Hamburger Beiträge
zur Angewandten Mathematik

Scattered Data Approximation by Positive Definite Kernel Functions

Armin Iske

Nr. 2011-13
July 2011

Armin Iske*

SCATTERED DATA APPROXIMATION BY POSITIVE DEFINITE KERNEL FUNCTIONS

Abstract. Kernel functions are suitable tools for scattered data interpolation and approximation. We first review basic features of kernel-based multivariate interpolation, before we turn to the construction and the characterization of positive definite kernels and their associated reproducing kernel Hilbert spaces. The optimality of the resulting kernel-based interpolation scheme is shown. Moreover, we analyze the conditioning of the reconstruction problem, before we prove stability estimates for the proposed interpolation method. We finally discuss kernel-based penalized least squares approximation, where we provide more recent results concerning the stability and the convergence of the approximation method.

1. Introduction

This contribution gives an introduction to selected aspects of multivariate scattered data approximation by positive definite kernel functions. For the convenience of the reader, we keep the presentation widely self-contained, where we first review basic features of kernel-based Lagrange interpolation, before we discuss more advanced topics of kernel-based interpolation and approximation. Relevant aspects of the subject are covered in five sections:

Lagrange interpolation from scattered data. We explain the problem of multivariate Lagrange interpolation. This leads us to positive definite functions, whose suitability for scattered data interpolation is demonstrated.

Native reproducing kernel Hilbert space. We introduce a native Hilbert space, whose reproducing kernel is given by the positive definite function of the interpolation scheme. We analyze the properties of the native space.

Optimality of the reconstruction scheme. We show that interpolation by positive definite kernels is optimal w.r.t. (a) energy minimization; (b) best approximation; (c) norm minimization of the pointwise error functionals.

Stability of the reconstruction scheme. We analyze the conditioning of the interpolation problem and we provide bounds on the associated Lebesgue constant. We prove useful stability estimates for the interpolation method.

Penalized least squares approximation. We discuss penalized least squares approximation by positive definite kernels. Recent results concerning the well-posedness, the stability, and the convergence of the approximation method are proven.

*This contribution is based on a *Lezione Lagrangiana* given in Turin on 23rd February 2011.

2. Lagrange Interpolation from Scattered Data

2.1. Discussion of the Interpolation Problem

To explain Lagrange interpolation from multivariate scattered data, suppose that a vector

$$f_X = (f(x_1), \dots, f(x_n))^T \in \mathbb{R}^n$$

of discrete function values, sampled from an unknown function $f : \mathbb{R}^d \to \mathbb{R}$ at a finite point set $X = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$, $d \geq 1$, is given. Since we do not make any assumptions on the distribution of the points in X, we say that the given data f_X is *scattered*.

Lagrange interpolation from scattered data requires computing a *suitable* interpolant $s : \mathbb{R}^d \to \mathbb{R}$ satisfying $s_X = f_X$, i.e.,

$$s(x_j) = f(x_j), \quad \text{for all } 1 \leq j \leq n. \tag{1}$$

A standard approach for doing so is to assume that s lies in a fixed finite dimensional linear function space

$$\mathcal{S} = \text{span}\{s_1, \dots, s_n\},$$

with basis $\mathcal{B} = \{s_1, \dots, s_n\}$, so that s can be represented as a unique linear combination

$$s = \sum_{j=1}^{n} c_j s_j \tag{2}$$

of the basis functions $s_1, \dots, s_n$.

In this case, solving the interpolation problem (1) requires solving the linear system

$$V_{\mathcal{B},X} \cdot c = f_X$$

for the unknown coefficients $c = (c_1, \dots, c_n)^T \in \mathbb{R}^n$ of s in (2), where

$$V_{\mathcal{B},X} = (s_j(x_k))_{1 \leq j,k \leq n} \in \mathbb{R}^{n \times n}$$

denotes the *Vandermonde matrix* for the basis $\mathcal{B}$.

It is desirable to select a basis $\mathcal{B}$, and thus a function space $\mathcal{S}$, such that the interpolation problem (1) has for any choice of scattered interpolation points X a unique solution. In other words, we require that the Vandermonde matrix $V_{\mathcal{B},X}$ is *regular* for any finite point set X. This leads us to the following observation.

THEOREM 1. *Let $\mathcal{S}$ denote a finite dimensional linear function space with basis $\mathcal{B} = \{s_1, \dots, s_n\}$. Moreover, let $X = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$ be a finite point set. Then the following properties are equivalent.*

(a) The Vandermonde matrix $V_{\mathcal{B},X}$ is regular.

(b) If $s \in \mathcal{S}$ vanishes on X, i.e, $s_X = 0$, then $s \equiv 0$.

(c) The interpolation problem $s_X = f_X$ has a unique solution $s \in \mathcal{S}$.

Proof. The linear map $L_X : \mathcal{S} \to \mathbb{R}^n$, defined by

$$L_X(s) = s_X \qquad \text{for } s \in \mathcal{S}$$

is, w.r.t. basis $\mathcal{B}$, represented by the Vandermonde matrix $V_{\mathcal{B},X}$. Note that the above properties (a)-(c) can be reformulated as three equivalent statements: (a) L_X is bijective; (b) L_X is injective; (c) L_X is surjective. □

DEFINITION 1. *A linear function space $\mathcal{S}$ satisfying the conditions (b),(c) of Theorem 1 for any finite point set $X \subset \mathbb{R}^d$ is said to be a* `Haar space`. *In this case, any basis $\mathcal{B}$ of $\mathcal{S}$ is called a* `Chebyshev system` *on $\mathbb{R}^d$.*

In the case of multivariate interpolation, however, the basis $\mathcal{B}$ should depend on X, i.e., $\mathcal{B} \equiv \mathcal{B}(X)$. This is due to the classical *Mairhuber-Curtis* theorem, named after J. Mairhuber and P.C. Curtis, as independently proven in their works [18] (in 1956) and [7] (in 1959).

THEOREM 2 (Mairhuber-Curtis). *For $d \geq 2$ and $n \geq 2$ there is no Chebyshev system $\mathcal{B} = \{s_1, \ldots, s_n\}$ on $\mathbb{R}^d$, i.e., there is $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, such that the Vandermonde matrix $V_{\mathcal{B},X}$ is singular.*

Proof. Suppose $\mathcal{B} = \{s_1, \ldots, s_n\}$, $n \geq 2$, is for $d \geq 2$ a Chebyshev system on $\mathbb{R}^d$. Moreover, let $X_1 = (x_1, x_2, x_3, \ldots, x_n)$ contain n pairwise distinct points in $\mathbb{R}^d$. Then there is a closed continuous curve $\gamma : [0,1] \to \mathbb{R}^d$ whose path $\Gamma = \{\gamma(t) : t \in [0,1]\}$ contains x_1 and x_2, but no other point from the set $\{x_3, \ldots, x_n\}$, i.e., $x_1, x_2 \in \Gamma$ and $x_j \notin \Gamma$ for all $3 \leq j \leq n$.

Now we can exchange the positions of x_1 and x_2 in X_1 by (continuously) moving the points x_1, x_2 along the path Γ, without any coincidence between x_1 and x_2. This exchange yields $X_2 = (x_2, x_1, x_3, \ldots, x_n)$. Note that $\det(V_{\mathcal{B},X_1}) = -\det(V_{\mathcal{B},X_2})$, since by the exchange of x_1 and x_2 the first two columns in $V_{\mathcal{B},X_1}$ are swapped.

Therefore, there is one point set $X_0 = \{\gamma(t_1), \gamma(t_2), x_3, \ldots, x_n\} \subset \mathbb{R}^d$, $t_1, t_2 \in [0,1]$, of pairwise distinct points, $\gamma(t_1) \neq \gamma_2(t_2)$, satisfying $\det(V_{\mathcal{B},X_0}) = 0$, due to the continuity of the determinant. But this is, for $\mathcal{B} = \{s_1, \ldots, s_n\}$, in contradiction to property (a) of Theorem 1. □

In conclusion, to ensure unique interpolation, the basis $\mathcal{B} = \{s_1, \ldots, s_n\}$ must necessarily depend on the interpolation points X, due to the Mairhuber-Curtis theorem. A straightforward approach for doing so is to let the j-th basis function s_j depend on the j-th interpolation point x_j, $1 \leq j \leq n$, i.e., we assume

$$s_j \equiv K(\cdot, x_j) \qquad \text{for } 1 \leq j \leq n$$

for some continuous function $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.

This leads us to the linear *reconstruction space*

$$\mathcal{S}_X = \text{span}\{K(\cdot, x_j) : 1 \leq j \leq n\} \qquad \text{for } X = \{x_1, \ldots, x_n\},$$

where in a modified ansatz for solving $s_X = f_X$ we now assume

$$s(x) = \sum_{j=1}^{n} c_j K(x, x_j) \tag{3}$$

for the form of the interpolant s. Therefore, solving $s_X = f_X$ with assuming (3) boils down to solving the linear system

$$A_X \cdot c = f_X$$

for $c = (c_1, \ldots, c_n)^T \in \mathbb{R}^n$, with a Vandermonde matrix

$$A_X = (K(x_j, x_k))_{1 \le j,k \le n} \in \mathbb{R}^{n \times n},$$

whose basis $\mathcal{B} = \{K(\cdot, x_j) : 1 \le j \le n\}$ does now depend on X.

2.2. Lagrange Interpolation by Positive Definite Functions

For the sake of unique interpolation, the matrix A_X must be regular for *all* possible choices of X, according to Theorem 1. To this end, it is sufficient to require that A_X is symmetric and positive definite for any X. To guarantee symmetry for A_X, the function K must be symmetric, i.e., $K(x,y) = K(y,x)$ for all $x, y \in \mathbb{R}^d$. To ensure that A_X is positive definite for any X, we moreover require that K is a positive definite function.

DEFINITION 2. *A symmetric function $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is said to be* `positive definite` *on $\mathbb{R}^d$, $K \in \mathbf{PD}_d$, iff the matrix $A_X = (K(x_j, x_k))_{1 \le j,k \le n}$ is positive definite for all possible choices of finite point sets $X \subset \mathbb{R}^d$.*

We conclude the above discussion as follows.

THEOREM 3. *Let $K \in \mathbf{PD}_d$. Then, for any finite point set $X \subset \mathbb{R}^d$, the following statements are true.*

(a) The matrix A_X is positive definite.

(b) If $s \in \mathcal{S}_X$ vanishes on X, i.e, $s_X = 0$, then $s = 0$.

(c) The interpolation problem $f_X = s_X$ has a unique solution

$$s(x) = \sum_{j=1}^{n} c_j K(x, x_j) \in \mathcal{S}_X,$$

whose coefficient vector $c = (c_1, \ldots, c_n)^T \in \mathbb{R}^n$ is the solution of the linear system

$$A_X \cdot c = f_X.$$

□

Relying on the above theorem, we can further conclude that – due to the well-posedness of the interpolation scheme – there is for any point set $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ a unique *Lagrange basis* $\{\ell_1, \ldots, \ell_n\} \subset \mathcal{S}_X$ satisfying

$$\ell_j(x_k) = \delta_{jk} = \begin{cases} 1 & \text{for } j = k \\ 0 & \text{for } j \neq k \end{cases} \qquad \text{for all } 1 \leq j,k \leq n.$$

PROPOSITION 1. *The Lagrange basis is uniquely given by the linear system*

$$A_X \ell(x) = R(x) \qquad \textit{for } x \in \mathbb{R}^d \tag{4}$$

where we let

$$\ell(x) = (\ell_1(x), \ldots, \ell_n(x))^T \in \mathbb{R}^n \textit{ and } R(x) = (K(x,x_1), \ldots, K(x,x_n))^T \in \mathbb{R}^n.$$

The unique interpolant $s \in \mathcal{S}_X$ satisfying $s_X = f_X$ can be represented as

$$s(x) = \langle f_X, \ell(x) \rangle, \tag{5}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of the Euclidean space $\mathbb{R}^n$.

Proof. Note that for any $x = x_j$, $1 \leq j \leq n$, the right hand side $R(x_j)$ in (4) coincides with the j-th column of A_X, and so $\ell(x_j) = e_j \in \mathbb{R}^n$ is the j-th unit vector in $\mathbb{R}^n$. Moreover, by $\ell(x) = A_X^{-1} R(x)$ any Lagrange basis function

$$\ell_j(x) = e_j^T A_X^{-1} R(x) \qquad \text{for } 1 \leq j \leq n \tag{6}$$

can be represented as a unique linear combination of the functions $K(x,x_1), \ldots, K(x,x_n)$ in $R(x)$, and so $\ell_j \in \mathcal{S}_X$ for all $1 \leq j \leq n$.

From property (c) in Theorem 3, the Lagrange representation of s in (5) can be obtained by

$$s(x) = \langle c, R(x) \rangle = \langle A_X^{-1} f_X, R(x) \rangle = \langle f_X, A_X^{-1} R(x) \rangle = \langle f_X, \ell(x) \rangle.$$

□

2.3. Basic Constructions of Positive Definite Functions

In this subsection, we discuss the construction of positive definite functions. But let us first list some of their elementary properties. To this end, let $K \in \mathbf{PD}_d$ and $X = \{x\}$ for some $x \in \mathbb{R}^d$. Then, the (diagonal) entry in $A_X \in \mathbb{R}^{1 \times 1}$ is positive, i.e., $K(x,x) > 0$. Next, if we let $X = \{x,y\}$ for some $x,y \in \mathbb{R}^d$, $x \neq y$, then $\det(A_X) > 0$, which implies $K(x,y)^2 < K(x,x)K(y,y)$.

In our subsequent construction of positive definite functions we assume

$$K(x,y) := \Phi(x - y) \qquad \text{for } x,y \in \mathbb{R}^d$$

for an even function $\Phi : \mathbb{R}^d \to \mathbb{R}$. Moreover, we say that Φ is *positive definite*, i.e., $\Phi \in \mathbf{PD}_d$, iff $K \in \mathbf{PD}_d$. We express the above properties for $K \in \mathbf{PD}_d$ as follows.

REMARK 1. Let $\Phi : \mathbb{R}^d \to \mathbb{R}$ be even and positive definite, i.e., $\Phi \in \mathbf{PD}_d$. Then,

(a) $\Phi(0) > 0$;

(b) $|\Phi(x)| < \Phi(0)$ for all $x \in \mathbb{R}^d \setminus \{0\}$.

In the subsequent discussion, we apply the normalization $\Phi(0) = 1$. □

Now let us discuss the construction of positive definite functions. This is done by using the Fourier transform, here written in the unsymmetric form

$$\hat{f}(\omega) := \int_{\mathbb{R}^d} f(x) e^{-i\langle x,\omega\rangle}\, dx \qquad \text{for } f \in L^1(\mathbb{R}^d).$$

The following basic result relies on the classical *Bochner theorem*, dating back to the 1932 lecture notes [5] of Salomon Bochner, a somewhat modified variant of which is as follows.

THEOREM 4 (Bochner). *Suppose that $\Phi : \mathbb{R}^d \to \mathbb{R}$ is even and continuous. Moreover, assume that Φ has a continuous Fourier transform $\hat{\Phi}$ satisfying the Fourier inversion formula*

$$\Phi(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} \hat{\Phi}(\omega) e^{i\langle x,\omega\rangle}\, d\omega.$$

If $\hat{\Phi} \not\equiv 0$ is non-negative on $\mathbb{R}^d$, then $K(x,y) = \Phi(x-y)$ is positive definite.

Proof. Suppose that $\hat{\Phi} \in \mathscr{C}(\mathbb{R}^d) \setminus \{0\}$ is non-negative on $\mathbb{R}^d$.

In this case, the quadratic form

$$c^T A_X c = \sum_{j,k=1}^{n} c_j c_k \Phi(x_j - x_k) = (2\pi)^{-d} \int_{\mathbb{R}^d} \left| \sum_{j=1}^{n} c_j e^{i\langle x_j,\omega\rangle} \right|^2 \hat{\Phi}(\omega)\, d\omega$$

is, for any $c = (c_1, \ldots, c_n)^T \in \mathbb{R}^n$ and any $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, non-negative, i.e., $c^T A_X c \geq 0$. If $c^T A_X c = 0$, then the (analytic) *symbol function*

$$S(\omega) \equiv S_{c,X}(\omega) = \sum_{j=1}^{n} c_j e^{i\langle x_j,\omega\rangle} \qquad \text{for } \omega \in \mathbb{R}^d$$

must vanish identically on $\mathbb{R}^d$, due to the non-negativity of the continuous Fourier transform $\hat{\Phi} \not\equiv 0$. But $S \equiv 0$ implies $c = 0$, which completes our proof. □

Now let us make three relevant examples for positive definite radial functions Φ, whose positive definiteness can be shown by using Bochner's characterization of Theorem 4.

EXAMPLE 1. The `Gaussian` function

$$\Phi(x) = e^{-\|x\|_2^2} \qquad \text{for } x \in \mathbb{R}^d$$

is for any $d \geq 1$ positive definite on $\mathbb{R}^d$. In this case, $\hat{\Phi}(\omega) = e^{-\|\omega\|_2^2/4} > 0$, and so $K(x,y) = \Phi(-\|x-y\|_2^2) \in \mathbf{PD}_d$, where $\|\cdot\|_2$ is the Euclidean norm on $\mathbb{R}^d$. □

EXAMPLE 2. The `inverse multiquadric`

$$\Phi(x) = \left(1 + \|x\|_2^2\right)^{-1/2}$$

is for any $d \geq 1$ positive definite on $\mathbb{R}^d$. In this case,

$$\hat{\Phi}(s) = I_{(d-1)/2}(s) \cdot s^{-(d-1)/2} > 0 \qquad \text{for } s = \|\omega\|_2,$$

where

$$I_\nu(z) = \sum_{j=0}^{\infty} \frac{(z/2)^{\nu+2j}}{j!\Gamma(\nu+j+1)} \qquad \text{for } z \in \mathbb{C} \setminus \{0\}$$

denotes the modified Bessel function of the third kind. For relevant properties of I_ν, we refer to [1]. □

EXAMPLE 3. The `radial characteristic functions` [3]

$$\Phi(x) = (1 - \|x\|_2)_+^\beta = \begin{cases} (1 - \|x\|_2)^\beta & \text{for } \|x\|_2 < 1 \\ 0 & \text{for } \|x\|_2 \geq 1 \end{cases}$$

are for $d \geq 2$ positive definite on $\mathbb{R}^d$, provided that $\beta \geq (d+1)/2$. In this case, the Fourier transform $\hat{\Phi}$ of Φ can (up to some positive constant) be represented as

$$\hat{\Phi}(s) = s^{-(d/2+\beta+1)} \int_0^s (s-t)^\beta t^{d/2} J_{(d-2)/2}(t)\, dt > 0 \qquad \text{for } s = \|\omega\|_2,$$

where

$$J_\nu(z) = \sum_{j=0}^{\infty} \frac{(-1)^j (z/2)^{\nu+2j}}{j!\Gamma(\nu+j+1)} \qquad \text{for } z \in \mathbb{C} \setminus \{0\}$$

is the usual Bessel function of the first kind. More details on these earlier examples for compactly supported multivariate positive definite functions can be found in [11]. □

Now that we have provided three explicit examples for positive definite (radial) functions, we remark that the characterization of Bochner's theorem allows us to construct even larger classes of positive definite functions. This is done by using convolutions. Recall that for any pair $f, g \in L^1(\mathbb{R}^d)$ of functions, the Fourier transform maps the *convolution product* $f * g \in L^1(\mathbb{R}^d)$,

$$(f * g)(x) = \int_{\mathbb{R}^d} f(x-y) g(y)\, dy \qquad \text{for } f, g \in L^1(\mathbb{R}^d),$$

of f and g onto the product of their Fourier transforms, i.e.,

$$\widehat{f * g} = \hat{f} \cdot \hat{g} \qquad \text{for } f, g \in L^1(\mathbb{R}^d).$$

If we let $g(x) = f^*(x) = f(-x)$, this yields

$$\widehat{f * f^*} = \hat{f} \cdot \overline{\hat{f}} = |\hat{f}|^2 \qquad \text{for } f \in L^1(\mathbb{R}^d),$$

which is a simple method to construct positive definite functions.

COROLLARY 1. *For any function $\Psi \in L^1(\mathbb{R}^d) \setminus \{0\}$ its autocorrelation*

$$\Phi(x) = (\Psi * \Psi^*)(x) = \int_{\mathbb{R}^d} \Psi(x-y)\Psi(-y)\, dy$$

is positive definite.

Proof. For $\Psi \in L^1(\mathbb{R}^d) \setminus \{0\}$, we have $\Phi \in L^1(\mathbb{R}^d) \setminus \{0\}$, and so $\hat{\Phi} \in \mathscr{C}(\mathbb{R}^d) \setminus \{0\}$. Moreover, the (continuous) Fourier transform $\hat{\Phi} = |\hat{\Psi}|^2 \geq 0$ of the (continuous) autocorrelation $\Phi = \Psi * \Psi^*$ is non-negative, so that Φ is positive definite by Bochner's theorem. □

The practical value of this construction, however, is rather limited. This is because the autocorrelations $\Psi * \Psi^*$ are rather awkward to evaluate. Instead, one would prefer to work with explicit analytic expressions for the functions $\Phi = \Psi * \Psi^*$.

We remark that the basic idea of Corollary 1 has led to the construction of *compactly supported* positive definite (radial) functions, dating back to earlier Göttingen works of Schaback & Wendland [21] (in 1993), Wu [25] (in 1994), and Wendland [22] (in 1995). In their constructions, explicit formulae were given for autocorrelations $\Phi = \Psi * \Psi^*$, whose generators $\Psi(x) = \psi(\|x\|_2)$, $x \in \mathbb{R}^d$, are specific *radially symmetric* and compactly supported ansatz functions $\psi : [0,\infty) \to \mathbb{R}$. This has provided a large family of continuous, radially symmetric, and compactly supported functions $\Phi = \Psi * \Psi^*$, as they were later popularized by Wendland [22], who used the radial characteristic functions of Example 3 for Ψ to obtain piecewise polynomial positive definite compactly supported radial functions of minimal degree. For further details concerning the construction of compactly supported positive definite radial functions, we refer to the survey [20] of Schaback.

3. Native Reproducing Kernel Hilbert Spaces

The discussion of this section is concerning *reproducing kernel Hilbert spaces $\mathcal{F}$* which are generated by positive definite functions K. In particular, for any fixed $K \in \mathbf{PD}_d$, the positive definite function K is shown to be the reproducing kernel of its associated Hilbert space $\mathcal{F} \equiv \mathcal{F}_K$, whose structure is entirely determined by the properties of K. Therefore, $\mathcal{F}$ is also referred to as the *native space* of K.

To introduce $\mathcal{F}$, we first define, for a fixed positive definite $K \in \mathbf{PD}_d$, the *reconstruction space*

$$\mathcal{S} = \{s \in \mathcal{S}_X : X \subset \mathbb{R}^d, |X| < \infty\}$$

containing all (potential) interpolants of the form

$$s(x) = \sum_{j=1}^{n} c_j K(x, x_j) \tag{7}$$

for some $c = (c_1, \ldots, c_n)^T \in \mathbb{R}^n$ and $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$.

Note that any $s \in \mathcal{S}$ in (7) can be rewritten as

$$s(x) \equiv s_\lambda(x) := \lambda^y K(x,y) \qquad \text{for } \lambda = \sum_{j=1}^{n} c_j \delta_{x_j} \tag{8}$$

where δ_x is the usual Dirac δ-functional, defined by $\delta_x(f) = f(x)$, and λ^y in (8) denotes action of the linear functional λ on variable y.

This gives rise to define the *dual space*

$$\mathcal{L} = \left\{ \lambda = \sum_{j=1}^{n} c_j \delta_{x_j} : c = (c_1, \ldots, c_n)^T \in \mathbb{R}^n \text{ and } X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d \right\}$$

containing all finite linear combinations of δ-functionals.

3.1. Topology of the Reconstruction Space and Duality Relation

Now the dual space $\mathcal{L}$ can be equipped with the inner product

$$(\lambda, \mu)_K := \lambda^x \mu^y K(x,y) = \sum_{j=1}^{n_\lambda} \sum_{k=1}^{n_\mu} c_j d_k K(x_j, y_k) \qquad \text{for } \lambda, \mu \in \mathcal{L},$$

where

$$\lambda = \sum_{j=1}^{n_\lambda} c_j \delta_{x_j} \in \mathcal{L} \quad \text{and} \quad \mu = \sum_{k=1}^{n_\mu} d_k \delta_{y_k} \in \mathcal{L}.$$

By $\|\cdot\|_K := (\cdot,\cdot)_K^{1/2}$, $\mathcal{L}$ is a normed linear space. Likewise, via the duality relation in (8), we can equip $\mathcal{S}$ with the inner product

$$(s_\lambda, s_\mu)_K := (\lambda, \mu)_K \qquad \text{for } s_\lambda, s_\mu \in \mathcal{S}$$

and norm $\|\cdot\|_K = (\cdot,\cdot)_K^{1/2}$. Note that the normed linear spaces $\mathcal{S}$ and $\mathcal{L}$ are isometric isomorphic, $\mathcal{S} \cong \mathcal{L}$, via the linear bijection $\lambda \mapsto s_\lambda$, and by

$$\|\lambda\|_K = \|s_\lambda\|_K \qquad \text{for all } \lambda \in \mathcal{L}.$$

Now let us make a few examples for inner products and norms in $\mathcal{L}$ and $\mathcal{S}$.

EXAMPLE 4. For any pair of point evaluation functionals $\delta_{z_1}, \delta_{z_2} \in \mathcal{L}$, with $z_1, z_2 \in \mathbb{R}^d$, their inner product is given by

$$(\delta_{z_1}, \delta_{z_2})_K = \delta_{z_2}^x \delta_{z_1}^y K(x,y) = K(z_2, z_1) = \Phi(z_2 - z_1).$$

Moreover, for the norm of any $\delta_z \in \mathcal{L}$, $z \in \mathbb{R}^d$, we obtain

$$\|\delta_z\|_K^2 = (\delta_z, \delta_z)_K = \delta_z^x \delta_z^y K(x,y) = K(z,z) = \Phi(0) = 1.$$

Likewise,

$$(K(\cdot,z_1),K(\cdot,z_2))_K = K(z_2,z_1) = \Phi(z_2 - z_1) \qquad \text{for all } z_1,z_2 \in \mathbb{R}^d$$

and

$$\|K(\cdot,z)\|_K = \|\delta_z\|_K = 1 \qquad \text{for all } z \in \mathbb{R}^d.$$

□

To extend this elementary example, we regard, for fixed $X = \{x_1,\dots,x_n\} \subset \mathbb{R}^d$, the linear bijection operator $G : \mathbb{R}^n \to \mathcal{S}_X$, defined as

$$G(c) = \sum_{j=1}^{n} c_j K(\cdot,x_j) \qquad \text{for } c = (c_1,\dots,c_n)^T \in \mathbb{R}^n. \tag{9}$$

PROPOSITION 2. *For any $X = \{x_1,\dots,x_n\} \subset \mathbb{R}^d$, we have*

$$(G(c),G(d))_K = \langle c,d\rangle_{A_X} \qquad \textit{for all } c,d \in \mathbb{R}^n,$$

where

$$\langle c,d\rangle_{A_X} := c^T A_X d \qquad \textit{for } c,d \in \mathbb{R}^n$$

is the inner product generated by the positive definite matrix A_X. In particular, G is an isometry by

$$\|G(c)\|_K = \|c\|_{A_X} \qquad \textit{for all } c \in \mathbb{R}^n,$$

where $\|\cdot\|_{A_X} := \langle\cdot,\cdot\rangle_{A_X}^{1/2}$.

Proof. Note that

$$(G(c),G(d))_K = \sum_{j,k=1}^{n} c_j d_k (K(\cdot,x_j),K(\cdot,x_k))_K = c^T A_X d = \langle c,d\rangle_{A_X}$$

holds for all $c = (c_1,\dots,c_n)^T \in \mathbb{R}^n$ and $d = (d_1,\dots,d_n)^T \in \mathbb{R}^n$. □

The above result gives rise to study the *dual operator* $G^* : \mathcal{S}_X \to \mathbb{R}^n$ of G.

PROPOSITION 3. *The dual operator $G^* : \mathcal{S}_X \to \mathbb{R}^n$ of G in (9), satisfying*

$$(G(c),s)_K = \langle c, G^*(s)\rangle \qquad \textit{for all } c \in \mathbb{R}^n \textit{ and all } s \in \mathcal{S}_X \tag{10}$$

is given by

$$G^*(s) = s_X \qquad \textit{for } s \in \mathcal{S}_X.$$

Proof. Note that for any $s \in \mathcal{S}_X$, there is a unique $d \in \mathbb{R}^n$ satisfying $G(d) = s$, so that

$$(G(c),s)_K = (G(c),G(d))_K = \langle c,d\rangle_{A_X} = \langle c, A_X d\rangle = \langle c, s_X\rangle \qquad \text{for all } c \in \mathbb{R}^n,$$

in which case our assertion follows from (10). □

Another important example is concerning inner products and norms of Lagrange basis functions.

PROPOSITION 4. *For $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, the inner product of the Lagrange basis functions $\ell_j \in \mathcal{S}_X$, satisfying $\ell_j(x_k) = \delta_{jk}$ for all $1 \leq j,k \leq n$, is given as*

$$(\ell_j, \ell_k)_K = a_{jk}^{-1} \qquad \text{for all } 1 \leq j,k \leq n,$$

where $A_X^{-1} = (a_{jk}^{-1})_{1 \leq j,k \leq n} \in \mathbb{R}^{n \times n}$. In particular, the norm of $\ell_j \in \mathcal{S}_X$ is given as

$$\|\ell_j\|_K^2 = a_{jj}^{-1} \qquad \text{for all } 1 \leq j \leq n.$$

□

Proof. Recalling the representation of the Lagrange basis functions ℓ_j in (6), we obtain

$$(\ell_j, \ell_k)_K = e_j^T A_X^{-1} A_X A_X^{-1} e_k = e_j^T A_X^{-1} e_k = a_{jk}^{-1} \qquad \text{for } 1 \leq j,k \leq n.$$

□

From Example 4 and Proposition 4 we see that

$$A_X = ((\delta_{x_j}, \delta_{x_k})_K)_{1 \leq j,k \leq n} \in \mathbb{R}^{n \times n} \quad \text{and} \quad A_X^{-1} = ((\ell_j, \ell_k)_K)_{1 \leq j,k \leq n} \in \mathbb{R}^{n \times n}$$

are Gramian matrices.

Next we discuss the construction of *orthonormal* bases.

PROPOSITION 5. *For any $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, let*

$$A_X = Q^T D Q$$

denote the eigendecomposition of the symmetric matrix $A_X \in \mathbb{R}^{n \times n}$ with orthogonal factor $Q \in \mathbb{R}^{n \times n}$ and diagonal $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \in \mathbb{R}^{n \times n}$, containing the positive eigenvalues $\sigma_1 > \sigma_2 > \ldots > \sigma_n > 0$ of A_X. Then, the functions

$$\varphi_j(x) = e_j^T D^{-1/2} Q \cdot R(x) \qquad \text{for } 1 \leq j \leq n \tag{11}$$

are an orthonormal basis of $\mathcal{S}_X$.

Proof. Due to their representation in (11), where $R(x) = (K(x,x_1), \ldots, K(x,x_n))^T$, any φ_j is a linear combination of the basis functions $\{K(\cdot,x_j)\}_{j=1}^n$, and so $\varphi_j \in \mathcal{S}_X$. Moreover, from Proposition 2 we obtain

$$(\varphi_j, \varphi_k)_K = e_j^T D^{-1/2} Q A_X Q^T D^{-1/2} e_k = \langle e_j, e_k \rangle = \delta_{jk} \quad \text{for all } 1 \leq j,k \leq n.$$

□

Let us finally express the inner product between functions $s, \tilde{s} \in \mathcal{S}_X$ in a different way (other than in Proposition 2). To this end, we work with their Lagrange representations

$$s = \sum_{j=1}^{n} s(x_j)\ell_j \quad \text{and} \quad \tilde{s} = \sum_{k=1}^{n} \tilde{s}(x_k)\ell_k.$$

Moreover, we define the inner product

$$\langle c, d \rangle_{A_X^{-1}} := c^T A_X^{-1} d \qquad \text{for } c, d \in \mathbb{R}^n$$

and norm $\|\cdot\|_{A_X^{-1}} = \langle \cdot, \cdot \rangle_{A_X^{-1}}^{1/2}$, generated by the positive definite matrix A_X^{-1}.

EXAMPLE 5. For $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, the inner product between any pair $s, \tilde{s} \in \mathcal{S}_X$ can be expressed as

$$(s, \tilde{s})_K = \langle s_X, \tilde{s}_X \rangle_{A^{-1}},$$

Indeed, from Proposition 4 we obtain

$$(s, \tilde{s})_K = \sum_{j,k=1}^{n} s(x_j)\tilde{s}(x_k)(\ell_j(x), \ell_k(x))_K = \sum_{j,k=1}^{n} s(x_j)\tilde{s}(x_k)a_{jk}^{-1} = s_X^T A^{-1} \tilde{s}_X = \langle s_X, \tilde{s}_X \rangle_{A^{-1}}.$$

In particular, we have

$$\|s\|_K = \|s_X\|_{A_X^{-1}} \qquad \text{for all } s \in \mathcal{S}_X.$$

□

3.2. Construction of the Reproducing Kernel Hilbert Space

By completion of the inner product spaces $\mathcal{L}$ and $\mathcal{S}$ w.r.t. their norms $\|\cdot\|_K$, we obtain the Hilbert spaces

$$\mathcal{D} := \overline{\mathcal{L}} \qquad \text{and} \qquad \mathcal{F} := \overline{\mathcal{S}}.$$

We extend the linear bijection $\lambda \mapsto s_\lambda$ to $\mathcal{D}$ and $\mathcal{F}$, which immediately yields

PROPOSITION 6. *The Hilbert spaces $\mathcal{D}$ and $\mathcal{F}$ are isometric isomorphic,*

$$\mathcal{D} \cong \mathcal{F},$$

via the bijection $\lambda \mapsto s_\lambda$ and by

$$\|\lambda\|_K = \|s_\lambda\|_K \qquad \textit{for all } \lambda \in \mathcal{D}.$$

□

REMARK 2. Any functional $\mu \in \mathcal{D}$ is continuous on the Hilbert space $\mathcal{F}$ by the Cauchy-Schwarz inequality

$$|\mu(s_\lambda)| = |\mu^x \lambda^y K(x,y)| = |(\mu, \lambda)_K| \leq \|\mu\|_K \cdot \|\lambda\|_K = \|\mu\|_K \cdot \|s_\lambda\|_K.$$

In particular, any point evaluation functional $\delta_x \in \mathcal{L}$, $x \in \mathbb{R}^d$, is continuous on $\mathcal{F}$, i.e.,

$$|\delta_x(f)| \leq \|\delta_x\|_K \cdot \|f\|_K = \|f\|_K \qquad \text{for all } f \in \mathcal{F}.$$

□

Resorting to basic functional analysis, we see that $\mathcal{F}$ is a reproducing kernel Hilbert space. But let us first recall some facts about *reproducing kernels* [2].

DEFINITION 3. *Let $\mathcal{H}$ denote a Hilbert space of functions $f : \mathbb{R}^d \to \mathbb{R}$. Then, a function $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is said to be a* `reproducing kernel` *for $\mathcal{H}$, iff $K(\cdot, x) \in \mathcal{H}$, for all $x \in \mathbb{R}^d$, and*

$$(K(\cdot, x), f)_{\mathcal{H}} = f(x) \qquad \textit{for all } f \in \mathcal{H} \textit{ and all } x \in \mathbb{R}^d.$$

Next we show a well-known result from functional analysis.

THEOREM 5. *A Hilbert space $\mathcal{H}$ of functions has a reproducing kernel, iff all point evaluation functionals are continuous on $\mathcal{H}$.*

Proof. Suppose K is a reproducing kernel for $\mathcal{H}$. Then, by

$$|\delta_x(f)| = |f(x)| = |(K(\cdot, x), f)_{\mathcal{H}}| \leq \|K(\cdot, x)\|_{\mathcal{H}} \cdot \|f\|_{\mathcal{H}} \qquad \text{for } x \in \mathbb{R}^d$$

any point evaluation functional δ_x is continuous on $\mathcal{H}$.

As for the converse, suppose that all point evaluation functionals δ_x are continuous on $\mathcal{H}$. Then, due to the *Riesz representation theorem*, there is, for any $x \in \mathbb{R}^d$, one function $k_x \in \mathcal{H}$ satisfying

$$f(x) = \delta_x(f) = (k_x, f)_{\mathcal{H}} \qquad \text{for all } f \in \mathcal{H}$$

and so $K(\cdot, x) = k_x$ is a reproducing kernel for $\mathcal{H}$. □

REMARK 3. A reproducing kernel K for $\mathcal{H}$ is unique. Indeed, if $\tilde{K}$ is another reproducing kernel for $\mathcal{H}$, then by

$$(\tilde{K}(\cdot, x), f)_K = f(x) \qquad \text{for all } f \in \mathcal{H}$$

we obtain

$$(k_x - \tilde{k}_x, f)_K = 0 \qquad \text{for all } f \in \mathcal{H},$$

where we let $k_x := K(\cdot, x)$ and $\tilde{k}_x := \tilde{K}(\cdot, x)$. But this implies $k_x \equiv \tilde{k}_x$ and so $K \equiv \tilde{K}$. □

3.3. The Madych-Nelson Theorem

Now we are in a position where we can show that the positive definite function K is the (unique) reproducing kernel for the Hilbert space $\mathcal{F} \equiv \mathcal{F}_K$. To this end, we apply the variational theory from the seminal papers by Madych and Nelson [15, 16, 17], whose central result relies on the representation

$$(\lambda^y K(\cdot,y), s_\mu)_K = (s_\lambda, s_\mu)_K = (\lambda,\mu)_K = \lambda^x \mu^y K(x,y) = \lambda(s_\mu) \tag{12}$$

for $\lambda \in \mathcal{L}$ and $s_\mu \in \mathcal{S}$. By continuous extension to $\lambda \in \mathcal{D}$ and $s_\mu \in \mathcal{F}$ in (12), we already obtain the important Madych-Nelson theorem.

THEOREM 6 (Madych-Nelson). *For any $\lambda \in \mathcal{D}$ and $f \in \mathcal{F}$ we have*

$$(\lambda^y K(\cdot,y), f)_K = \lambda(f).$$

□

This allows us to show the stated result.

COROLLARY 2. *The positive definite function K is the unique reproducing kernel of the Hilbert space $\mathcal{F}$.*

Proof. On the one hand, for $\delta_x \in \mathcal{L}$, $x \in \mathbb{R}^d$, we find

$$\delta_x^y K(\cdot,y) = K(\cdot,x) \in \mathcal{F} \qquad \text{for all } x \in \mathbb{R}^d,$$

on the other hand, letting $\lambda = \delta_x \in \mathcal{L}$ in the Madych-Nelson theorem, we obtain

$$(K(\cdot,x), f)_K = f(x) \qquad \text{for all } f \in \mathcal{F} \text{ and all } x \in \mathbb{R}^d.$$

□

Another useful consequence of the Madych-Nelson theorem is as follows.

COROLLARY 3. *Every function $f \in \mathcal{F}$ is continuous on $\mathbb{R}^d$, i.e., $\mathcal{F} \subset \mathscr{C}(\mathbb{R}^d)$.*

Proof. Recall that we assume continuity for K. Then, by

$$|f(x) - f(y)| = |(K(\cdot,x) - K(\cdot,y), f)_K| \le \|f\|_K \cdot \|K(\cdot,x) - K(\cdot,y)\|_K$$

and

$$\begin{aligned} &\|K(\cdot,x) - K(\cdot,y)\|_K^2 \\ &\quad = \quad (K(\cdot,x),K(\cdot,x))_K - 2(K(\cdot,x),K(\cdot,y))_K + (K(\cdot,y),K(\cdot,y))_K \\ &\quad = \quad K(x,x) - 2K(x,y) + K(y,y) \end{aligned}$$

any $f \in \mathcal{F}$ is a continuous function. □

4. Optimality of the Reconstruction Scheme

In this section, we discuss some approximation properties of the proposed scattered data reconstruction method. To this end, we prove a sequence of corollaries from the Madych-Nelson theorem to show that the kernel-based Lagrange interpolation scheme is *optimal* in three different senses.

4.1. Variational Property

The first optimality result relies on the Pythagoras theorem, here stated as follows.

COROLLARY 4. *For $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, let $s \in \mathcal{S}_X$ denote the unique interpolant to $f \in \mathcal{F}$ satisfying $s_X = f_X$. Then, the Pythagoras theorem*

$$\|f\|_K^2 = \|s\|_K^2 + \|f - s\|_K^2$$

holds.

Proof. Recall that $s \in \mathcal{S}_X$ can be represented as $s = \lambda^y K(\cdot, y)$, with a functional $\lambda \in \mathcal{L}$ supported on X, i.e., $\mathrm{supp}(\lambda) = X$. Moreover, by the Madych-Nelson theorem we find

$$(s, g)_K = 0 \quad \text{for all } g \in \mathcal{F} \text{ with } \lambda(g) = 0,$$

i.e., s is orthogonal to the kernel of $\lambda \in \mathcal{L}$. But this implies

$$(s, f - s)_K = 0,$$

since $f_X - s_X = 0$ and $\mathrm{supp}(\lambda) = X$. Therefore, we have

$$\|f\|_K^2 = \|f - s + s\|_K^2 = \|f - s\|_K^2 + 2(f - s, s)_K + \|s\|_K^2 = \|f - s\|_K^2 + \|s\|_K^2$$

which completes our proof. □

The result of Corollary 4 immediately yields the following variational property of the interpolation scheme.

COROLLARY 5. *For $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ and $f_X \in \mathbb{R}^n$, the interpolant $s \in \mathcal{S}_X$ satisfying $s_X = f_X$ is the unique minimizer of the energy functional $\|\cdot\|_K$ among all interpolants from $\mathcal{F}$ to the data f_X, i.e.,*

$$\|s\|_K \leq \|g\|_K \qquad \text{for all } g \in \mathcal{F} \text{ with } g_X = f_X.$$

□

Now we are in a position where we can compute the native space norm

$$\|I_X(f)\|_K := \sup_{f \in \mathcal{F} \setminus \{0\}} \frac{\|I_X(f)\|_K}{\|f\|_K}$$

of the interpolation operator $I_X : \mathcal{F} \to \mathcal{S}_X$, which returns on any argument $f \in \mathcal{F}$ its unique interpolant $s \in \mathcal{S}_X$ satisfying $s_X = f_X$.

THEOREM 7. *For $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, the norm $\|I_X\|_K$ of the interpolation operator $I_X : \mathcal{F} \to \mathcal{S}_X$ is one, i.e.,*

$$\|I_X\|_K = 1.$$

Proof. The variational property in Corollary 5 implies

$$\|I_X(f)\|_K \le \|f\|_K \qquad \text{for all } f \in \mathcal{F},$$

and so $\|I_X\|_K \le 1$. Due to the projection property $I_X(s) = s$, for all $s \in \mathcal{S}_X$, equality is attained by any $s \in \mathcal{S}_X$, i.e.,

$$\|I_X(s)\|_K = \|s\|_K \qquad \text{for all } f \in \mathcal{S}_X,$$

and therefore $\|I_X\|_K = 1$. □

The above result allows us to conclude that Lagrange interpolation w.r.t. the native space norm $\|\cdot\|_K$ is well-conditioned.

4.2. Orthogonality and Best Approximation

Our next result shows that the proposed interpolation scheme provides, on input f_X, the unique *best approximation* to f according to the following definition.

DEFINITION 4. *A function $s^* \in \mathcal{S}_X$ is said to be the* `best approximation` *to $f \in \mathcal{F}$ from $\mathcal{S}_X$ w.r.t. $\|\cdot\|_K$, iff*

$$\|s^* - f\|_K < \|s - f\|_K \qquad \textit{for all } s \in \mathcal{S}_X \setminus \{s^*\}.$$

Now we can show the following properties of the interpolation scheme.

COROLLARY 6. *Let $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ and $f \in \mathcal{F}$. Then, the interpolant $s^* \in \mathcal{S}_X$ satisfying $s^*_X = f_X$ is*

(a) the unique orthogonal projection of $f \in \mathcal{F}$ onto $\mathcal{S}_X$;

(b) the unique best approximation to $f \in \mathcal{F}$ from $\mathcal{S}_X$ w.r.t. $\|\cdot\|_K$.

Proof. We first show property (a), i.e., $s^* - f \perp \mathcal{S}_X$, which is equivalent to

$$(K(\cdot, x_j), s^* - f)_K = 0 \qquad \text{for all } 1 \le j \le n,$$

or, by using the Madych-Nelson theorem, equivalent to

$$s^*(x_j) = (K(\cdot, x_j), s^*)_K = (K(\cdot, x_j), f)_K = f(x_j) \qquad \text{for all } 1 \le j \le n,$$

i.e., $s^*_X = f_X$, as covered by our assumption on $s^* \in \mathcal{S}$.

To prove (b), let $s \in \mathcal{S}_X$. Then $(s^* - s, s^* - f)_K = 0$, due to (a), and so

$$\|s - f\|_K^2 = \|s - s^* + s^* - f\|_K^2 = \|s - s^*\|_K^2 + \|s^* - f\|_K^2.$$

But this already implies

$$\|s^* - f\|_K < \|s - f\|_K \qquad \text{for } s \neq s^*.$$

□

4.3. Norm Minimality of the Pointwise Error Functional

Our next optimality result is concerning pointwise error estimates. To this end, we regard, for any fixed $x \in \mathbb{R}^d$, the pointwise error

$$\varepsilon_x(f) = f(x) - s(x) \tag{13}$$

between $f \in \mathcal{F}$ and the interpolant $s \in \mathcal{S}_X$, satisfying $s_X = f_X$.

Due to the Lagrange representation of s in (5) the error functional ε_x in (13) is given as

$$\varepsilon_x = \delta_x - \ell(x)^T \delta_X \in \mathcal{L} \qquad \text{where } \delta_X = (\delta_{x_1}, \ldots, \delta_{x_n})^T. \tag{14}$$

Moreover, by the Madych-Nelson theorem, we obtain the representation

$$\varepsilon_x(f) = (\varepsilon_x^y K(\cdot, y), f)_K \qquad \text{for } f \in \mathcal{F}, \tag{15}$$

which in turn allows us to bound the pointwise error $\varepsilon_x(f)$ as follows.

COROLLARY 7. *Let $s \in \mathcal{S}_X$ satisfy $s_X = f_X$. Then, the pointwise error $\varepsilon_x(f)$ satisfies the estimate*

$$|f(x) - s(x)| = |\varepsilon_x(f)| \leq \|\varepsilon_x\|_K \cdot \|f\|_K, \tag{16}$$

where the norm of the error functional can be represented as

$$\|\varepsilon_x\|_K^2 = 1 - \ell(x)^T A_X \ell(x) = 1 - \|\ell(x)\|_{A_X}^2,$$

so that

$$0 \leq \|\varepsilon_x\|_K \leq 1 \qquad \textit{for all } x \in \mathbb{R}^d. \tag{17}$$

The error bound in (16) is sharp, where equality is attained for the function

$$f_x = \varepsilon_x^y K(\cdot, y) \in \mathcal{F}.$$

Proof. The pointwise error bound for $|\varepsilon_x(f)|$ in (16) follows directly from the representation (15) in combination with the Cauchy-Schwarz inequality.

The norm of the error functional ε_x can be computed by

$$\begin{aligned} \|\varepsilon_x\|_K^2 &= (\varepsilon_x, \varepsilon_x)_K = (\delta_x - \ell(x)^T \delta_X, \delta_x - \ell(x)^T \delta_X)_K \\ &= 1 - 2\ell(x)^T R(x) + \ell(x)^T A_X \ell(x) = 1 - \ell(x)^T A_X \ell(x), \end{aligned}$$

where we have used (4) and $K(x,x) = \Phi(0) = 1$. The bound on $\|\varepsilon_x\|_K$ in (17) follows then from the positive definiteness of A_X.

To show sharpness for (16), consider the function $f_x = \varepsilon_x^y K(\cdot,y)$, for which, by the Madych-Nelson theorem, equality is attained:

$$|\varepsilon_x(f_x)| = |(\varepsilon_x^y K(\cdot,y), f_x)_K| = (f_x, f_x)_K = (\varepsilon_x, \varepsilon_x)_K = \|\varepsilon_x\|_K \cdot \|f_x\|_K.$$

□

Now let us finally turn to the pointwise optimality of the interpolation scheme. To this end, we consider, for fixed $x \in \mathbb{R}^d$, the variation of the coefficients $\ell \equiv \ell(x)$ for the *quasi interpolants*

$$s = \ell^T f_X = \sum_{j=1}^{n} \ell_j f(x_j) \qquad \text{for } \ell = (\ell_1, \ldots, \ell_n)^T \in \mathbb{R}^n$$

to show the norm minimality of the error functional ε_x in (14). This immediately leads us to the unconstrained optimization problem

$$1 - 2\ell^T R(x) + \ell^T A_X \ell \longrightarrow \min_{\ell \in \mathbb{R}^n}$$

whose unique solution is given by the solution of the system $A_X \ell = R(x)$.

COROLLARY 8. *For any $x \in \mathbb{R}^d$, the error functional $\varepsilon_x \equiv \varepsilon_x(s)$ in (14) is the unique norm minimizer among all error functionals of the form*

$$\varepsilon_x^{(\ell)}(f) = \delta_x - \ell^T \delta_X \in \mathcal{L} \qquad \textit{for } \ell \in \mathbb{R}^n,$$

i.e.,

$$\|\varepsilon_x\|_K < \|\varepsilon_x^{(\ell)}\|_K \qquad \textit{for all } \ell \in \mathbb{R}^n \textit{ with } A_X \ell \neq R(x).$$

□

5. Stability of the Reconstruction Scheme

The discussion of this section is devoted to the stability of the Lagrange interpolation scheme. To this end, we first provide some basic Riesz stability estimates, before we analyze the conditioning of the interpolation problem.

5.1. Riesz Bases and Basic Stability Estimates

In this section, we follow along the lines of basic wavelet analysis, where, for the sake of stability, the construction of Riesz bases is of vital importance. It is rather straightforward to show that, for any point set $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, the basis functions $\mathcal{B} = \{K(\cdot,x_j)\}_{j=1}^n$ are a *Riesz basis* of $\mathcal{S}_X$, so that the stability estimate

$$A\|c\|_2^2 \leq \left\| \sum_{j=1}^{n} c_j K(\cdot,x_j) \right\|_K^2 \leq B\|c\|_2^2 \quad \text{for all } c = (c_1, \ldots, c_n)^T \in \mathbb{R}^n \tag{18}$$

is satisfied with *Riesz constants* $0 < A \leq B < \infty$.

To be more precise on this, we provide the following theorem.

THEOREM 8. *For $X = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$, the functions $\mathcal{B} = \{K(\cdot, x_j)\}_{j=1}^n$ are a Riesz basis of $\mathcal{S}_X$, i.e., the Riesz stability estimate (18) holds, where the Riesz constants in (18) are given by the smallest eigenvalue $A = \sigma_{\min}(A_X)$ and the largest eigenvalue $B = \sigma_{\max}(A_X)$ of the matrix A_X.*

Proof. Recall the linear bijection operator $G : \mathbb{R}^n \to \mathcal{S}_X$ from Proposition 2, where

$$\|G(c)\|_K^2 = \|c\|_{A_X}^2 = c^T A_X c \qquad \text{for all } c \in \mathbb{R}^n$$

which implies (18) with optimal constants $A = \sigma_{\min}(A_X)$ and $B = \sigma_{\max}(A_X)$. □

Next we recall that any Riesz basis $\mathcal{B}$ has a unique dual Riesz basis $\tilde{\mathcal{B}}$, satisfying a biorthonormality relation, where the Riesz constants of $\tilde{\mathcal{B}}$ are given by the reciprocal values of the Riesz constants for $\mathcal{B}$. In our particular situation, the dual Riesz basis of $\mathcal{B}$ in Theorem 8 is given by the Lagrange basis of the interpolation scheme.

THEOREM 9. *For $X = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$, the Lagrange functions $\tilde{\mathcal{B}} = \{\ell_j\}_{j=1}^n$ are the unique dual Riesz basis of $\mathcal{B} = \{K(\cdot, x_j)\}_{j=1}^n$, satisfying the biorthonormality relation*

$$(K(\cdot, x_j), \ell_k)_K = \delta_{jk} \qquad \textit{for all } 1 \leq j, k \leq n. \tag{19}$$

Moreover, the stability estimate

$$\frac{1}{B}\|f_X\|_2^2 \leq \left\| \sum_{j=1}^n f(x_j)\ell_j \right\|_K^2 \leq \frac{1}{A}\|f_X\|_2^2 \qquad \textit{for all } f_X \in \mathbb{R}^n \tag{20}$$

holds with Riesz constants $0 < 1/B \leq 1/A < \infty$, where $A = \sigma_{\min}(A_X)$ and $B = \sigma_{\max}(A_X)$.

Proof. The biorthonormality (19) follows from the Madych-Nelson theorem,

$$(K(\cdot, x_j), \ell_k)_K = \ell_k(x_j) = \delta_{jk} \qquad \text{for all } 1 \leq j, k \leq n.$$

Moreover, the stability estimate in (20) is due the representation

$$\left\| \sum_{j=1}^n f(x_j)\ell_j \right\|_K^2 = \|f_X\|_{A_X^{-1}}^2 = f_X^T A_X^{-1} f_X \qquad \text{for all } f_X \in \mathbb{R}^n,$$

in combination with the bounds

$$\sigma_{\min}(A_X^{-1})\|f_X\|_2^2 \leq f_X^T A_X^{-1} f_X \leq \sigma_{\max}(A_X^{-1})\|f_X\|_2^2 \qquad \text{for all } f_X \in \mathbb{R}^n,$$

so that (20) holds with

$$1/B = \sigma_{\max}^{-1}(A_X) = \sigma_{\min}(A_X^{-1}) \quad \text{and} \quad 1/A = \sigma_{\min}^{-1}(A_X) = \sigma_{\max}(A_X^{-1}).$$

□

The above result implies the following dual stability estimate to (18).

COROLLARY 9. *For $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, the stability estimate*

$$A\|s\|_K^2 \le \|s_X\|_2^2 \le B\|s\|_K^2 \qquad \text{for all } s \in \mathcal{S}_X$$

holds with $A = \sigma_{\min}(A_X)$ and $B = \sigma_{\max}(A_X)$.

Proof. Letting $f = s \in \mathcal{S}_X$ in the stability estimate (20), we obtain

$$\frac{1}{B}\|s_X\|_2^2 \le \|s\|_K^2 = \left\| \sum_{j=1}^n s(x_j)\ell_j \right\|_K^2 \le \frac{1}{A}\|s_X\|_2^2 \qquad \text{for all } f \in \mathcal{S}_X,$$

with $A = \sigma_{\min}(A_X)$ and $B = \sigma_{\max}(A_X)$, or, in other words,

$$\frac{1}{B}\|G^*(s)\|_2^2 \le \|s\|_K^2 \le \frac{1}{A}\|G^*(s)\|_2^2 \qquad \text{for all } f \in \mathcal{S}_X,$$

where $G^* : \mathcal{S}_X \to \mathbb{R}^n$, $s \mapsto s_X$, is the dual operator of G in (9), see Proposition 3. □

REMARK 4. Due to the duality of the Riesz basis $\mathcal{B} = \{K(\cdot, x_j)\}_{j=1}^n$ and the Riesz basis $\tilde{\mathcal{B}} = \{\ell_j\}_{j=1}^n$ in $\mathcal{S}_X$, any $f \in \mathcal{S}_X$ can uniquely be represented w.r.t. $\mathcal{B}$ and $\tilde{\mathcal{B}}$ as

$$f = \sum_{j=1}^n (f, K(\cdot, x_j))_K \ell_j = \sum_{j=1}^n (f, \ell_j)_K K(\cdot, x_j) \qquad \text{for all } f \in \mathcal{S}_X.$$

We can immediately see this by the two relations $(f, K(\cdot, x_j))_K = f(x_j)$ and

$$(f, \ell_j)_K = e_j^T A_X^{-1} f_X = \langle e_j, c \rangle = c_j \qquad \text{for all } 1 \le j \le n.$$

□

5.2. Absolute Condition Number and Lebesgue Constant

In this section we discuss, for some closed domain $\Omega \subset \mathbb{R}^d$, the conditioning of the interpolation problem w.r.t. the L_∞-topology of the continuous functions $\mathscr{C}(\Omega)$. To this end, let for any finite set $X = \{x_1, \ldots, x_n\} \subset \Omega$ of interpolation points, $I_X : \mathscr{C}(\Omega) \to \mathcal{S}_X$ denote the interpolation operator, which returns on any argument $f \in \mathscr{C}(\Omega)$ its unique interpolant $s \in \mathcal{S}_X$ satisfying $s_X = f_X$.

Recall that the *absolute condition number* of the interpolation problem is given by the smallest number $\kappa_\infty \equiv \kappa_{\infty,X}$ satisfying

$$\|I_X f\|_{L_\infty(\Omega)} \le \kappa_\infty \cdot \|f\|_{L_\infty(\Omega)} \qquad \text{for all } f \in \mathscr{C}(\Omega).$$

Therefore, κ_∞ is the operator norm $\|I_X\|_\infty$ of I_X w.r.t. the norm $\|\cdot\|_{L_\infty(\Omega)}$ on $\mathscr{C}(\Omega)$, i.e., $\kappa_\infty = \|I_X\|_\infty$. Now, the operator norm $\|I_X\|_\infty$, and so the absolute condition number κ_∞, can be computed as follows.

THEOREM 10. *For $X = \{x_1, \ldots, x_n\} \subset \Omega$, the norm $\|I_X\|_\infty$ of the interpolation operator $I_X : \mathscr{C}(\Omega) \to \mathcal{S}_X$ is given by the Lebesgue constant*

$$\Lambda_\infty := \max_{x\in\Omega} \sum_{j=1}^{n} |\ell_j(x)| = \max_{x\in\Omega} \|\ell(x)\|_1. \tag{21}$$

Proof. For any $f \in \mathscr{C}(\Omega)$, let $s = I_X(f) \in \mathcal{S}_X \subset \mathscr{C}(\Omega)$ denote the unique interpolant to f on X satisfying $f_X = s_X$. Using the Lagrange representation of s in (5), we obtain the estimate

$$\|I_X(f)\|_{L_\infty(\Omega)} = \|s\|_{L_\infty(\Omega)} \le \max_{x\in\Omega} \sum_{j=1}^{n} |\ell_j(x)| \cdot |f(x_j)| \le \Lambda_\infty \cdot \|f\|_{L_\infty(\Omega)},$$

and therefore $\|I_X\|_\infty \le \Lambda_\infty$.

In order to see that $\|I_X\|_\infty \ge \Lambda_\infty$, suppose that the maximum of Λ_∞ in (21) is attained at $x^* \in \Omega$. Moreover, let $g \in \mathscr{C}(\Omega)$ be a function satisfying $g(x_j) = \mathrm{sign}(\ell_j(x^*))$, for all $1 \le j \le n$, and $\|g\|_{L_\infty(\Omega)} = 1$. Then,

$$\|I_X(g)\|_{L_\infty(\Omega)} \ge (I_X(g))(x^*) = \sum_{j=1}^{n} \ell_j(x^*) g(x_j) = \sum_{j=1}^{n} |\ell_j(x^*)| = \Lambda_\infty$$

and so $\|I_X(g)\|_{L_\infty(\Omega)} \ge \Lambda_\infty$, which implies $\|I_X\|_\infty \ge \Lambda_\infty$. Altogether, $\|I_X\|_\infty = \Lambda_\infty$. □

Now let us compute an upper bound for the Lebesgue constant Λ_∞.

THEOREM 11. *For $X = \{x_1, \ldots, x_n\} \subset \Omega$, the Lebesgue constant Λ_∞ is bounded above by*

$$\Lambda_\infty \le \sum_{j=1}^{n} \sqrt{a_{jj}^{-1}},$$

where $a_{jj}^{-1} > 0$ is the j-th diagonal element of the matrix A_X^{-1}, for $1 \le j \le n$.

Proof. Suppose that the maximum of Λ_∞ in (21) is attained at $x^* \in \Omega$. Then, by using Example 4 and Proposition 4, we obtain

$$\Lambda_\infty = \sum_{j=1}^{n} |\ell_j(x^*)| = \sum_{j=1}^{n} |\delta_{x^*}(\ell_j)| \le \sum_{j=1}^{n} \|\delta_{x^*}\|_K \cdot \|\ell_j\|_K = \sum_{j=1}^{n} \|\ell_j\|_K = \sum_{j=1}^{n} \sqrt{a_{jj}^{-1}}.$$

□

REMARK 5. Note that by

$$1 = |\delta_{x_j}(\ell_j)| \le \|\delta_{x_j}\|_K \cdot \|\ell_j\|_K = \|\ell_j\|_K = \sqrt{a_{jj}^{-1}} \qquad \text{for } 1 \le j \le n$$

we have $a_{jj}^{-1} \ge 1$, for all $1 \le j \le n$, which implies

$$\Lambda_\infty \le \sum_{j=1}^{n} \sqrt{a_{jj}^{-1}} \le \sum_{j=1}^{n} a_{jj}^{-1} = \mathrm{trace}(A_X^{-1}) \le n \cdot \sigma_{\max}(A_X^{-1}).$$

□

To compute a lower bound for Λ_∞, we use the estimate

$$\begin{aligned}\|\ell(x)\|_2^2 &\geq \sigma_{\max}^{-1}(A_X)\ell(x)^T A_X \ell(x) = \sigma_{\min}(A_X^{-1})\ell(x)^T A_X \ell(x)\\ &= \sigma_{\min}(A_X^{-1}) R(x)^T A_X^{-1} R(x) \geq \sigma_{\min}^2(A_X^{-1})\|R(x)\|_2^2,\end{aligned}$$

which, in combination with $\|\ell(x)\|_1 \geq \|\ell(x)\|_2$, yields

$$\Lambda_\infty = \max_{x\in\Omega}\|\ell(x)\|_1 \geq \max_{x\in\Omega}\|\ell(x)\|_2 \geq \sigma_{\min}(A_X^{-1})\max_{x\in\Omega}\|R(x)\|_2 \geq \sigma_{\min}(A_X^{-1}),$$

where we used

$$\max_{x\in\Omega}\|R(x)\|_2 \geq \max_{1\leq j\leq n}\|R(x_j)\|_2 \geq 1.$$

We conclude this section as follows.

THEOREM 12. *For $X = \{x_1, \ldots, x_n\} \subset \Omega$, the Lebesgue constant Λ_∞ is bounded below by*

$$\Lambda_\infty \geq \sigma_{\min}(A_X^{-1}) = \sigma_{\max}^{-1}(A_X).$$

□

We finally remark that the bounds in Remark 5 and Theorem 12 are rather coarse. But the purpose of our basic discussion is to relate the range of the Lebesgue constant Λ_∞ to the spectrum of the matrix A_X.

6. Penalized Least Squares Approximation

This section is devoted to *penalized least squares approximation* (PLSA), an alternative approach for scattered data fitting other than Lagrange interpolation. PLSA makes sense especially in situations where the given data is contaminated with noise, or, when the given data is very large. We first prove the well-posedness of the PLSA problem, before we characterize its solution. Then, we turn to the sensitivity of PLSA, and, moreover, we prove the convergence of the PLSA solution to the solution of the classical *least squares approximation* (LSA) problem. For a more general account to PLSA problems, we refer to [10].

6.1. Problem Formulation and Characterization of the Solution

To explain PLSA, let a finite point set $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$ be given. Moreover, suppose that $Y = \{y_1, \ldots, y_n\}$ is a subset of X, $Y \subset X$, whose size $|Y| = n$ is much smaller than the size $|X| = N$ of X, i.e., $n \ll N$. Then, our aim is to reconstruct an unknown function $f \in \mathcal{F}$ from its values $f_X \in \mathbb{R}^N$ by solving the following problem of *penalized least squares approximation*.

Problem ($\mathbf{P}_\alpha$). On given function values $f_X \in \mathbb{R}^N$ and for $\alpha \geq 0$, find $s_\alpha \in \mathcal{S}_Y$ satisfying

$$\frac{1}{N}\|(f - s_\alpha)_X\|_2^2 + \alpha\|s_\alpha\|_K^2 = \min_{s\in\mathcal{S}_Y}\left(\frac{1}{N}\|(f-s)_X\|_2^2 + \alpha\|s\|_K^2\right). \tag{22}$$

Before we discuss the well-posedness of problem (P_α), let us first provide some comments. Note that the first term in the cost functional (22),

$$E_X(f,s) = \frac{1}{N}\|(f-s)_X\|_2^2 = \frac{1}{N}\sum_{k=1}^{N}|f(x_k)-s(x_k)|^2 \qquad \text{for } s \in \mathcal{S}_Y$$

is the usual *mean square error* between f and s on X. Moreover, the *energy functional*

$$J(s) = \|s\|_K^2 \qquad \text{for } s \in \mathcal{S}_Y$$

in (22) measures the *smoothness* of s. Therefore, the *penalty parameter* $\alpha \geq 0$ in (22) serves to balance between the smoothness of the solution s_α and its mean square error to f on X.

In fact, note that for $\alpha = 0$, the corresponding optimization problem (P_0) coincides with the classical problem of linear *least squares approximation* (LSA) [4, 14]. In contrast, for large α, the smoothing term $J(s)$ will dominate over the mean square error $E_X(f,s)$. In particular, for $\alpha \to \infty$, the solution s_α in (22) tends to zero.

Now let us show that the problem (P_α) has a unique solution. To this end, we first rewrite the mean square error as

$$E_X(f,s) = \frac{1}{N}\|f_X - A_{X,Y}c\|_2^2 \qquad \text{for } s \in \mathcal{S}_Y, \tag{23}$$

where

$$A_{X,Y} = (K(x_k,y_j))_{1\leq k\leq N; 1\leq j\leq n} \in \mathbb{R}^{N\times n},$$

and where $c = (c_1,\ldots,c_n)^T \in \mathbb{R}^n$ are the coefficients of

$$s = \sum_{j=1}^{n} c_j K(\cdot,y_j) \in \mathcal{S}_Y. \tag{24}$$

The representation of $E_X(f,s)$ in (23), in combination with the representation $J(s) = c^T A_Y c$ of the energy functional, allows us to rewrite the cost functional in (22) as

$$E_X(f,s) + \alpha J(s) = \frac{1}{N}\|f_X - A_{X,Y}c\|_2^2 + \alpha c^T A_Y c \qquad \text{for } s \in \mathcal{S}_Y. \tag{25}$$

Now we see that the problem (P_α) is well-posed.

THEOREM 13. *For any $\alpha \geq 0$, the penalized least squares approximation problem (P_α) has a unique solution $s_\alpha \in \mathcal{S}_Y$ of the form (24), where the coefficients $c_\alpha \in \mathbb{R}^n$ of s_α are given by the unique solution of the normal equations*

$$\left[\frac{1}{N}A_{X,Y}^T A_{X,Y} + \alpha A_Y\right] c_\alpha = \frac{1}{N}A_{X,Y}^T f_X. \tag{26}$$

Proof. Requiring a vanishing gradient of the cost functional in (25), this immediately leads us to the normal equations (26), whose coefficient matrix is, for any $\alpha \geq 0$, positive definite. Therefore, problem (P_α) has a unique solution. $\square$

An alternative characterization for the unique solution s_α to the optimization problem (22) relies on classical approximation theory.

THEOREM 14. *For any $\alpha \geq 0$, the solution $s_\alpha \equiv s_\alpha(f) \in \mathcal{S}_Y$ of problem (P_α) can be characterized by the conditions*

$$\frac{1}{N}\langle (f-s_\alpha)_X, s_X\rangle = \alpha(s_\alpha, s)_K \qquad \textit{for all } s \in \mathcal{S}_Y. \tag{27}$$

Proof. We introduce a semi-inner product $[\cdot,\cdot]_\alpha$ on $\mathcal{F} \times \mathcal{F}$ by

$$[(f,g),(\tilde{f},\tilde{g})]_\alpha := \frac{1}{N}\langle f_X, \tilde{f}_X\rangle + \alpha(g,\tilde{g})_K \qquad \text{for } f,g,\tilde{f},\tilde{g} \in \mathcal{F},$$

which yields the semi-norm $\|\cdot\|_\alpha$ on $\mathcal{F} \times \mathcal{F}$ by

$$\|(f,g)\|_\alpha^2 = \frac{1}{N}\|f_X\|_2^2 + \alpha\|g\|_K^2 \qquad \text{for } f,g \in \mathcal{F}.$$

Note that problem (P_α) is equivalent to finding a *best approximation* $s_\alpha^* \in \mathcal{S}_Y$ satisfying

$$\|(f,0)-(s_\alpha^*,s_\alpha^*)\|_\alpha^2 = \inf_{s\in\mathcal{S}_Y} \|(f,0)-(s,s)\|_\alpha^2.$$

Resorting to standard approximation theory in Euclidean spaces, and since $\mathcal{S}_Y$ is a Hilbert space w.r.t. $[\cdot,\cdot]_\alpha$, the unique best approximation $s_\alpha^* \in \mathcal{S}_Y$ can be characterized by the orthogonality relations

$$[(f,0)-(s_\alpha^*,s_\alpha^*),(s,s)]_\alpha = 0 \qquad \text{for all } s \in \mathcal{S}_Y,$$

which are equivalent to the conditions in (27) for $s_\alpha^* = s_\alpha$. $\square$

Note that the characterizations in Theorems 13 and 14 are equivalent. Indeed, with replacing $s \in \mathcal{S}_Y$ in (27) by the basis functions $K(\cdot,y_k) \in \mathcal{S}_Y$, for $1 \leq k \leq n$, the conditions in (27) can be rewritten as

$$\frac{1}{N}\langle (f-s_\alpha)_X, R(y_k)\rangle = \alpha(s_\alpha, K(\cdot,y_k))_K \qquad \text{for all } 1 \leq k \leq n, \tag{28}$$

where

$$R^T(y_k) = (K(x_1,y_k),\ldots,K(x_N,y_k)) = e_k^T A_{X,Y}^T.$$

Assuming (24) for the form of s_α, with coefficients $c_\alpha \in \mathbb{R}^n$, we have

$$(s_\alpha)_X = A_{X,Y}c_\alpha \in \mathbb{R}^N,$$

and so we can derive the normal equations from (28): On the one hand, the left hand side in (28) can be rewritten as

$$\begin{aligned}\frac{1}{N}\langle (f-s_\alpha)_X, R(y_k)\rangle &= \frac{1}{N}\left[R^T(y_k)f_X - R^T(y_k)A_{X,Y}c_\alpha\right] \\ &= \frac{1}{N}\left[e_k^T A_{X,Y}^T f_X - e_k^T A_{X,Y}^T A_{X,Y}c_\alpha\right].\end{aligned}$$

On the other hand, the right hand side in (28) can be rewritten as

$$\alpha(s_\alpha, K(\cdot,y_k))_K = \alpha s_\alpha(y_k) = \alpha e_k^T A_Y c_\alpha,$$

where we used the identity $(s_\alpha, K(\cdot,y_k))_K = s_\alpha(y_k)$ from the Madych-Nelson theorem.

6.2. Stability, Sensitivity, Error Bounds, and Convergence

Let us turn to the stability of penalized least squares approximation. To this end, we bound the minimizer s_α of the cost functional in (22) as follows.

THEOREM 15. *For any* $\alpha \geq 0$*, the solution* $s_\alpha \equiv s_\alpha(f) \in \mathcal{S}_Y$ *of problem* (P_α) *satisfies the stability estimate*

$$\frac{1}{N}\|(s_\alpha - f)_X\|_2^2 + \alpha\|s_\alpha\|_K^2 \leq (1+\alpha)\|f\|_K^2.$$

Proof. Let $s_f \in \mathcal{S}_Y$ denote the (unique) interpolant to f at Y satisfying $(s_f - f)_Y = 0$. Recall $\|s_f\|_K \leq \|f\|_K$ from Corollary 5. Then,

$$\begin{aligned}\frac{1}{N}\|(s_\alpha - f)_X\|_2^2 + \alpha\|s_\alpha\|_K^2 &= \frac{1}{N}\sum_{k=1}^{N}|s_\alpha(x_k) - f(x_k)|^2 + \alpha\|s_\alpha\|_K^2 \\ &\leq \frac{1}{N}\sum_{k=1}^{N}|s_f(x_k) - f(x_k)|^2 + \alpha\|s_f\|_K^2 \\ &\leq \frac{1}{N}\sum_{x\in X\setminus Y}\|\varepsilon_x\|_K^2\cdot\|f\|_K^2 + \alpha\|f\|_K^2 \\ &= \left(\frac{1}{N}\sum_{x\in X\setminus Y}\|\varepsilon_x\|_K^2 + \alpha\right)\|f\|_K^2 \\ &\leq \left(\frac{N-n}{N} + \alpha\right)\|f\|_K^2 \leq (1+\alpha)\|f\|_K^2,\end{aligned}$$

where we have used the pointwise error estimate (16) from Subsection 4.3 and the uniform bound $\|\varepsilon_x\|_K \leq 1$ in (17). □

Next we analyze the sensitivity of problem (P_α) under the variation of the smoothing parameter $\alpha \geq 0$. To this end, we first observe that the solution $s_\alpha \equiv s_\alpha(f)$ of problem (P_α) coincides with that for the target function s_0, i.e., $s_\alpha(s_0) = s_\alpha(f)$.

LEMMA 1. *For any $\alpha \geq 0$, the solution $s_\alpha \equiv s_\alpha(f)$ of problem (P_α) satisfies*

(a) the Pythagoras theorem,

$$\|(f-s_\alpha)_X\|_2^2 = \|(f-s_0)_X\|_2^2 + \|(s_0-s_\alpha)_X\|_2^2,$$

(b) the best approximation property,

$$\|(s_0-s_\alpha)_X\|_2^2 + \alpha\|s_\alpha\|_K^2 = \min_{s\in\mathcal{S}_Y} \|(s_0-s)_X\|_2^2 + \alpha\|s\|_K^2,$$

i.e., $s_\alpha(s_0) = s_\alpha(f)$.

Proof. Recall that the solution $s_\alpha(f)$ to (P_α) is characterized by the conditions (27) in Theorem 14, where for $\alpha = 0$ we obtain the characterization

$$\frac{1}{N}\langle (f-s_0)_X, s_X\rangle = 0 \qquad \text{for all } s \in \mathcal{S}_Y. \tag{29}$$

This implies, for any $s \in \mathcal{S}_Y$, the relation

$$\begin{aligned}
\|(f-s)_X\|_2^2 &= \langle (f-s_0+s_0-s)_X, (f-s_0+s_0-s)_X\rangle \\
&= \|(f-s_0)_X\|_2^2 + 2\langle (f-s_0)_X, (s_0-s)_X\rangle + \|(s_0-s)_X\|_2^2 \\
&= \|(f-s_0)_X\|_2^2 + \|(s_0-s)_X\|_2^2,
\end{aligned}$$

and so in particular, with letting $s = s_\alpha$, (a) is proven.

As regards property (b), we subtract (29) from (27) to obtain by

$$\frac{1}{N}\langle (s_0-s_\alpha)_X, s_X\rangle = \alpha(s_\alpha, s)_K \qquad \text{for all } s \in \mathcal{S}_Y \tag{30}$$

the characterization (27) for the solution $s_\alpha(s_0)$ of (P_α) from Theorem 14. □

Next we aim to analyze the convergence of $\{s_\alpha\}_\alpha$ for $\alpha \searrow 0$. To this end, we first prove a stability estimate for s_α and an error bound for $s_\alpha - s_0$.

THEOREM 16. *Let $f \in \mathcal{F}$ and $\alpha \geq 0$. Then, the solution $s_\alpha \equiv s_\alpha(f)$ of problem (P_α) and the solution $s_0 \equiv s_0(f)$ of problem (P_0) satisfy*

(a) the stability estimate

$$\|s_\alpha\|_K \leq \|s_0\|_K,$$

(b) the error bound

$$\frac{1}{N}\|(s_\alpha-s_0)_X\|_2^2 \leq \alpha\|s_0\|_K^2.$$

Proof. Letting $s = s_0 - s_\alpha$ in (30), we obtain

$$\frac{1}{N}\|(s_0 - s_\alpha)_X\|_2^2 + \alpha\|s_\alpha\|_K^2 = \alpha(s_\alpha, s_0)_K. \tag{31}$$

Using the Cauchy-Schwarz inequality, this yields

$$\frac{1}{N}\|(s_0 - s_\alpha)_X\|_2^2 + \alpha\|s_\alpha\|_K^2 \leq \alpha\|s_\alpha\|_K \cdot \|s_0\|_K,$$

which immediately implies properties (a) and (b). □

Let us finally show the convergence of s_α to s_0 for $\alpha \searrow 0$.

THEOREM 17. *For $\alpha \searrow 0$, the solution s_α of problem (P_α) converges to the solution s_0 of problem (P_0) at the following asymptotic convergence rates.*

(a) Convergence in the native space norm,

$$\|s_\alpha - s_0\|_K^2 = O(\alpha) \qquad \textit{for } \alpha \searrow 0,$$

(b) convergence of the mean square error,

$$\frac{1}{N}\|(s_\alpha - s_0)_X\|_2^2 = o(\alpha) \qquad \textit{for } \alpha \searrow 0.$$

Proof. To prove (a), first note that

$$\|s\|_X := \|s_X\|_2 \qquad \text{for } s \in \mathcal{S}_Y$$

is a norm on $\mathcal{S}_Y$. To see the definiteness of $\|\cdot\|_X$ on $\mathcal{S}_Y$, note that $\|s\|_X = 0$ implies $s_X = 0$, in particular $s_Y = 0$, since $Y \subset X$, in which case $s = 0$.

Moreover, since $\mathcal{S}_Y$ has finite dimension, the norms $\|\cdot\|_X$ and $\|\cdot\|_K$ are equivalent on $\mathcal{S}_Y$, so that there exists a constant $C > 0$ satisfying

$$\|s\|_K \leq C\|s\|_X \qquad \text{for all } s \in \mathcal{S}_Y.$$

This, in combination with part (b) of Theorem 16, implies property (a) by

$$\|s_\alpha - s_0\|_K^2 \leq C^2\|(s_\alpha - s_0)_X\|_2^2 \leq C^2 N\alpha\|s_0\|_K^2.$$

To prove (b), we first recall (31) to obtain

$$(s_\alpha, s_0)_K = \frac{1}{\alpha}\left[\frac{1}{N}\|(s_0 - s_\alpha)_X\|_2^2 + \alpha\|s_\alpha\|_K^2\right] \qquad \text{for } \alpha > 0.$$

This in turn implies the identity

$$\|s_\alpha - s_0\|_K^2 = \|s_0\|_K^2 - \|s_\alpha\|_K^2 - \frac{2}{\alpha N}\|(s_0 - s_\alpha)_X\|_2^2 \tag{32}$$

by

$$
\begin{aligned}
\|s_\alpha - s_0\|_K^2 &= \|s_\alpha\|_K^2 - 2(s_\alpha, s_0)_K + \|s_0\|_K^2 \\
&= \|s_\alpha\|_K^2 + \|s_0\|_K^2 - \frac{2}{\alpha}\left[\frac{1}{N}\|(s_0 - s_\alpha)_X\|_2^2 + \alpha\|s_\alpha\|_K^2\right] \\
&= \|s_0\|_K^2 - \|s_\alpha\|_K^2 - \frac{2}{\alpha N}\|(s_0 - s_\alpha)_X\|_2^2.
\end{aligned}
$$

To conclude our proof, note that, by property (a), the left hand side in (32) tends to zero as $\alpha \searrow 0$, and so does the right hand side in (32) tend to zero. By the stability estimate in Theorem 16 (a), we find

$$0 \le \|s_0\|_K - \|s_\alpha\|_K \le \|s_\alpha - s_0\|_K \longrightarrow 0 \quad \text{for } \alpha \searrow 0,$$

so that $\|s_\alpha\|_K \longrightarrow \|s_0\|_K$ for $\alpha \searrow 0$. Therefore,

$$\frac{2}{\alpha N}\|(s_0 - s_\alpha)_X\|_2^2 \longrightarrow 0 \quad \text{for } \alpha \searrow 0,$$

which completes our proof for (b). □

7. Conclusions and Final Remarks

We have introduced the reader to selected aspects of kernel-based scattered data approximation. To this end, we have explained basic principles of multivariate Lagrange interpolation by positive definite kernels. This includes the construction and characterization of their associated native reproducing kernel Hilbert spaces, along with a discussion concerning the optimality and the stability of the recovery method. Moreover, we have explained the solution of penalized least squares approximation problems, where we have placed special emphasis on the stability and the convergence of the proposed kernel-based approximation method.

For the sake of simplicity, and in order to keep the presentation within reasonable page limits, we have restricted ourselves to approximation methods by *positive definite kernels*. For a comprehensive discussion on scattered data approximation by *conditionally positive definite* (radial) kernels and their applications to meshfree approximation methods, we refer the interested reader to the research monographs [6, 8, 13, 23].

Further directions for possible extensions and generalizations are concerning kernel-based approximation from scattered *Hermite-Birkhoff data* [12, 24] (rather than from plain Lagrange data), reconstruction methods for vector-valued functions (rather than for scalar-valued functions) via matrix-valued kernels [19], as well as kernel-based *multiscale methods* [9, 13], to mention but a few. For a general account to (penalized) least squares approximation problems, we refer the reader to [4, 10, 14].

Acknowledgment

This paper is based on an invited lecture in the series *Lezione Lagrangiana*, which I gave at the University of Turin on 23rd February 2011. I gratefully thank the organizers of the *Lezione Lagrangiana* lecture series for their generous support and their kind hospitality. Moreover, I wish to thank Dennis von Grudzinski and Alexander Yefremov for proofreading the paper.

References

[1] ABRAMOWITZ M. AND STEGUN I.A., *Handbook of Mathematical Functions*, Dover Publications, New York, 1970.

[2] ARONSZAJN N., *Theory of reproducing kernels*, Transactions of the AMS **68** (1950), 337–404.

[3] ASKEY R., *Radial characteristic functions*, technical report TSR # 1262, University of Wisconsin, Madison, 1973.

[4] BJØRCK Å, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.

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

[6] BUHMANN M.D., *Radial Basis Functions: Theory and Implementations*, Cambridge University Press, Cambridge, UK, 2003.

[7] CURTIS P.C., *N-parameter families and best approximation*, Pacific J. Math. **9** (1959), 1013–1027.

[8] FASSHAUER G.E., *Meshfree Approximation Methods with Matlab*, Interdisciplinary Mathematical Sciences **6**, World Scientific Publishing, Singapore, 2007.

[9] FLOATER M.S. AND ISKE A., *Multistep scattered data interpolation using compactly supported radial basis functions*, Journal of Computational and Applied Mathematics **73** 5 (1996), 65–78.

[10] V. GOLITSCHEK M., *Penalized least squares approximation problems*, Jaen Journal on Approximation Theory **1** 1 (2009), 83–96.

[11] ISKE A., *Charakterisierung bedingt positiv definiter Funktionen für multivariate Interpolationsmethoden mit radialen Basisfunktionen*. Dissertation, University of Göttingen, 1994.

[12] ISKE A., *Reconstruction of functions from generalized Hermite-Birkhoff data*, in: Approximation Theory VIII, Vol. 1: Approximation and Interpolation, C.K. Chui and L.L. Schumaker (eds.), World Scientific, Singapore, 1995, 257–264.

[13] ISKE A., *Multiresolution Methods in Scattered Data Modelling*, Lecture Notes in Computational Science and Engineering **37**, Springer-Verlag, Berlin, 2004.

[14] LAWSON C.L. AND HANSON R.J., *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, N.J., 1974.

[15] MADYCH W.R. AND NELSON S.A., *Multivariate interpolation: a variational theory*, manuscript, Iowa State University, 1983.

[16] MADYCH W.R. AND NELSON S.A., *Multivariate interpolation and conditionally positive definite functions*, Approx. Theory Appl. **4** (1988), 77–89.

[17] MADYCH W.R. AND NELSON S.A., *Multivariate interpolation and conditionally positive definite functions II*, Math. Comp. **54** (1990), 211–230.

[18] MAIRHUBER J., *On Haar's theorem concerning Chebysheff problems having unique solutions*, Proc. Am. Math. Soc. **7** (1956), 609–615.

[19] NARCOWICH F.J. AND WARD J.D., *Generalized Hermite interpolation via matrix-valued conditionally positive definite functions*, Math. Comp. **63** (1994), 661–687.

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

[21] SCHABACK R. AND WENDLAND H., *Special cases of compactly supported radial basis functions*, manuscript, University of Göttingen, 1993.

[22] WENDLAND H., *Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree*, Advances in Comp. Math. **4** (1995), 389–396.

[23] WENDLAND H., *Scattered Data Approximation*, Cambridge University Press, Cambridge, UK, 2005.

[24] WU Z., *Hermite-Birkhoff interpolation of scattered data by radial basis functions*, Approx. Theory Appl. **8** (1992), 1–10.

[25] WU Z., *Multivariate compactly supported positive definite radial functions*, Advances in Comp. Math. **4** (1995), 283–292.

AMS Subject Classification: 65D15, 65D05; 65-02

Armin ISKE
Department of Mathematics, University of Hamburg
Bundesstraße 55, D-20146 Hamburg, GERMANY
e-mail: `iske@math.uni-hamburg.de`