Near–optimal Data–independent Point Locations for Radial Basis Function Interpolation*

Stefano De Marchi
University of Verona
S.da Le Grazie, 15 - Ca' Vignal 2
I-37137 - Verona, Italy

Robert Schaback Holger Wendland
Universität Göttingen
Lotzestrasse 16-18, D-37083 Göttingen, Germany

Abstract

The goal of this paper is to construct data–independent optimal point sets for interpolation by radial basis functions. The interpolation points are chosen to be uniformly good for all functions from the associated native Hilbert space. To this end we collect various results on the power function, which we use to show that good interpolation points are always uniformly distributed in a certain sense. We also prove convergence of two different greedy algorithms for the construction of near-optimal sets which lead to stable interpolation. Finally, we provide several examples.

Keywords: Radial basis function interpolation, optimal points, greedy algorithms.
2000 Mathematics Subject Classification: 41A05,41063, 41065, 65D05, 65D15.

1 Introduction

Interpolation by radial basis functions is a state-of-the-art and well-established method for reconstructing multivariate functions from scattered data.

*This work has been done with the support of the Vigoni CRUI-DAAD programme, for the years 2001-02, between the Universities of Verona and Göttingen.

The general concept can be described as follows. Suppose a set $X = \{x_1, \ldots, x_N\} \subseteq \Omega \subseteq \mathbb{R}^d$ of N distinct points coming from a compact subset Ω of $\mathbb{R}^d$ is given. These points will be called *data sites*. Suppose further that N *data values* $f_1, \ldots, f_N$ should be interpolated at the data sites. The easiest approach is to fix a symmetric kernel function $\Phi : \Omega \times \Omega \to \mathbb{R}$ and to form an interpolant

$$s_{f,X} = \sum_{j=1}^{N} \alpha_j \Phi(\cdot, x_j). \tag{1}$$

Obviously, the coefficients $\{\alpha_j\}$ are determined by the interpolation conditions $s_{f,X}(x_j) = f_j$, $1 \leq j \leq N$. They are unique, if the interpolation matrix $A_{\Phi,X} := (\Phi(x_i, x_j))_{1 \leq i,j \leq N}$ is invertible. If the matrices $A_{\Phi,X}$ are even positive definite for all possible point sets $X \subseteq \Omega$, regardless of the number of points, the function Φ is called a *positive definite kernel*. It is often *radial* in the sense $\Phi(x, y) = \phi(\|x - y\|_2)$ and therefore defined on $\mathbb{R}^d \times \mathbb{R}^d$.

In this paper we will confine ourselves to the case of positive definiteness, neglecting conditionally positive definite functions. This is not a serious restriction, because every conditionally positive definite kernel has an associated, normalized positive definite kernel (see for example [7] and [1]). Moreover, the experienced reader will notice without problems that many of our proofs also work in the conditionally positive definite case.

The generality of this interpolation method allows to deal with any set of data sites. Thus it leads immediately to the problem of finding good or even optimal point sets for the reconstruction process. But despite of the natural character of this problem, it was only addressed in [2, 3, 4]. While the thesis [2] concentrated on numerical results, the other two publications tried to investigate the problem theoretically, at least in certain special cases. In particular, [3] shows how difficult it is to find truly optimal interpolation points. The reason is simple: one has to minimize a highly nonlinear function of Nd unknowns. Hence, we will use another approach here, which is based on power function estimates and geometric arguments.

The paper is organized as follows. In the next section we collect all necessary material on positive definite kernels, power functions, and native Hilbert spaces. In the third section, we introduce our concept of uniformly distributed points for a general region $\Omega \subseteq \mathbb{R}^d$ and show that the geometry of the data set is determined by a global interpolation error. In the next two section we introduce two different greedy methods for finding good points and prove their convergence. The final section deals with numerical examples.

2 Power Functions and Native Hilbert Spaces

Throughout this paper we assume $\Omega \subseteq \mathbb{R}^d$ to be a compact subset of $\mathbb{R}^d$, which satisfies an interior cone condition. On Ω we have a positive definite kernel $\Phi : \Omega \times \Omega \to \mathbb{R}$. The interpolant (1) can also be written in the following form. Let $u_j \in V_X = \text{span}\{\Phi(\cdot, x) : x \in X\}$ denote the cardinal functions, i.e. u_j satisfies $u_j(x_k) = \delta_{j,k}$. Then the interpolant takes the form

$$s_{f,X} = \sum_{j=1}^{N} f(x_j) u_j. \tag{2}$$

Moreover, the kernel Φ defines on the space

$$V_\Omega = \text{span}\{\Phi(\cdot, x) : x \in \Omega\}$$

an inner product via

$$\left(\sum_{j=1}^{N} \alpha_j \Phi(\cdot, x_j), \sum_{k=1}^{M} \beta_k \Phi(\cdot, y_k) \right)_\Phi := \sum_{j=1}^{N} \sum_{k=1}^{M} \alpha_j \beta_k \Phi(x_j, y_k).$$

It can easily be seen that Φ is the reproducing kernel of V_Ω with respect to this inner product, i.e. every $f \in V_\Omega$ can be represented by

$$f(x) = (f, \Phi(\cdot, x))_\Phi.$$

The closure of V_Ω leads to a Hilbert space with reproducing kernel Φ, since the above reproduction formula stays valid in the closure, showing how the abstract elements in the closure can be interpreted as functions. This Hilbert space is often called the *native Hilbert space* to Φ and we will denote it by $\mathcal{N}_\Phi(\Omega)$.

Obviously, an inclusion of the form $\Omega_1 \subseteq \Omega_2$ leads to a continuous embedding $\mathcal{N}_\Phi(\Omega_2) \subseteq \mathcal{N}_\Phi(\Omega_1)$ by restriction, whenever the kernel Φ is defined on the bigger set. In particular, if Φ is a translation invariant or even radial kernel, we have $\mathcal{N}_\Phi(\mathbb{R}^d) \subseteq \mathcal{N}_\Phi(\Omega)$ for every $\Omega \subseteq \mathbb{R}^d$.

For a function $f \in \mathcal{N}_\Phi(\Omega)$ we can express the interpolation error using the cardinal representation (2) and the reproducing kernel property by

$$f(x) - s_{f,X}(x) = \left(f, \Phi(\cdot, x) - \sum_{j=1}^{N} u_j(x) \Phi(\cdot, x_j) \right)_\Phi,$$

so that an application of the Cauchy-Schwarz inequality immediately leads to

$$|f(x) - s_{f,X}(x)| \leq P_{\Phi,X}(x)\|f\|_\Phi \tag{3}$$

with the *power function* $P_{\Phi,X}$ taking the explicit form

$$\begin{aligned} P^2_{\Phi,X}(x) &:= \left\| \Phi(\cdot,x) - \sum_{j=1}^N u_j(x)\Phi(\cdot,x_j) \right\|^2_\Phi \\ &= \Phi(x,x) - 2\sum_{j=1}^N u_j(x)\Phi(x,x_j) \\ &\quad + \sum_{j,k=1}^N u_j(x)u_k(x)\Phi(x_j,x_k). \end{aligned} \tag{4}$$

In other words, the power function is nothing but the norm of the pointwise error functional, and it can numerically be evaluated using the Lagrange basis.

Typically, error estimates lead to the problem of bounding the power function in terms of the *fill distance*,

$$h_{X,\Omega} = \sup_{x\in\Omega} \min_{x_j\in X} \|x - x_j\|_2.$$

This is, in general, done by considering the square of the power function for a fixed $x \in \Omega$ as a quadratic form of the coefficient vector $(u_1(x), \ldots, u_N(x))$. One can show that this vector minimizes the quadratic form over all vectors $u \in \mathbb{R}^N$. Then a suitable vector $(\widetilde{u}_1(x), \ldots, \widetilde{u}_N(x))$ is constructed which allows an easier way of handling the quadratic form and yields the error bound in terms of the fill distance, because it provides an upper bound on the square of the power function. We will not discuss the details here but refer the reader to [9]. Instead, we remark that this minimization property has another consequence. If $X \subseteq Y$ are two point sets, then the associated power functions must necessarily satisfy

$$P_{\Phi,X}(x) \geq P_{\Phi,Y}(x), \qquad x \in \Omega,$$

because the Lagrange basis for X, extended by some zero functions, is admissible for the minimization of the quadratic form with respect to Y. Note that the above inequality holds pointwise and everywhere in Ω. It will be an important ingredient for the greedy methods to be described later.

To provide the reader with examples, we take a shift-invariant kernel $\Phi(x,y) = \phi(x-y)$, where $\phi : \mathbb{R}^d \to \mathbb{R}$ is integrable and thus has a Fourier transform $\widehat{\phi}$ defined by

$$\widehat{\phi}(\omega) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \phi(x) e^{-ix^T\omega} dx.$$

In this case, the native space norm of f becomes

$$\|f\|_\Phi^2 = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{|\widehat{f}(\omega)|^2}{\widehat{\Phi}(\omega)} d\omega.$$

Hence, if $\widehat{\phi}$ satisfies

$$c_\phi (1 + \|\omega\|_2^2)^{-\beta} \le \widehat{\phi}(\omega) \le C_\Phi (1 + \|\omega\|_2^2)^{-\beta} \tag{5}$$

with $\beta > d/2$ and two constants $C_\phi \ge c_\phi > 0$, the global native Hilbert space $\mathcal{N}_\Phi(\mathbb{R}^d)$ is norm–equivalent to the Sobolev space $W_2^\beta(\mathbb{R}^d)$. Moreover, the interpolation error can be bounded by

$$\|f - s_{f,X}\|_{L_\infty(\Omega)} \le C h_{X,\Omega}^{\beta - d/2} \|f\|_{W_2^\beta(\mathbb{R}^d)}. \tag{6}$$

In case of infinitely often differentiable basis functions ϕ, such as Gaussians $\phi(x) = e^{-\alpha\|x\|_2}$, $\alpha > 0$, or inverse multiquadrics $\phi(x) = 1/\sqrt{\|x\|_2^2 + 1}$ the error even decays exponentially, but only for functions from a rather small native space.

3 Near–optimally Distributed Data Sites

Throughout this section we shall assume that our kernel Φ is translation invariant, integrable, and has a Fourier transform satisfying (5).

Our first goal is to show that a data set which allows good approximation for all functions from the native space cannot have a large region in Ω without centers. In other words the fill distance $h_{X,\Omega}$ must necessarily be sufficiently small.

Theorem 3.1 *Let Ω be the closure of an open and bounded region in $\mathbb{R}^d$ satisfying an interior cone condition. Suppose that the kernel Φ is translation invariant and its generating function ϕ is integrable with a Fourier transform satisfying (5) with $\beta > d/2$. Then for every $\alpha > \beta$ there exists a constant*

$M_\alpha > 0$ with the following property. If $\epsilon > 0$ and $X = \{x_1, \ldots, x_N\} \subseteq \Omega$ are given such that

$$\|f - s_{f,X}\|_{L_\infty(\Omega)} \leq \epsilon \|f\|_\Phi, \qquad \text{for all } f \in W_2^\beta(\mathbb{R}^d), \tag{7}$$

then the fill distance of X satisfies

$$h_{X,\Omega} \leq M_\alpha \epsilon^{\frac{1}{\alpha - d/2}}.$$

Proof: Without restriction, we can assume that Ω contains a sufficiently small ball. We fix an integer k such that $2k > \beta + d/2$. Next, we choose a function $\psi \in C_0^{2k}(\mathbb{R}^d)$ from [8] having support in the open unit ball and being bounded by $\|\psi\|_{L_\infty(\mathbb{R}^d)} = |\psi(0)| = 1$. Elementary calculus gives

$$\widehat{\Delta^\ell \psi}(\omega) = \sum_{|\alpha|=\ell} \frac{\ell!}{\alpha!} \widehat{D^\alpha \psi}(\omega) = \widehat{\psi}(\omega) \sum_{|\alpha|=\ell} \frac{\ell!}{\alpha!} (i\omega)^{2\alpha} = (-1)^\ell \|\omega\|_2^{2\ell} \widehat{\psi}(\omega)$$

for all $0 \leq \ell \leq k$, where we have used the standard notation for partial derivatives and the iterated Laplacian. Hence, if we define the constant

$$K_\psi := \left\| \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^\ell \widehat{\Delta^\ell \psi} \right\|_{L_\infty(\mathbb{R}^d)}$$

we get

$$\left| (1 + \|\omega\|_2^2)^k \widehat{\psi}(\omega) \right| = \left| \sum_{\ell=0}^{k} \binom{k}{\ell} \|\omega\|_2^{2\ell} \widehat{\psi}(\omega) \right| = \left| \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^\ell \widehat{\Delta^\ell \psi}(\omega) \right| \leq K_\psi,$$

which shows for every $s \leq k$ that

$$|\widehat{\psi}(\omega)| \leq K_\psi (1 + \|\omega\|_2^2)^{-k} \leq K_\psi (1 + \|\omega\|_2^2)^{-s}.$$

Moreover, if we scale ψ by setting $\psi_h := \psi(\cdot/h)$, we find

$$\begin{aligned} |\widehat{\psi_h}(\omega)| &= h^d |\widehat{\psi}(h\omega)| \\ &\leq K_\psi h^d (1 + h^2 \|\omega\|_2^2)^{-s} \\ &= K_\psi h^{d-2s} \left(\frac{1}{h^2} + \|\omega\|_2^2 \right)^{-s} \\ &\leq K_\psi h^{d-2s} \left(1 + \|\omega\|_2^2 \right)^{-s}, \end{aligned}$$

whenever $h \in (0,1)$. Thus, if $2s > \beta + d/2$, the function ψ_h belongs to the native space $W_2^\beta(\mathbb{R}^d)$ of Φ and its norm can be bounded by

$$\begin{aligned}
\|\psi_h\|_\Phi^2 &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{|\widehat{\psi_h}(\omega)|^2}{\widehat{\phi}(\omega)} d\omega \\
&\leq (2\pi)^{-d/2} \frac{K_\psi}{c_\phi} h^{2d-4s} \int_{\mathbb{R}^d} (1 + \|\omega\|_2^2)^{\beta-2s} d\omega \\
&=: M^2 h^{2d-4s}.
\end{aligned}$$

Finally, if we set $2s = \alpha + d/2$ with $\alpha > \beta$ we have

$$\|\psi_h\|_\Phi \leq M h^{\frac{d}{2}-\alpha}.$$

Given an arbitrary X with fill distance $h = h_{X,\Omega}$ we can find a point $x \in \Omega$ such that the ball of radius h around x contains no data point. We interpolate the function ψ_h shifted to that point. Then all data sites from X are outside the support of ψ_h so that $s_{\psi_h,X} = 0$. Then the standard error bound leads to

$$1 = \|\psi_h\|_{L_\infty(\Omega)} = \|\psi_h - s_{\psi_h,X}\|_{L_\infty(\Omega)} \leq \epsilon \|\psi_h\|_\Phi \leq M\epsilon h^{\frac{d}{2}-\alpha},$$

which immediately gives the stated result. □

Unfortunately, the constant M_α tends to infinity when α tends to β. Hence, we do not get

$$h_{X,\Omega}^{\beta-d/2} \leq C\epsilon$$

as we would have expected from (6), but we get as close as possible. Moreover, our proof does not work in case of the Gaussian, since it follows from the Paley-Wiener theory that there cannot be a compactly supported function in the native space of the Gaussian. Both drawbacks can be remedied if we make the additional assumption that X is already quasi-uniform, which means that the fill distance $h_{X,\Omega}$ essentially behaves like the *separation distance* $q_X := \min_{j \neq k} \|x_j - x_k\|_2$.

In this particular case we can define the function

$$f_y = \Phi(\cdot, y) - \sum_{j=1}^N u_j(y) \Phi(\cdot, x_j)$$

for every $y \in \Omega$. For this function we obviously have

$$|f_y(y) - s_{f_y,X}(y)| = P_{\Phi,X}(y) \|f_y\|_\Phi,$$

i.e. there is equality in (3). Hence, the assumption on the approximation properties of the set X gives

$$\epsilon \geq P_{\Phi,X}(y)$$

and the desired results follow from lower bounds on the power function. Such lower bounds have been thoroughly studied (see for example [5]) but are also in general given in terms of the separation distance. The paper [6] contains lower bounds for all relevant basis functions. Quasi-uniformity brings these bounds back to the fill distance.

4 A Greedy Data–independent Method

For numerical purposes it is crucial to have a practical method that produces well–distributed point sets. This can for example be achieved by a greedy algorithm that generates larger and larger point sets by adding a point where the power function of the previous set attains its maximum.

We start with $X_1 = \{x_1\}$ for $x_1 \in \Omega$ arbitrary. Then

$$X_j := X_{j-1} \cup \{x_j\} \text{ with } P_{\Phi,X_{j-1}}(x_j) = \|P_{\Phi,X_{j-1}}\|_{L_\infty(\Omega)},\ j \geq 2. \tag{8}$$

Practically, we maximize over some very large discrete set $X \subset \Omega$ instead of maximizing on Ω.

It is the goal of this section to show convergence of this method in the sense $\|P_{\Phi,X_j}\|_{L_\infty(\Omega)} \to 0$ for $j \to \infty$. To this end we make the additional assumption that either Ω is convex and $\Phi \in C^2(\Omega \times \Omega)$ or, alternatively, that $\Phi \in C^2(\Omega_1 \times \Omega_1)$ with $\Omega_1 \supseteq \Omega$ being convex.

We start our convergence analysis by two lemmas on the power function.

Lemma 4.1 *The power function (4) has the alternative representations*

$$\begin{aligned} P^2_{\Phi,X}(x) &= \Phi(x,x) - \sum_{j=1}^N u_j(x)\Phi(x,x_j) \\ &= \left\| \Phi(x,\cdot) - \sum_{j=1}^N \Phi(x,x_j)u_j \right\|^2_\Phi . \end{aligned}$$

Proof: Since the cardinal functions u_j satisfy

$$\sum_{k=1}^N u_k(x)\Phi(x_j,x_k) = \Phi(x,x_j), \qquad 1 \leq j \leq N, \tag{9}$$

the first equality follows immediately from the explicit form of the power function given in (4). Moreover, (9) implies

$$u_j(x) = \sum_{k=1}^{N} \alpha_{jk} \Phi(x, x_k),$$

if (α_{jk}) denotes the inverse to $(\Phi(x_j, x_k))$, so that

$$(u_j, u_k)_\Phi = \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_{jn} \alpha_{km} \Phi(x_n, x_m) = \alpha_{jk}.$$

This shows together with the reproducing property of the kernel that

$$\begin{aligned} \left\| \Phi(x, \cdot) - \sum_{j=1}^{N} \Phi(x, x_j) u_j \right\|_\Phi^2 &= \Phi(x, x) - 2 \sum_{j=1}^{N} \Phi(x, x_j) (\Phi(x, \cdot), u_j)_\Phi \\ &\quad + \sum_{j,k=1}^{N} \Phi(x, x_j) \Phi(x, x_k) (u_j, u_k)_\Phi \\ &= \Phi(x, x) - 2 \sum_{j=1}^{N} u_j(x) \Phi(x, x_j) \\ &\quad + \sum_{j=1}^{N} u_j(x) \Phi(x, x_j), \end{aligned}$$

which is $P_{\Phi,X}^2(x)$ by the first stated equality. □

Note that in the second equality of Lemma 4.1 the argument x has moved from u_j to Φ when compared to the definition of the power function.

In the following result we use the notation $\partial_j f$ to indicate the first order partial derivative of f with respect to the coordinate j. Moreover, $\partial_j^1 \Phi(x, y)$ shall mean that this derivative is taken with respect to the first argument of Φ. The gradient will, as usual, be denoted by ∇.

Lemma 4.2 *Suppose $\Phi \in C^2(\Omega \times \Omega)$. Then for every $X = \{x_1, \ldots, x_N\} \subseteq \Omega$ and every $1 \leq k \leq d$ we have*

$$|\partial_k P_{\Phi,X}^2(x)| \leq 2 P_{\Phi,X}(x) \sqrt{\partial_k^1 \partial_k^2 \Phi(x, x)}, \qquad x \in \Omega.$$

Proof: The second representation of the power function in Lemma 4.1 allows us to derive

$$\partial_k P^2_{\Phi,X}(x) = 2 \left(\partial^1_k \Phi(x,\cdot) - \sum_{j=1}^N \partial^1_k \Phi(x,x_j) u_j, \Phi(x,\cdot) - \sum_{j=1}^N u_j \Phi(x,x_j) \right)_\Phi .$$

This implies

$$\begin{aligned} |\partial_k P^2_{\Phi,X}(x)| &\leq 2 \left\| \partial^1_k \Phi(x,\cdot) - \sum_{j=1}^N \partial^1_k \Phi(x,x_j) u_j \right\|_\Phi P_{\Phi,X}(x) \\ &\leq 2 P_{\Phi,X}(x) \left\| \partial^1_k \Phi(x,\cdot) \right\|_\Phi \\ &= 2 P_{\Phi,X}(x) \sqrt{\partial^1_k \partial^2_k \Phi(x,x)}, \end{aligned}$$

where the last inequality holds since $\sum_{j=1}^N \partial^1_k \Phi(x,x_j) u_j$ is the interpolant to $\partial^1_k \Phi(x,x_j)$ and hence equals its best approximant from V_X with respect to the native space norm. □

Now it is time to come back to our greedy method described in (8). For simplicity we define $P_j := P_{\Phi,X_j}$. Remember that the point x_{j+1} is chosen in such a way that $P_j(x_{j+1}) = \|P_j\|_{L_\infty(\Omega)}$. Moreover, since $X_j \subseteq X_{j+1}$ we have $P_j(x) \geq P_{j+1}(x) \geq 0$ for all $x \in \Omega$.

Theorem 4.3 *Suppose $\Omega \subseteq \mathbb{R}^d$ is compact and satisfies an interior cone condition. Suppose further that $\Phi \in C^2(\Omega_1 \times \Omega_1)$ is a positive definite kernel defined on a convex and compact region $\Omega_1 \supseteq \Omega$. Then, the greedy algorithm defined in (8) converges at least like*

$$\|P_j\|_{L_\infty(\Omega)} \leq C j^{-1/d}$$

with a constant $C > 0$.

Proof: From Lemma 4.2 it follows that there exists a constant $C_\nabla > 0$ such that

$$\|\nabla P^2_j(x)\|_2 \leq C_\nabla P_j(x), \qquad x \in \Omega, j \in \mathbb{N}.$$

Hence, there exists ξ and η on the line segment between x and x_{j+1} with

$$\begin{aligned} P^2_j(x) &= P^2_j(x_{j+1}) + \nabla P^2_j(\xi) \cdot (x - x_{j+1}) \\ &\geq \|P_j\|^2_{L_\infty(\Omega)} - \|\nabla P^2_j(\xi)\|_2 \|x - x_{j+1}\|_2 \\ &\geq \|P_j\|^2_{L_\infty(\Omega)} - C_\nabla \|P_j\|_{L_\infty(\Omega)} \|x - x_{j+1}\|_2 \\ &= \|P_j\|_{L_\infty(\Omega)} \left(\|P_j\|_{L_\infty(\Omega)} - C_\nabla \|x - x_{j+1}\|_2 \right) \end{aligned}$$

and, because of $P_{j+1}(x_{j+1}) = 0$, we have

$$\begin{aligned} P_{j+1}^2(x) &= P_{j+1}^2(x_{j+1}) + \nabla P_j^2(\eta) \cdot (x - x_{j+1}) \\ &\leq 0 + C_\nabla \|P_{j+1}\|_{L_\infty(\Omega)} \|x - x_{j+1}\|_2 \\ &\leq C_\nabla \|P_j\|_{L_\infty(\Omega)} \|x - x_{j+1}\|_2. \end{aligned}$$

Both inequalities together yield

$$\begin{aligned} P_j^2(x) - P_{j+1}^2(x) &\geq \|P_j\|_{L_\infty(\Omega)} \left(\|P_j\|_{L_\infty(\Omega)} - 2C_\nabla \|x - x_{j+1}\|_2 \right) \\ &\geq \frac{1}{2} \|P_j\|_{L_\infty(\Omega)}^2 \end{aligned}$$

provided that

$$\|x - x_{j+1}\|_2 \leq \frac{\|P_j\|_{L_\infty(\Omega)}}{4C_\nabla} =: \delta_j.$$

For such an x we can continue by

$$\begin{aligned} \frac{1}{2} \|P_j\|_{L_\infty(\Omega)}^2 &\leq P_j^2(x) - P_{j+1}^2(x) \\ &= (P_j(x) - P_{j+1}(x))(P_j(x) + P_{j+1}(x)) \\ &\leq 2\|P_j\|_{L_\infty(\Omega)} (P_j(x) - P_{j+1}(x)), \end{aligned}$$

which leads to

$$\|P_j\|_{L_\infty(\Omega)} \leq 4 \left(P_j(x) - P_{j+1}(x) \right), \qquad \text{for } \|x - x_{j+1}\|_2 \leq \delta_j. \tag{10}$$

If we look at $x = x_k$ for some $k \leq j$, we see that $P_j(x_k) = P_{j+1}(x_k) = 0$, and thus $\|x_k - x_{j+1}\|_2 > \delta_j$ by (10), provided $\|P_j\|_{L_\infty(\Omega)} > 0$. Since we can assume that all $\|P_j\|_{L_\infty(\Omega)}$ are positive, and since the δ_j are nonincreasing, we have $\|x_\ell - x_k\|_2 > \delta_{\ell-1} \geq \delta_j$ for all $k < \ell \leq j$. Hence, all x_k with $k \leq j$ are centers of disjoint balls $B(x_k, \delta_j/2)$ of radius $\delta_j/2$. Moreover, the union of all these balls is contained in the bounded region $\Omega^* = \cup_{x \in \Omega} B(x, \delta_1/2)$. Hence, if we compare the volume of the union of all balls $B(x_k, \delta_j/2)$, $1 \leq k \leq j$, to the volume of Ω^*, we see that there is a constant C, independent of j, such that $j(\delta_j/2)^d \leq C$, and we finally get

$$\|P_j\|_{L_\infty(\Omega)} = 4C_\nabla \delta_j \leq 8C_\nabla C^{1/d} j^{-1/d}.$$

□

5 Geometric Greedy Method

Practical examples show that the greedy minimization of the power function usually just tries to fill the currently largest hole in the data by placing a new data point close to the center of that hole. This strategy is independent of the radial basis in question:

- Let Ω be a compact set in $\mathbb{R}^d$, and start with $X_0 = \emptyset$. Define the distance of $x \in \Omega$ to $X_0 = \emptyset$ as some value which is not smaller than the diameter of Ω, e.g. the diameter of the bounding box.
- If X_n is a finite subset of Ω consisting of n points, pick $x_{n+1} \in \Omega \setminus X_n$ such that it has maximal distance to X_n and form $X_{n+1} := X_n \cup \{x_{n+1}\}$.

If Ω is finite, each step of the algorithm can be carried out in $\mathcal{O}(|\Omega|)$ operations, because one can keep for each $x \in \Omega$ the current distance to its nearest neighbor within X_n. Updating this array of length $|\Omega|$ is done by first calculating the $|\Omega|$ values $\|x - x_{n+1}\|_2$ and taking the componentwise minimum with the existing distance array. The next point is then easily found by picking the maximum of the array. This technique can be refined, but we omit such arguments here.

It turns out that the above greedy algorithm works nicely when it comes to finding subsets of Ω of size n with small fill distance $h_{X,\Omega}$ and large separation distance q_X. Define

$$q_n := \frac{1}{2} \min_{x \neq y \in X_n} \|x - y\|_2,$$

$$d_n(x) := \min_{y \in X_n} \|x - y\|_2,$$

$$h_n := \max_{x \in \Omega} d_n(x) = \max_{x \in \Omega} \min_{y \in X_n} \|x - y\|_2 = d_n(x_{n+1}) = h_{X_n,\Omega}.$$

Lemma 5.1 *The geometric greedy algorithm produces point sets which are quasi-uniform. To be more precise,*

$$h_n \geq q_n \geq \frac{1}{2} h_{n-1} \geq \frac{1}{2} h_n \text{ for all } n \geq 2.$$

Proof: The left–hand and right–hand sides are clear. For X_2 we have

$$q_2 := \frac{1}{2}\|x_1 - x_2\|_2 = \frac{1}{2} d_1(x_2) = \frac{1}{2} h_1.$$

Assume $q_n \geq \frac{1}{2} h_{n-1}$ and look at

$$\begin{aligned} q_{n+1} &= \min\left(q_n, \frac{1}{2} \min_{x \in X_n} \|x_{n+1} - x\|_2\right) \\ &= \min\left(q_n, \frac{1}{2} d_n(x_{n+1})\right) \\ &= \min\left(q_n, \frac{1}{2} h_n\right) \end{aligned}$$

to get

$$q_{n+1} \geq \min\left(\frac{1}{2} h_{n-1}, \frac{1}{2} h_n\right) \geq \frac{1}{2} h_n.$$

□

If Ω is a bounded region in $\mathbb{R}^d$, the geometric greedy method constructs asymptotically uniformly distributed data sets that cover Ω in an asymptotically optimal way. In fact, the balls with centers in X_n and radius h_n cover Ω, while those with radius q_n are disjoint. With

$$\Omega_n := \{y \in \mathbb{R}^d \ : \ dist(y, \Omega) \leq q_n\}$$

we find

$$\begin{aligned} n q_n^d v_1 &\leq vol(\Omega_n) \\ vol(\Omega) &\leq n h_n^d v_1, \end{aligned}$$

where v_1 denotes the volume of the unit ball in $\mathbb{R}^d$, and see that both h_n and q_n decay asymptotically like $n^{-1/d}$. Note that this rate also occurs in the proof of Theorem 4.3, but only for the δ_n and q_n.

6 Examples

We start with two examples on the domain $[-1, 1]^2$. Everything was discretized on a regular set of $71 \times 71 = 5041$ points, and the greedy method was executed until the norm of the power function fell under $2 \cdot 10^{-5}$. For the Gaussian with scale 1, this required 48 points distributed as shown in Figure 1. The "error" in the right–hand figure is $\|P_N\|^2_{L_\infty(\Omega)}$ with decay as a function of the number N of data points. As determined by the regression line in the figure, the decay is like $N^{-7.2}$.

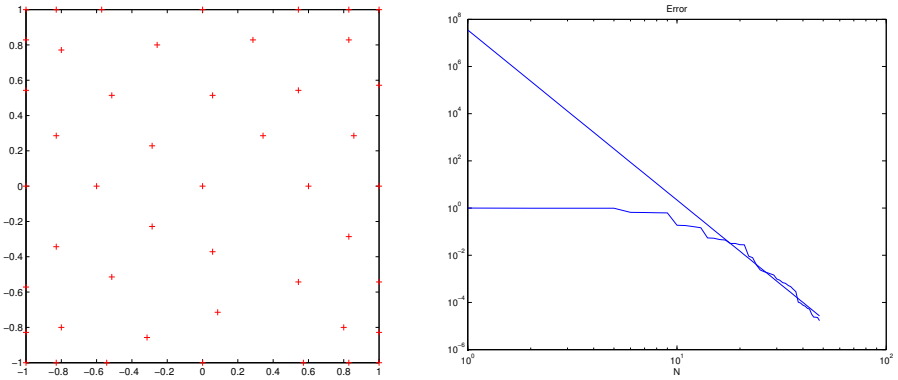

Figure 1: Gaussian

Figure 2 uses the C^2 Wendland function with scale 15. It requires $N = 100$ points to depress the power function down to $2 \cdot 10^{-5}$. The error decays like $N^{-1.9}$ as determined by the regression line in the figure.

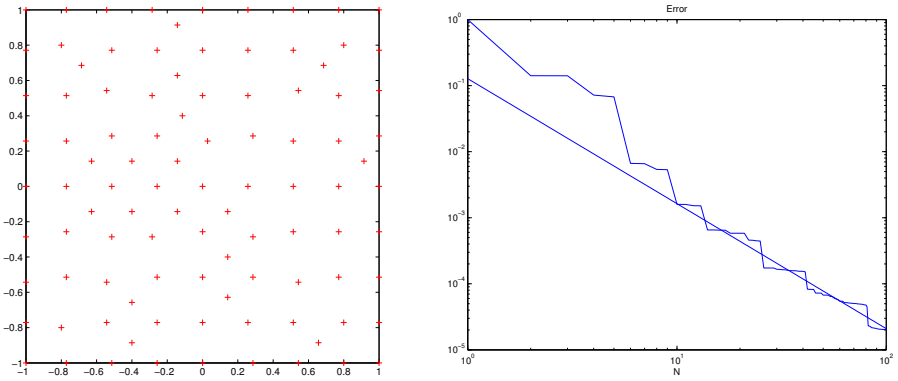

Figure 2: Wendland's function

Figure 3 shows the error decay when the Gaussian power function is evaluated on the data supplied by the geometric greedy method up to X_{48}. The final error is larger by a factor of 4, and the estimated decrease of the error is only like $N^{-6.1}$. In case of Wendland's function in Figure 4 the error factor is only 1.4, while the estimated decay order is -1.72.

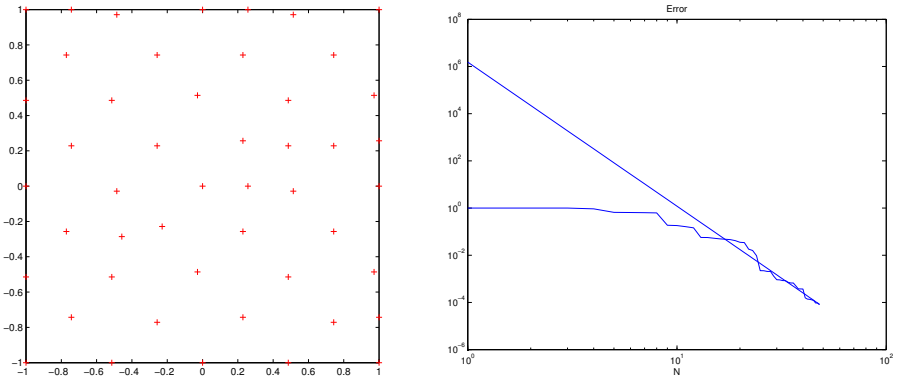

Figure 3: Gaussian, geometric greedy data

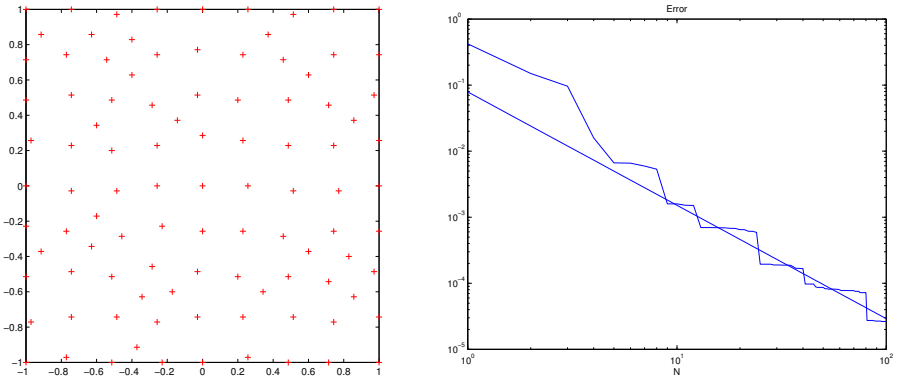

Figure 4: Wendland function, geometric greedy data

Final remarks.

- Note that for asymptotically uniformly distributed points we would theoretically get an arbitrarily high decay rate in case of the Gaussian, while for Wendland's function we expect a decay rate of -1.5 due to the limited smoothness (see [8]).

- Both algorithms were run starting from a regular discretization of the square $[-1,1]^2$. One intuitively would expect that since the domain is symmetric, the kernels are symmetric functions as well as the power functions, the resulting points should be symmetric. The

non-symmetric distribution is essentially due to the non-uniqueness in the choices where maxima are attained. To be more precise, since the implementation of our numerical experiments has been done by Matlab, the `max` function works so that *if the values along the first non-singleton dimension contain more than one maximal element, the index of the first one is returned*, which explains why there is no symmetry in the points distribution. But this is also clear from looking at our two methods: both require to maximize a multivariate function whose relative maxima are in general not unique.

- Though the geometric greedy algorithm can be proven to generate an asymptotically optimal sequence, its independence of the kernel makes it still inferior to the greedy algorithm that takes maxima of the power function. However, so far there is no proof that the latter algorithm generates a sequence with $h_N \leq CN^{-1/d}$, as required for asymptotic optimality.

References

[1] Beatson, R. K., Light, W. A., and Billings, S. Fast solution of the radial basis function interpolation equations: Domain decomposition methods. *SIAM J. Sci. Comput. 22* (2000), 1717–1740.

[2] Beyer, A. Optimale Centerverteilung bei Interpolation mit radialen Basisfunktionen. Diplomarbeit, Universität Göttingen, 1994.

[3] Bos, L. P., and Maier, U. On the asymptotics of points which maximize determinants of the form $\det(g(|x_i - x_j|))$. In *Advances in Multivariate Approximation* (Berlin, 1999), W. Haussmann, K. Jetter, and M. Reimer, Eds., vol. 107 of *Math. Res.*, Wiley-VCH., pp. 1–22.

[4] Iske, A. Optimal distribution of centers for radial basis function methods. Tech. Rep. M0004, Technische Universität München, 2000.

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

[6] Schaback, R. Multivariate interpolation and approximation by translates of a basis function. In *Approximation Theory VIII*, C. K. Chui and L. Schumaker, Eds., vol. 1: Approximation and interpolation. World Scientific Publishing, 1995, pp. 491–514.

[7] SCHABACK, R. Native Hilbert spaces for radial basis functions I. In *New Developments in Approximation Theory*, M. Buhmann, D. H. Mache, M. Felten, and M. Müller, Eds., no. 132 in International Series of Numerical Mathematics. Birkhäuser Verlag, 1999, pp. 255–282.

[8] WENDLAND, H. Error estimates for interpolation by compactly supported radial basis functions of minimal degree. *Journal of Approximation Theory 93* (1998), 258–272.

[9] WU, Z., AND SCHABACK, R. Local error estimates for radial basis function interpolation of scattered data. *IMA Journal of Numerical Analysis 13* (1993), 13–27.