Approximation by Positive Definite Kernels

Robert Schaback and Holger Wendland

TEXed on February 18, 2002

Abstract

This contribution extends earlier work [16] on interpolation/approximation by positive definite basis functions in several aspects. First, it works out the relations between various types of kernels in more detail and more generality. Second, it uses the new generality to exhibit the first example of a discontinuous positive definite function. Third, it establishes the first link from (radial) basis function theory to n–widths, and finally it uses this link to prove quasi–optimality results for approximation rates of interpolation processes and decay rates for eigenvalues of integral operators having smooth kernels.

1 Kernel Functions

Let Ω be a domain in $\mathbb{R}^d$. We want to work with large–dimensional data–dependent spaces of functions on Ω. A simple way to do this is to consider functions of the form

$$s_{\alpha,X} := \sum_{j=1}^{M} \alpha_j \Phi(x_j, \cdot) \tag{1}$$

for "data sets" $X = \{x_1, \ldots, x_M\} \subseteq \Omega \subseteq \mathbb{R}^d$, coefficients $\alpha \in \mathbb{R}^M$ and a "kernel" function

$$\Phi \;:\; \Omega \times \Omega \to \mathbb{R}, \quad \Omega \subseteq \mathbb{R}^d. \tag{2}$$

We start with a short review of the basic features of kernels, but we do not follow the standard path. The functions (1) form a finite–dimensional space

$$S_{X,\Phi} := \text{ span } \{\Phi(x, \cdot) \;:\; x \in X\} \tag{3}$$

of dimension at most M. These spaces are the ones we want to work with. The union of these spaces is

$$\mathcal{S}_\Phi := \text{ span } \{\Phi(x, \cdot) \;:\; x \in \Omega\}. \tag{4}$$

Of course, everything is useless if $\Phi = 0$, and at least one would like to have linear independence of functions $\Phi(x, \cdot)$ for $x \in \Omega$.

Definition 1 *We call a kernel* (2) nondegenerate *on Ω, if for all finite data sets $X = \{x_1, \ldots, x_M\}$ the functions $\Phi(x_j, \cdot)$, $x_j \in X = \{x_1, \ldots, x_M\}$ are linearly independent over Ω.*

There are plenty of such kernels, e.g. $\Phi(x, y) := \exp(x^T y)$, $x, y \in \mathbb{R}^d$ is nondegenerate on every subset of $\mathbb{R}^d$ that contains at least an interior point.

But we also want to have a norm structure on the space (4). The simplest axiomatic way to do this is to use the kernel itself:

Definition 2 *A function* (2) *on $\Omega \subseteq \mathbb{R}^d$ that generates an inner product of the form*

$$(\Phi(x, \cdot), \Phi(y, \cdot))_\Phi = \Phi(x, y) \textit{ for all } x, y \in \Omega \tag{5}$$

on the space $\mathcal{S}_\Phi$ will be called a reproducing kernel *on Ω.*

Clearly, a reproducing kernel has simple properties like

$$\begin{array}{lcll} \Phi(x,x) & \geq & 0 & \text{for all } x \in \Omega, \\ \Phi(x,y) & = & \Phi(y,x) & \text{for all } x,y \in \Omega, \\ \Phi(x,y)^2 & \leq & \Phi(x,x)\Phi(y,y) & \text{for all } x,y \in \Omega, \end{array}$$

but we just note them in passing. Equation (5) turns $\mathcal{S}_\Phi$ into a pre–Hilbert space, and it allows to write

$$(f(\cdot),\Phi(y,\cdot))_\Phi := f(y) \text{ for all } y \in \Omega, f \in \mathcal{S}_\Phi,$$

because the equation holds for all functions $f_x(y) := \Phi(x,y)$ and thus on the whole space $\mathcal{S}_\Phi$.

The formal closure $\mathcal{N}_\Phi$ of $\mathcal{S}_\Phi$ under the inner product $(.,.)_\Phi$ will be a Hilbert space, and an abstract element f of $\mathcal{N}_\Phi$ can be interpreted as a function on Ω by

$$(f,\Phi(y,\cdot))_\Phi =: f(y) \text{ for all } y \in \Omega, f \in \mathcal{N}_\Phi, \tag{6}$$

because the left–hand side makes sense on the closure. Equation (6) is the reason why a kernel Φ is usually called reproducing with respect to a specific Hilbert space of functions: it allows to recover the function values of an element f of the Hilbert space by (6). Standard sources for results on reproducing kernel Hilbert spaces are [1, 8], while results on native spaces are compiled in [7, 15].

Definition 3 *If Φ is a reproducing kernel on $\Omega \subseteq I\!R^d$, we call the space*

$$\mathcal{N}_\Phi := \textit{clos}_{\,(.,.)_\Phi}\mathcal{S}_\Phi := \textit{clos}_{\,(.,.)_\Phi}\ \textit{span}\ \{\Phi(x,\cdot)\ :\ x \in \Omega\}$$

the native space *for Φ.*

If the kernel is just reproducing and possibly degenerate or even zero, we cannot get a rich native Hilbert space. But in many situations we have both properties, and then we get another useful notion:

Theorem 1 *If a kernel* (2) *is reproducing and nondegenerate on $\Omega \subseteq I\!R^d$, it is* (strictly) positive definite *there. This means that for all finite data sets $X = \{x_1,\ldots,x_M\}$ the matrices*

$$A_{X,\Phi} := (\Phi(x_j,x_k))_{1\leq j,k\leq M}$$

are symmetric and positive definite. The converse is also true: a positive definite kernel is nondegenerate and reproducing.

Proof: If we have a reproducing kernel Φ, the matrices $A_{X,\Phi}$ are Gramians and thus positive semidefinite. If the kernel is nondegenerate, the matrices must be positive definite, because Gramians of linearly independent functions are positive definite.

For the converse, we start with a positive definite kernel Φ and consider functions $s_{\alpha,X}$ of the form (1). We have

$$s_{\alpha,X}(x_k) = \alpha^T A_{X,\Phi} e_k,\ 1 \leq k \leq M$$

using the k–th unit vector $e_k \in I\!R^M$. By positive definiteness we can conclude that such a function can only vanish on X if the coefficients are zero. This proves that the kernel is nondegenerate, and it implies that finitely generated functions $s_{\alpha,X}$ from (4) are uniquely determined by α and X. Thus we can define a bilinear form by (5) on the functions $\Phi(x,\cdot)$ that generate $\mathcal{S}_\Phi$ and use (1) again to write

$$\alpha^T A_{X,\Phi}\alpha = \sum_{j,k=1}^{M} \alpha_j\alpha_k\Phi(x_j,x_k) = \left(\sum_{j=1}^{M}\alpha_j\Phi(x_j,\cdot),\sum_{k=1}^{M}\alpha_k\Phi(x_k,\cdot)\right)_\Phi = \|s_{\alpha,X}\|_\Phi^2 \geq 0$$

for all $\alpha \in I\!R^M, X = \{x_1,\ldots,x_M\} \subseteq \Omega \subseteq I\!R^d$ to conclude the definiteness of the bilinear form. □

There are other equivalent formulations for positive definiteness of a kernel Φ on $\Omega \subseteq I\!R^d$:

Theorem 2 *If a kernel Φ is reproducing on $\Omega \subseteq I\!R^d$, then the following properties are equivalent:*

1. *The functions $\Phi(x_j, \cdot)$ are linearly independent on Ω for all finite data sets $X = \{x_1, \ldots, x_M\} \subseteq \Omega$.*
2. *For all finite data sets $X = \{x_1, \ldots, x_M\}$ the matrices $A_{X,\Phi}$ are positive definite.*
3. *All point evaluation functionals for distinct points in Ω are linearly independent in the dual of $\mathcal{N}_\Phi$.*
4. *The native space $\mathcal{N}_\Phi$ separates points of Ω, i.e. for all finite data sets $X = \{x_1, \ldots, x_M\} \subseteq \Omega$ and all points $x_j \in X$ there is a function $f_j \in \mathcal{N}_\Phi$ such that $f_j(x_k) = \delta_{jk},\ 1 \leq j, k \leq M$.*

Proof: We already know the equivalence of properties 1 and 2. In the dual of the native space, we can use (6) to see that the Riesz representer of the point evaluation functional $\delta_x : f \mapsto f(x)$ is the function $\Phi(x, \cdot)$, and thus

$$(\delta_x, \delta_y)_{\mathcal{N}_\Phi^*} = \Phi(x, y) \text{ for all } x, y \in \Omega$$

holds in the dual of the native space. Thus the matrices $A_{X,\Phi}$ are Gramians of the point evaluation functionals $\delta_{x_j},\ x_j \in X$ in the dual of the native space, and linear independence of the functionals is equivalent to positive definiteness of the matrix.

If we have property 4, we can easily see that the point evaluation functionals for any finite point set are linearly independent, since for a vanishing linear combination we get

$$0 = \left(\sum_{k=1}^{M} \alpha_k \delta_{x_k} \right) f_j = \sum_{k=1}^{M} \alpha_k f_j(x_k) = \alpha_j,\ 1 \leq j \leq M.$$

Conversely, property 4 follows from property 2 by interpolation. We define vectors and functions

$$a_j := A_{X,\Phi}^{-1} e_j,\ f_j := s_{a_j, X},\ 1 \leq j \leq M$$

via (1). This gives

$$f_j(x_k) = s_{a_j,X}(x_k) = a_j^T A_{X,\Phi} e_k = e_j^T A_{X,\Phi}^{-1} A_{X,\Phi} e_k = e_j^T e_k = \delta_{jk},\ 1 \leq j, k \leq M.$$

□

For completeness, we add a standard observation that goes the other way round:

Theorem 3 *If $\mathcal{H}$ is a Hilbert space of functions on Ω such that all point evaluation functionals for distinct points in Ω are linearly independent in the dual of $\mathcal{H}$, then $\mathcal{H}$ is the native space of a nondegenerate reproducing kernel.*

Proof: We define Φ as the Riesz representer for the point evaluation functionals, i.e. by (6) for all $f \in \mathcal{H}$. Then we get (5) by putting $f_x(\cdot) := \Phi(x, \cdot)$ into (6), and the previous theorems yield that Φ is a positive definite reproducing kernel on Ω with its native space $\mathcal{N}_\Phi$ being necessarily a closed subspace of $\mathcal{H}$. But we can use (6) to show that an element f of $\mathcal{H}$ which is orthogonal to all $\Phi(y, \cdot)$ must vanish on Ω, and thus the spaces $\mathcal{H}$ and $\mathcal{N}_\Phi$ coincide. □

This result shows that reproducing positive definite kernels are not exotic. They automatically arise for any Hilbert space of functions where point evaluation is a continuous and nondegenerate operation.

We list a series of important special forms of kernels:

Radial Basis Functions	$\Phi(x, y)$	$=$	$\phi(\|x - y\|_2)$	for all $x, y \in I\!R^d$
Translation–invariant Kernels on $I\!R^d$	$\Phi(x, y)$	$=$	$\Psi(x - y)$	for all $x, y \in I\!R^d$
Zonal Kernels on Spheres	$\Phi(x, y)$	$=$	$\phi(x^T y)$	for all $x, y \in S^{d-1}$
Periodic Kernels on Tori	$\Phi(x, y)$	$=$	$\Psi(x - y)$	for all $x, y \in [0, 2\pi]^d$
Convolution Kernels	$\Phi(x, y)$	$=$	$\int_\Sigma \Psi(x, s)\Psi(y, s) d\mu(s)$	for all $x, y \in \Omega$
			$\Psi : \Omega \times \Sigma \to I\!R$	
Hilbert–Schmidt Kernels	$\Phi(x, y)$	$=$	$\sum_{i \in I} \lambda_i \varphi_i(x) \varphi_i(y)$	for all $x, y \in \Omega$
			$\varphi_i : \Omega \to I\!R,\ \lambda_i > 0$	for all $i \in I$

This paper focuses on Hilbert–Schmidt kernels, because it turns out that they are quite general, though they look rather special. This will be topic of the next section. But we should add some remarks on the other cases. Translation–invariant kernels occur as reproducing kernels of translation–invariant Hilbert spaces of functions on $\mathbb{R}^d$. They allow Fourier transform methods and are positive definite in $\mathbb{R}^d$, if their Fourier transform exists and is positive almost everywhere. Radial basis functions additionally have rotational symmetry. By replacing Fourier transforms by other transforms, one can deal with the other cases. Zonal kernels $\phi(x^T y)$ are positive definite, if their symmetrized spherical transform, i.e. their expansion into Legendre polynomials as functions of the cosine of the angle θ between x and y has positive coefficients. For periodic kernels on tori, one simply uses positivity of the coefficients of the Fourier series representation.

These observations immediately show that many kernels have series expansions with positive coefficients, and thus they come close to the Hilbert–Schmidt kernel form that we want to study in the next section.

2 Hilbert–Schmidt Kernels

Before we delve into the standard way of looking at those kernels, i.e. by introducing an integral operator in $L_2(\Omega)$, we want to focus on a somewhat more abstract view that does not require a link to embeddings into L_2 spaces.

Definition 4 *For each index i from a countable index set I let there be a positive weight λ_i and a function $\varphi_i : \Omega \to \mathbb{R}$ such that for all $x \in \Omega$ the condition*

$$\sum_{i \in I} \lambda_i \varphi_i^2(x) < \infty \tag{7}$$

is satisfied and such that any finite subset of the φ_i is linearly independent over Ω. Then the function

$$\Phi(x,y) = \sum_{i \in I} \lambda_i \varphi_i(x) \varphi_i(y) \ : \ \Omega \times \Omega \to \mathbb{R} \tag{8}$$

is called a Hilbert–Schmidt kernel.

Theorem 4 *Any Hilbert–Schmidt kernel Φ is a reproducing kernel on the native space*

$$\mathcal{N}_\Phi := \left\{ \sum_{i \in I} c_i \varphi_i \ : \ c_i \in \mathbb{R}, \ \sum_{i \in I} \frac{c_i^2}{\lambda_i} < \infty \right\}. \tag{9}$$

Proof: Note first that our summability condition (7) implies that the kernel series is summable. Furthermore, the functions in $\mathcal{N}_\Phi$ are well–defined because of

$$\sum_{i \in I} |c_i \varphi_i(x)| = \sum_{i \in I} \frac{|c_i|}{\sqrt{\lambda_i}} \sqrt{\lambda_i} |\varphi_i(x)|| \le \sqrt{\sum_{i \in I} \frac{c_i^2}{\lambda_i}} \sqrt{\sum_{i \in I} \lambda_i \varphi_i^2(x).}$$

By our assumption on linear independence, all finite linear combinations of the φ_i have unique coefficients, and we can define the inner product

$$\begin{aligned} (\varphi_i, \varphi_j)_\Phi &:= \frac{\delta_{ij}}{\lambda_i} \\ \left(\sum_{i \in I} c_i \varphi_i, \sum_{j \in I} d_j \varphi_j \right)_\Phi &:= \sum_{i \in I} \frac{c_i d_i}{\lambda_i} \end{aligned}$$

on these functions. We get a pre–Hilbert space whose closure is $\mathcal{N}_\Phi$. By easy calculations, all $\Phi(x, \cdot)$ are in $\mathcal{N}_\Phi$ and both (6) and (5) hold. □

Unfortunately, the linear independence assumptions of Definitions 1 and 4 differ, and we cannot conclude that a Hilbert–Schmidt kernel is nondegenerate in general. For example, if all φ_i have a common zero, the nondegeneracy fails.

Theorem 5 *If the space of all finite linear combinations of the generating functions φ_i of a Hilbert–Schmidt kernel Φ of the form* (8) *separates points of Ω in the sense of assertion 4 of Theorem 2, the kernel is nondegenerate.*

Proof: Assume there is a vanishing linear combination $s_{\alpha,X}$ for some $X = \{x_1, \ldots, x_M\} \subseteq \Omega$. Then

$$0 = \|s_{\alpha,X}\|_\Phi^2 = \sum_{j,k=1}^{M} \alpha_j \alpha_k \Phi(x_j, x_k) = \sum_{i \in I} \lambda_i \left(\sum_{j=1}^{M} \alpha_j \varphi_i(x_j) \right)^2$$

implies that all sums $\sum_{j=1}^{M} \alpha_j \varphi_i(x_j)$ are zero. Taking linear combinations with the coefficients of point–separating functions, we can conclude that α vanishes. □

We now know that under mild assumptions all Hilbert–Schmidt kernels are positive definite reproducing kernels of some Hilbert space. We now assert the converse, but we need some tool to proceed from a fairly general kernel Φ, e.g. a radial basis function on $I\!R^d$, to certain functions φ_i and positive weights λ_i that allow to rewrite Φ in the form (8). This will be done by going back to the origin of Hilbert–Schmidt theory, i.e. eigenfunction expansions of kernels of compact integral operators.

Definition 5 *Let $\Phi : \Omega \times \Omega \to I\!R$ be a kernel. If the integral operator*

$$\mathcal{I}_\Phi(f) := \int_\Omega f(t)\Phi(t, \cdot)dt \tag{10}$$

maps $L_2(\Omega)$ into itself and is **c***ompact,* **i***njective,* **p***ositive, and* **s***elfadjoint, we say that Φ is a* CIPS *kernel on $L_2(\Omega)$.*

Theorem 6 *Any* **CIPS** *kernel on $L_2(\Omega)$ has an absolutely and uniformly convergent representation* (8) *with $I := I\!N$ and*

$$\lambda_1 \geq \lambda_2 \geq \ldots > 0 \text{ and } \lambda_i \to 0 \text{ for } i \to \infty$$

and a complete orthonormal system $\{\varphi_i\}_{i \in I\!N}$ in $L_2(\Omega)$ of eigenfunctions, i.e.

$$\mathcal{I}_\Phi(\varphi_i) = \lambda_i \varphi_i \text{ for all } i \in I\!N.$$

Proof: The existence of the eigenfunctions and the series representation is a consequence of standard ([12]) spectral theory of selfadjoint compact operators on $L_2(\Omega)$. Uniform convergence of the series follows from the theorem of Mercer, and we get (7). □

Definition 6 *A Hilbert–Schmidt kernel on Ω that has the properties asserted in Theorem 6 will be called a* positive *Hilbert–Schmidt kernel (***PHS***) on $L_2(\Omega)$.*

Note that positivity and injectivity of the integral operator means that

$$(f, g)_{\mathcal{I}_\Phi} := (\mathcal{I}_\Phi(f), g)_2 = (f, \mathcal{I}_\Phi(g))_2 \text{ for all } f, g \in L_2(\Omega)$$

is an inner product on $L_2(\Omega)$. The notion of positive definiteness of a kernel is different, and it does not seem easy to connect these properties. We further note that for PHS kernels we also have

$$(f, \mathcal{I}_\Phi(g))_\Phi = (f, g)_2 \text{ for all } f \in \mathcal{N}_\Phi,\ g \in L_2(\Omega), \tag{11}$$

and the native space (9) is embedded into $L_2(\Omega)$ as

$$\mathcal{N}_\Phi = \left\{ f \in L_2(\Omega) \ : \ \sum_{i \in I\!N} \frac{(f, \varphi_i)_2^2}{\lambda_i} < \infty \right\} \tag{12}$$

with the inner product taking the form

$$(f, g)_\Phi = \sum_{i \in I\!N} \frac{(f, \varphi_i)_2 (g, \varphi_i)_2}{\lambda_i} \text{ for all } f, g \in \mathcal{N}_\Phi. \tag{13}$$

Theorem 7 *The following are equivalent:*

1. *The kernel Φ is* **PHS** *in $L_2(\Omega)$.*
2. *The kernel Φ is reproducing on Ω with the above native space $\mathcal{N}_\Phi \subseteq L_2(\Omega)$ and a complete L_2–orthonormal system of functions φ_i such that (7) holds.*
3. *The kernel Φ is a* **CIPS** *kernel on $L_2(\Omega)$.*

Proof sketch: The implication $3 \Rightarrow 1$ is Theorem 6, while the implication $1 \Rightarrow 2$ follows from Theorem 4. If 2 holds, the integral operator is the limit of integral operators whose kernels are the finite partial sums of Φ, and thus is compact. Injectivity and positivity follow easily, because all λ_i are positive. □

Theorem 8 *If Φ is a reproducing kernel on Ω such that*

$$\begin{aligned}\int_\Omega \Phi(y,y)dy &< \infty\\ \int_\Omega\int_\Omega \Phi(x,y)^2 dxdy &< \infty\\ \int_\Omega \Phi(x,y)f(y)dy &= 0 \text{ for all } x\in\Omega \text{ implies } f=0 \text{ in } L_2(\Omega)\end{aligned}$$

then Φ is a **CIPS** *kernel on $L_2(\Omega)$.*

Proof sketch: The first additional hypothesis guarantees that the native space of Φ can be embedded into $L_2(\Omega)$. The second ensures compactness of the integral operator in $L_2(\Omega)$. Then spectral theory [12] allows to conclude the existence of an expansion (8) with L_2–orthogonal φ_i and rather general weights, but the reproduction property implies that the weights are nonnegative. The third additional hypothesis guarantees injectivity of the integral operator, positivity of all weights, and completeness of the system of orthogonal eigenfunctions. Details are in [16]. □

Note that injectivity of $\mathcal{I}_\Phi$ is essential here, but the nondegeneracy of the kernel and the separation property are not mentioned at all. Theorem 8 shows that very many kernels have a positive Hilbert–Schmidt form, and this motivates our concentration on those kernels in the remaining sections. We close this section by noting that we are still lacking useful conditions that allow to relate properties of Φ like positive definiteness or nondegeneracy to properties of $\mathcal{I}_\Phi$ like positivity or injectivity.

3 A Discontinuous Example

The techniques of the previous section allow to construct new kernels from expansions. These expansions may be based on a complete set of L_2–orthonormal functions, but they can also be quite general as in Definition 4 and Theorem 4. So far, all known kernels are at least continuous, but we can use the new technique to present a discontinuous case as an example. We modify an approach due to Fabien Hinault (private communication, 2000).

Let us mimic part of a Haar basis on $\mathbb{R}$ by taking scaled and shifted characteristic functions

$$H_k^j(x) := \chi_{[0,1)}\left(2^j x - k\right) = \chi_{[k2^{-j},(k+1)2^{-j})}(x) \text{ for all } k\in\mathbb{Z}, j\geq 0,\ x\in\mathbb{R}.$$

They have the properties

$$\begin{aligned} H_k^j(x) &= 1 \text{ iff } k = \lfloor 2^j x\rfloor & &\text{else } = 0,\\ H_k^j(x)H_k^j(y) &= 1 \text{ iff } k = \lfloor 2^j x\rfloor = \lfloor 2^j y\rfloor & &\text{else } = 0.\end{aligned}$$

With a summable sequence of positive weights ρ_j, $j\geq 0$ we define

$$\begin{aligned}\Phi(x,y) &:= \sum_{j=0}^\infty \rho_j \sum_{k=-\infty}^\infty H_k^j(x)H_k^j(y)\\ &= \sum_{\substack{j=0\\ \lfloor 2^j x\rfloor = \lfloor 2^j y\rfloor}}^\infty \rho_j\end{aligned}$$

for all $x, y \in I\!R$. Note now that $\lfloor 2^j x \rfloor = \lfloor 2^j y \rfloor$ for some $j \geq 0$ can hold only if x and y are of the same sign and do not differ by 1 or more. Moreover, the identity $\lfloor 2^j x \rfloor = \lfloor 2^j y \rfloor$ means that x and y coincide in their binary expansions in all of the pre–period digits and in the first j post–period digits. This means

$$\Phi(x, y) = \begin{cases} \sum_{j=0}^{m} \rho_j & x, y \text{ coincide in sign and all leading binary digits up to the } m\text{–th after the period} \\ 0 & \text{else} \end{cases}$$

and in particular

$$\Phi(x, x) = \sum_{j=0}^{\infty} \rho_j.$$

Thus the kernel is piecewise constant and has a finite evaluation scheme, if the sum over the ρ_j has a known value.

Theorem 9 *The kernel Φ is positive definite.*

Proof: In view of Theorems 4 and 5 we only have to show that the functions H_k^j separate points. Take a set $X = \{x_1, \ldots, x_M\} \subseteq I\!R$, pick an arbitrary index $s \in \{1, \ldots, M\}$ and a $j > 0$ such that

$$|x_r - x_s| > 2^{-j} \text{ for all } r \neq s,\ 1 \leq r \leq M.$$

This implies $\lfloor 2^j x_s \rfloor \neq \lfloor 2^j x_r \rfloor$ for all $r \neq s$. Then we pick $k = \lfloor 2^j x_s \rfloor$ and find that $H_k^j(x_s) = 1$ while $H_k^j(x_r) = 0$ for all $r \neq s$, and we get the separation. □

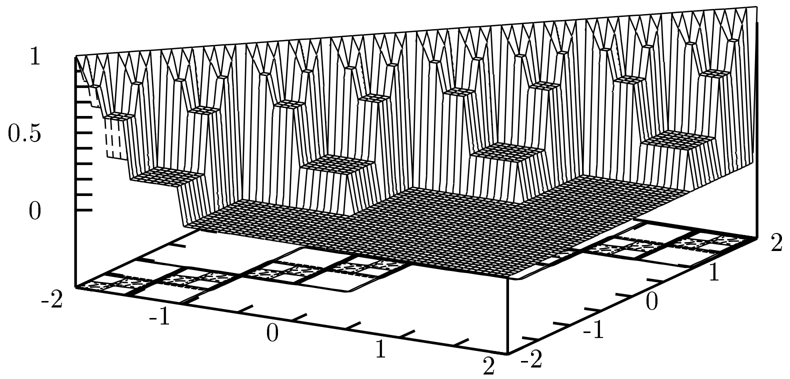

Figure 1: The case $\rho_j = 2^{-j-1}$

We remark that one can construct plenty of other examples using other bases, in particular wavelet bases. We hope to find time to follow the open road towards "refinable kernels" elsewhere.

4 Native Space and Range

From here on we always assume a positive definite kernel Φ that is a positive Hilbert–Schmidt kernel on $L_2(\Omega)$, and in particular we consider the native space (12) and the inner product (13) there. Note that the action of the integral operator $\mathcal{I}_\Phi$ of (10) on a function f with expansion coefficients $(f, \varphi_i)_2$ just consists of a multiplication of the coefficients by λ_i.

The range of the integral operator $\mathcal{I}_\Phi$ of (10) then is

$$\mathcal{R}_\Phi := \left\{ f \in L_2(\Omega) \ : \ \sum_{i\in I\!N} \frac{(f,\varphi_i)_2^2}{\lambda_i^2} < \infty \right\},$$

and it is the native space of the convolution kernel

$$\begin{array}{rcl} (\Phi * \Phi)(x,y) & := & \displaystyle\int_\Omega \Phi(x,t)\Phi(y,t)dt \\ & = & \displaystyle\sum_{i=1}^\infty \lambda_i^2 \varphi_i(x)\varphi_i(y) \end{array}$$

Consequently we have the inclusions

$$\mathcal{R}_\Phi \quad = \quad \mathcal{N}_{\Phi*\Phi} \quad \subseteq \quad \mathcal{N}_\Phi \quad \subseteq \quad L_2(\Omega).$$

The subspace $\mathcal{R}_\Phi$ of the native space $\mathcal{N}_\Phi$ is of quite some importance. For completeness, we add a result from [16] that generalizes [14]:

Theorem 10 *The convergence order of interpolants to functions from $\mathcal{R}_\Phi$ is twice the convergence order of functions from the native space $\mathcal{N}_\Phi$.*

Proof: The interpolant $s_{f,X,\Phi} \in S_{X,\Phi}$ of (3) to a function f from $\mathcal{N}_\Phi$ in data locations $X = \{x_1,\ldots,x_M\}$ with *fill distance*

$$h_X := \sup_{x\in\Omega} \min_{1\le j\le M} \|x - x_j\|_2$$

has a standard [13] error bound

$$\|f - s_{f,X,\Phi}\|_2^2 \le F_\Phi(h_X)\|f - s_{f,X,\Phi}\|_{\mathcal{N}_\Phi}^2 \le F_\Phi(h_X)\|f\|_{\mathcal{N}_\Phi}^2 \tag{14}$$

for all $x \in \Omega$ with a certain function F_Φ that depends on the smoothness of Φ. We assert that for $f = \mathcal{I}_\Phi(g) \in \mathcal{R}_\Phi$ there is an improved bound

$$\|f - s_{f,X,\Phi}\|_2^2 \le F_\Phi(h_X)^2\|f\|_{\Phi*\Phi}^2 = F_\Phi(h_X)^2\|g\|_2^2.$$

To this end, we use the standard [15] orthogonality relation

$$(f - s_{f,X,\Phi}, s_{f,X,\Phi})_{\mathcal{N}_\Phi} = 0$$

and the property (11) of $\mathcal{I}_\Phi$ for $f = \mathcal{I}_\Phi(g) \in \mathcal{R}_\Phi = \mathcal{N}_{\Phi*\Phi}$ to find

$$\begin{array}{rcl} \|f - s_{f,X,\Phi}\|_{\mathcal{N}_\Phi}^2 & = & (f - s_{f,X,\Phi}, f)_{\mathcal{N}_\Phi} \\ & = & (f - s_{f,X,\Phi}, \mathcal{I}_\Phi(g))_{\mathcal{N}_\Phi} \\ & = & (f - s_{f,X,\Phi}, g)_2 \\ & \le & \|f - s_{f,X,\Phi}\|_2\|g\|_2 \\ & \le & \sqrt{F_\Phi(h_X)}\|f - s_{f,X,\Phi}\|_{\mathcal{N}_\Phi}\|g\|_2 \\ \|f - s_{f,X,\Phi}\|_{\mathcal{N}_\Phi} & = & \sqrt{F_\Phi(h_X)}\|g\|_2 \end{array}$$

and we can plug this into the standard error bound (14) to arrive at

$$\begin{array}{rcl} \|f - s_{f,X,\Phi}\|_2^2 & \le & F_\Phi(h_X)\|f - s_{f,X,\Phi}\|_{\mathcal{N}_\Phi}^2 \\ & \le & F_\Phi(h_X)^2\|g\|_2^2 \end{array}$$

with

$$\|g\|_2^2 = (g,g)_2 = (\mathcal{I}_\Phi(g), g)_{\mathcal{N}_\Phi} = (\mathcal{I}_\Phi(g), \mathcal{I}_\Phi(g))_{\Phi*\Phi} = \|f\|_{\Phi*\Phi}^2.$$

□

If we ask somewhat more than (7), i.e.

$$\sum_{i\in I} \sqrt{\lambda_i}\varphi_i^2(x) < \infty \tag{15}$$

we can define the *convolution square–root* of Φ by the kernel

$$\sqrt{\Phi}(x,y) := \sum_{i=1}^{\infty} \sqrt{\lambda_i}\varphi_i(x)\varphi_i(y)$$

and get

$$\mathcal{R}_\Phi \quad = \quad \mathcal{N}_{\Phi*\Phi} \quad \subseteq \quad \mathcal{R}_{\sqrt{\Phi}} \quad = \quad \mathcal{N}_\Phi \quad \subseteq \quad \mathcal{N}_{\sqrt{\Phi}} \quad \subseteq \quad L_2(\Omega).$$

5 n–Widths

From now on we let Φ be a positive Hilbert–Schmidt kernel on $L_2(\Omega)$ and assume (15) to play safe. We make use of the fact that we have integral operators related to $\sqrt{\Phi}(x,y)$ or $\Phi(x,y)$ that map $L_2(\Omega)$ into $\mathcal{N}_\Phi$ or $\mathcal{R}_\Phi$. This opens the road for applications of the theory of n–widths [11]. For the convenience of the reader, we will review that part that is of interest for us. For a subset A of a Hilbert space H, the Kolmogorov n–width is defined by

$$d_n(A;H) := \inf_{V_n} \sup_{f\in H} \inf_{s\in V_n} \|f-s\|_H.$$

Here, the outer infimum is taken over all n-dimensional subspaces V_n of H. An n-dimensional space V_n^* is called optimal if

$$E(A;V_n^*) := \sup_{f\in A} \inf_{s\in V_n^*} \|f-s\|_H = d_n(A;H).$$

In our case, the Hilbert space H will always be $H = L_2(\Omega)$ and the set A will essentially be either $\mathcal{N}_\Phi$ or $\mathcal{R}_\Phi$. Actually, to avoid problems with scaling we will take A rather to be the unit ball in that space, i.e. $A = S(\mathcal{N}_\Phi)$ or $A = S(\mathcal{R}_\Phi)$, where we used the general notation $S(H) = \{h \in H : \|h\|_2 \leq 1\}$. This perfectly fits into the theory of n–width of compact operators, where A is the image of the unit ball of the linear space H under a continuous mapping T. In our case, the mapping is given by $\mathcal{I}_{\sqrt{\Phi}}$ and $\mathcal{I}_\Phi$, respectively.

Lemma 1 *The unit ball of the native space $\mathcal{N}_\Phi$ is the image of the unit ball of $L_2(\Omega)$ under the operator $\mathcal{I}_{\sqrt{\Phi}}$, i.e. $S(\mathcal{N}_\Phi) = \mathcal{I}_{\sqrt{\Phi}}(S(L_2(\Omega)))$. Similarly, we have for $\mathcal{R}_\Phi$ that $S(\mathcal{R}_\Phi) = \mathcal{I}_\Phi(S(L_2(\Omega)))$.*

Proof: If $f = \mathcal{I}_{\sqrt{\Phi}}v$ with $v \in S(L_2(\Omega))$, then, by definition of the native space norm, $\|f\|_\Phi = \|v\|_2$. The same holds in the second case. □

The results of Pinkus' book [11], in particular, Corollary 2.6 of Chapter IV yield:

Theorem 11 *Let Φ be a positive Hilbert–Schmidt kernel on $L_2(\Omega)$ with (15). Then, the n–widths for the unit ball in $\mathcal{N}_\Phi$ and $\mathcal{R}_\Phi$ are given by*

$$\begin{aligned} d_n(S(\mathcal{N}_\Phi);L_2(\Omega)) &= \sqrt{\lambda_{n+1}}, \\ d_n(S(\mathcal{R}_\Phi);L_2(\Omega)) &= \lambda_{n+1}, \end{aligned}$$

respectively. In both cases, the subspace

$$V_n^* := \text{ span } \{\varphi_1,\ldots,\varphi_n\}$$

is optimal. The associated optimal data functionals have the form $\rho_k(f) := (f,\varphi_k)_2$ for all $f \in L_2(\Omega)$.

As said before, the proof can be found in Pinkus' book, but it is also not too difficult. For example, to see that V_n^* is optimal for $S(\mathcal{N}_\Phi)$ we simply use $f_n = \sum_{j=1}^n (f,\varphi_j)_2\varphi_j \in V_n^*$ as the approximant to $f \in S(\mathcal{N}_\Phi)$ to get

$$\|f-f_n\|_2^2 = \sum_{j=n+1}^{\infty} (f,\varphi_j)_2^2 = \sum_{j=n+1}^{\infty} \lambda_j \frac{(f,\varphi_j)_2^2}{\lambda_j} \leq \lambda_{n+1} \sum_{j=n+1}^{\infty} \frac{(f,\varphi_j)_2^2}{\lambda_j} \leq \sqrt{\lambda_{n+1}},$$

since $\|f\|_\Phi = \sum \frac{(f,\varphi_j)_2^2}{\lambda_j} \leq 1$. □

The good news here is that we have found best rates for n–term approximation. The bad news is that for standard radial cases neither the φ_i nor the λ_i are known. Furthermore, the optimal functionals are not easily accessible numerically. Thus the next section tries to compare the optimal n–width errors with the behaviour of standard interpolation in n data locations or with simple approximation schemes.

6 Quasi–optimal Processes

Here, we shall look at approximation or interpolation schemes to see whether they realize the optimal behaviour outlined in Theorem 11 or not. Since the eigenfunctions are not accessible in many cases, and since the inner products with eigenfunctions are not practically relevant as data functionals, we have to be satisfied with quasi-optimal subspaces instead of optimal subspaces.

Definition 7 *An n dimensional subspace $V_n \subseteq H$ is called quasi–optimal for $A \subseteq H$ if there exists a constant $C > 0$, independent of n, such that*

$$E(A; V_n) \leq C d_n(A; H).$$

Since $E(A; V_n) \geq d_n(A; H)$ is always satisfied, both quantities are equivalent, which we will also denote by $E(A; V_n) \sim d_n(A; H)$.

We now look at some special cases from the literature, and we start with approximation on the sphere $S^{d-1} = \{x \in I\!R^d : \|x\|_2 = 1\}$. Here, things are generally presented upside down, i.e. one starts with a family of orthonormal functions, namely spherical harmonics and defines the kernel Φ by its expanding series so that the eigenvalues of the corresponding integral operator are the Fourier coefficients of the kernel. To be more precise, let $\{Y_{\ell,k} : 1 \leq k \leq N(d,\ell)\}$ denote the usual orthonormal basis for the space of spherical harmonics of degree ℓ (cf. [10]), where

$$N(d, 0) = 1, \quad \text{and} \quad N(d,\ell) = \frac{2\ell + d - 2}{\ell}\binom{\ell + d - 3}{\ell - 1}, \ \ell > 0.$$

Then the kernel has an expansion of the form

$$\Phi(p, q) = \sum_{\ell=0}^{\infty} \sum_{k=1}^{N(d,\ell)} a_{\ell,k} Y_{\ell,k}(p) Y_{\ell,k}(q). \tag{15}$$

For simplicity, we will assume that the kernel is radial or zonal, which is equivalent to the fact that for a fixed ℓ all coefficients $a_{\ell,k}$, $1 \leq k \leq N(d,\ell)$, are the same, i.e. $a_\ell := a_{\ell,k}$, $1 \leq k \leq N(d,\ell)$.

Under this assumption, it is actually more natural to look at the space of spherical harmonics up to order ℓ,

$$V_\ell := \operatorname{span}\{Y_{\lambda,k} : 0 \leq \lambda \leq \ell, 1 \leq k \leq N(d,\lambda)\},$$

which is the restriction of the space of d-variate polynomials of degree at most ℓ to the sphere and has dimension $\dim V_\ell = N(d+1,\ell)$. The n–width theory gives

Corollary 1 *If the coefficients $\alpha_\ell = \alpha_{\ell,k}$, $1 \leq k \leq N(d,\ell)$, of the kernel (15) form a sufficiently fast decaying, nonincreasing, and positive sequence, then*

$$d_n(S(\mathcal{N}_\Phi), L_2(S^{d-1})) = \sqrt{a_\ell},$$

for $N(d+1,\ell) \leq n < N(d+1,\ell+1)$.

This is the result to which we have to compare the known estimates for interpolation by positive definite kernels. In the latter context it is usual to assume that

$$N(d,\ell) a_\ell \leq C(1+\ell)^{-\alpha}$$

which is, since $N(d,\ell)$ grows like $\mathcal{O}(\ell^{d-2})$, equivalent to $a_\ell = \mathcal{O}(\ell^{-\alpha-d+2})$. The reason for looking at $N(d,\ell)a_\ell$ rather than a_ℓ is that this number appears naturally for "radial" kernels, since the addition theorem (cf. [10]) yields

$$\Phi(p, q) = \sum_{\ell=0}^{\infty} \frac{N(d,\ell) a_\ell}{\omega_{d-1}} P_\ell(p \cdot q),$$

where ω_{d-1} denotes the surface area of S^{d-1} and P_ℓ is the Legendre polynomial of degree ℓ in d dimensions, normalized by $P_\ell(1) = 1$.

In case of interpolation by positive definite kernels it is usual to measure the approximation orders in terms of the so-called fill distance, which is in this context $h_X := \sup_{x \in S^{d-1}} \min_{x_j \in X} \operatorname{dist}(x, x_j)$. Here, dist is the usual spherical distance.

The following result comes from Dyn/Narcowich/Ward [4], Jetter/Stöckler/Ward [6], and Morton/Neamtu [9].

Theorem 12 *Suppose Φ is a radial positive definite kernel on the sphere with $a_\ell = \mathcal{O}(\ell^{-\alpha})$, $\ell \to \infty$, with $\alpha > d$. Then, the interpolation error can be bounded by*

$$\|f - s_{f,X}\|_\infty \le C h_X^{\frac{\alpha-1}{2}} \|f\|_\Phi.$$

The L_∞-error bound leads immediately to an L_2-error bound, which we now want to compare with the results from n–width theory. To acchieve this, we have to relate h_X to ℓ, since by Corollary 1 the n–width is rather related to ℓ than to n in this situation,

$$d_n(S(\mathcal{N}_\Phi); L_2(S^{d-1})) = \mathcal{O}(\ell^{-\frac{\alpha-d-2}{2}}).$$

This is hopeless in the general case, but the situation changes in case of quasi-uniform data sets. A set $X \subseteq S^{d-1}$ of n points is said to be quasi-uniform if $h_X^{d-1} \sim 1/n$. Since we also know that $n \sim N(d+1,\ell) \sim \ell^{d-1}$ we can conclude

$$\|f - s_{f,X}\|_2 = \mathcal{O}(\ell^{-\frac{\alpha-1}{2}}).$$

Corollary 2 *Interpolation of function values in quasi-uniform data locations by positive definite "radial" kernels on the sphere may fail to be quasi–optimal by order at most $\frac{d-1}{2}$ if the kernel has eigenvalues with algebraic decay.*

Our formulation of the corollary just poses an upper bound on the deviation from quasi–optimality, but we think that we actually have a quasi–optimal approximation scheme. The reason for our optimistic point of view is the following. We gained the L_2 approximation error simply by integrating the L_∞-error. In the light of the $\mathbb{R}^d$ theory, this seems to be too naive. In the $\mathbb{R}^d$ case it is, in a similar situation, possible to gain an additional $d/2$ in the order by using a localization trick, which dates back to Duchon's initial work on thin-plate splines (cf. [2, 3]). This trick should also work in the sphere setting, but so far nobody has ever tried it.

Note that in the just described situation the native space is actually the Sobolev space $H^s(S^{d-1})$ with $s = \frac{\alpha+d}{2} - 1$.

For Euclidean space $\mathbb{R}^d$ and bounded domains Ω therein, we usually do not know the orthogonal Hilbert–Schmidt expansions in $L_2(\Omega)$. Thus we cannot assess the optimality of the known error bounds. The state–of–the–art in results on optimality of rates of approximation provided by interpolation is in [17, 20]. Instead of optimality results for approximations, we here get upper bounds on the decay of the unknown eigenvalues. Curiously enough, this means that approximation theory provides results on the spectrum of integral operators.

On $\mathbb{R}^d$ we make the following assumptions:

- the kernel $\Phi(x,y) = \phi(x-y)$ is symmetric and Fourier–transformable,
- we consider interpolation by translates of ϕ on n asymptotically quasi-uniform data locations in a bounded domain $\Omega \subseteq \mathbb{R}^d$, which has a sufficiently smooth boundary.

Let us look at the case of limited smoothness (e.g. [13]) first. For

$$\hat{\phi}(\omega) \sim (1 + \|\omega\|_2)^{-d-\beta},\ \|\omega\|_2 \to \infty, \tag{16}$$

there is an error bound

$$\|f - s_{f,X}\|_\infty \le C h^{\beta/2} \|f\|_\Phi$$

This error bound can be improved by Duchon's localization trick as mentionened earlier (see for example [19]) to

$$\|f - s_{f,X}\|_2 \le C h^{(\beta+d)/2} \|f\|_\Phi,$$

provided that the boundary of Ω is sufficiently smooth.

In case of quasi-uniform data, which now becomes $h_{X,\Omega}^d \sim 1/n$, the latter means in terms of n,

$$\|f - s_{f,X}\|_2 \le C n^{-(\beta+d)/2d} \|f\|_\Phi.$$

The error of the optimal process must be asymptotically smaller, and this implies

Theorem 13 *The eigenvalues of the Hilbert–Schmidt operator $\mathcal{I}_\Phi$ with kernel Φ on $L_2(\Omega)$ and Fourier transform satisfying* (16) *for a bounded domain $\Omega \subseteq I\!R^d$ satisfy*

$$\lambda_{n+1} \leq C n^{-(\beta+d)/d}$$

for $n \to \infty$. □

Again, as in the case of the sphere, the native space is a Sobolev space $H^s(\Omega)$, $s = (\beta + d)/2$. For Sobolev spaces, the optimal n–widths are known (Jerome 1970 [5]):

$$d_n(S(H^s(\Omega); L_2(\Omega)) = \sqrt{\lambda_{n+1}} = \mathcal{O}(n^{-s/d}) \text{ for } n \to \infty$$

and we can compare with the interpolation error bounds for $H^s(\Omega)$ with $\Omega \subseteq I\!R^d$. They have the form (14) with $s = (\beta + d)/2 > 0$, and we get

Theorem 14 *Interpolation in quasi-uniform locations by translates of reproducing kernels that generate Sobolev spaces is quasi–optimal.*

Since Sobolev kernels and Wendland functions [18, 19] reproduce spaces that are norm–equivalent to Sobolev spaces, we have

Corollary 3 *Interpolation in asymptotically regular data locations by translates of Sobolev kernels or Wendland functions is quasi–optimal.* □

Generalizations to other radial basis functions are not known, but would be welcome.

The case of unlimited smoothness occurs for inverse multiquadrics and Gaussians, and it leads to Fourier transforms with a decay like

$$\hat{\phi}(\omega) \leq C \exp(-c\|\omega\|_2), \ \|\omega\|_2 \to \infty. \tag{17}$$

Then there is an error bound [13]

$$\|f - s_{f,n}\|_\infty \leq C \exp(-c/h)\|f\|_\Phi \leq C \exp(-cn^{1/d})\|f\|_\Phi.$$

Theorem 15 *For a kernel Φ with exponential decay* (17) *of its Fourier transform, the eigenvalues of the integral operator $\mathcal{I}_\Phi$ in $L_2(\Omega)$ for a bounded domain $\Omega \subseteq I\!R^d$ satisfy*

$$\lambda_{n+1} \leq C \exp(-cn^{1/d})$$

for $n \to \infty$. □

References

[1] M. Atteia. *Hilbertian kernels and spline functions.* North-Holland, Amsterdam, 1992. Studies in Computational Mathematics (4).

[2] J. Duchon. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *Rev. Française Automat. Informat. Rech. Opér. Anal. Numer.*, 10:5–12, 1976.

[3] J. Duchon. Sur l'erreur d'interpolation des fonctions de plusieurs variables pas les D^m–splines. *Rev. Française Automat. Informat. Rech. Opér. Anal. Numer.*, 12(4):325–334, 1978.

[4] N. Dyn, F.J. Narcowich, and J.D. Ward. Variational principles and Sobolev–type estimates for generalized interpolation on a riemannian manifold. *Constructive Approximation*, 15(2):174–208, 1999.

[5] J.W. Jerome. On n-widths in Sobolev spaces and applications to elliptic boundary value problems. *Journal of Mathematical Analysis and Applications*, 29:201–215, 1970.

[6] K. Jetter, J. Stöckler, and J.D. Ward. Error estimates for scattered data interpolation on spheres. *Mathematics of Computation*, 68:733–747, 1999.

[7] L.-T. Luh. *Characterizations of native spaces.* Dissertation, Universität Göttingen, 1998.

[8] H. Meschkowski. *Hilbertsche Räume mit Kernfunktion.* Springer, Berlin, 1962.

[9] Tanya M. Morton and M. Neamtu. Error bounds for solving pseudodifferential equations on spheres by collocation with zonal kernels. Preprint Dept. of Math. Vanderbilt University, Nashville TN 37240, 2000.

[10] C. Müller. *Spherical Harmonics.* Springer, Berlin, 1966.

[11] A. Pinkus. *n–widths in Approximation Theory.* Springer, 1985.

[12] D. Porter and D.G. D.G. Stirling. *Integral equations: a practical treatment from spectral theory to applications.* Cambridge Texts in Applied Mathematics. Cambridge University Press, 1990.

[13] R. Schaback. Error estimates and condition numbers for radial basis function interpolation. *Advances in Computational Mathematics*, 3:251–264, 1995.

[14] R. Schaback. Improved error bounds for scattered data interpolation by radial basis functions. *Mathematics of Computation*, 68:201–216, 1999.

[15] R. Schaback. Native Hilbert spaces for radial basis functions i. In M.D. Buhmann, D. H. Mache, M. Felten, and M.W. Müller, editors, *New Developments in Approximation Theory*, number 132 in International Series of Numerical Mathematics, pages 255–282. Birkhäuser Verlag, 1999.

[16] R. Schaback. A unified theory of radial basis functions. *J. Comp. Appl. Math.*, pages 165–177, 2000.

[17] R. Schaback and H. Wendland. Inverse and saturation theorems for radial basis function interpolation. to appear in Math. Comp.

[18] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4:389–396, 1995.

[19] H. Wendland. Sobolev–type error estimates for interpolation by radial basis functions. In A. LeMéhauté, C. Rabut, and L.L. Schumaker, editors, *Surface Fitting and Multiresolution Methods*, pages 337–344. Vanderbilt University Press, Nashville, TN, 1997.

[20] H. Wendland. Optimal approximation orders in L_p for radial basis functions. *East J. Approx.*, 6:87–102, 2000.