf{x})$, $j = 1, \dots, N$, be the cardinal basis functions for interpolation with Φ . Then*

$$\left| f(\mathbf{x}) - \sum_{j=1}^N f(\mathbf{x}_j) u_j^*(\mathbf{x}) \right| \leq \left| f(\mathbf{x}) - \sum_{j=1}^N f(\mathbf{x}_j) u_j \right|$$

for all choices of $u_1, \dots, u_N \in \mathbb{R}$ with $\sum_{j=1}^N u_j p(\mathbf{x}_j) = p(\mathbf{x})$ for any $p \in P$.

Theorem 5.5.6 is proved in [634]. It says in particular that the minimum norm interpolant $\mathcal{P}f$ is also more accurate (in the pointwise sense) than any linear combination of the given data values.