Optimal data-independent point locations for RBF interpolation

S. De Marchi, *R. Schaback and H. Wendland*

Università di Verona (Italy), *Universität Göttingen (Germany)*

Preliminaries

- $X = \{x_1, ..., x_n\} \subseteq \Omega \subseteq \mathbb{R}^d$, distinct, *data sites*.
- $\{f_1, ..., f_N\}$, *data values* to be interpolated.

RBF interpolation (easiest): fix a symmetric PD kernel $\Phi : \Omega \times \Omega \to \mathbb{R}$ and form

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

$A_{\Phi,X} := (\Phi(x_i, x_j))_{1 \le i,j \le N}$: the interpolation matrix, invertible. If $A_{\Phi,X}$ is *even positive definite* $\forall\, X \subseteq \Omega$, then Φ is called a *positive definite, PD* kernel. It is often *radial*, $\Phi(x, y) = \phi(\|x - y\|_2)$, and therefore defined on $\mathbb{R}^d \times \mathbb{R}^d$ because *every CPD kernel has an associated normalized PD kernel*.

Some useful notations

- Take $V_X = \text{span}\{\Phi(\cdot, x) : x \in X\}$. The interpolant $s_{f,X}$ can be written in terms of *cardinal functions* $u_j \in V_X$, $u_j(x_i) = \delta_{ji}$, i.e. $s_{f,X} = \sum_{j=1}^{N} f(x_j) u_j$.
- For the purpose of stability and error analysis the following quantities are important:

  separation distance: $q_X = \min_{x_j, x_k \in X, j \neq k} \|x_j - x_k\|_2$;

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

  uniformity: $\rho_{X,\Omega} = \dfrac{q_X}{h_{X,\Omega}}$.

The problem

$$\boxed{\text{Are there any good or even optimal point sets for the interpolation problem?}}$$

Literature

1. BEYER, A. Optimale Centerverteilung bei Interpolation mit radialen Basisfunktionen. Diplomarbeit, Universität Göttingen, 1994.
2. 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.
3. ISKE, A., Optimal distribution of centers for radial basis function methods. Tech. Rep. M0004, Technische Universität München, 2000.

Literature

1. BEYER, A. considered numerical aspects of the problem.

2. BOS, L. P., AND MAIER, U., investigated on Fekete-type points for univariate RBFs for a broad class of functions Φ, proving that: *Equally spaced points give asymptotically largest determinants for the interpolation matrix* $A_{\Phi,X}$.

3. ISKE, A. constructed and characterized *admissible* sets by varying the centers for stability and quality of approximation by RBF, proving that *uniformly* distributed points gives better results. He also provided a bound for the *uniformity*: $\rho_{X,\Omega} \le \sqrt{\frac{2(d+1)}{d}}$, d= space dimension.

Our approach

(I) **Power function estimates.**

(II) **Geometric arguments.**

Power function estimates

The kernel Φ defines on the space $V_\Omega = \mathrm{span}\{\Phi(\cdot, x) : x \in \Omega\}$ an inner product

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

so that Φ is a reproducing kernel of V_Ω. Set $\bar{V_\Omega} := \mathcal{N}_\Phi(\Omega)$, the *native Hilbert space*. If $f \in \mathcal{N}_\Phi(\Omega)$, then

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

and by Cauchy-Schwarz inequality

$$|f(x) - s_{f,X}(x)| \le P_{\Phi,X}(x) \|f\|_\Phi \tag{2}$$

Some properties of the Power function

1. $P_{\Phi,X}(x)$ is the norm of the *pointwise error functional*;
2. Error estimates bound $P_{\Phi,X}(x)$ in terms of the *fill distance* $h_{X,\Omega}$;
3. If $X \subseteq Y$ then $P_{\Phi,X}(x) \geq P_{\Phi,Y}(x), \ \ \forall x \in \Omega$.

• • •

If Φ is translation invariant, integrable and has Fourier transform such that $c_\phi(1+\|\omega\|_2^2)^{-\beta} \leq \widehat{\phi}(\omega) \leq C_\Phi(1+\|\omega\|_2^2)^{-\beta}$ with $\beta > d/2$, $C_\phi \geq c_\phi > 0$, then $\mathcal{N}_\Phi(\mathbb{R}^d)$ is norm-equivalent to the space $W_2^\beta(\mathbb{R}^d)$. Therefore

$$\|f - s_{f,X}\|_{L_\infty(\Omega)} \leq C h_{X,\Omega}^{\beta-d/2} \|f\|_{W_2^\beta(\mathbb{R}^d)}. \tag{3}$$

Main result

The hypotheses on Ω and Φ are as before.

Theorem 1 *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*

$$\|f - s_{f,X}\|_{L_\infty(\Omega)} \leq \epsilon \|f\|_\Phi, \qquad \text{for all } f \in W_2^\beta(\mathbb{R}^d), \tag{4}$$

then the fill distance of X satisfies

$$h_{X,\Omega} \leq M_\alpha \epsilon^{\frac{1}{\alpha - d/2}}. \tag{5}$$

Comment: optimally distributed data sites are sets that cannot have a large region in Ω without centers, i.e. $h_{X,\Omega}$ is sufficiently small.

Quasi-uniformity and fill-distance

The previous theorem fails in two situations:

$(\mathbf{a})$ When $\alpha \to \beta$ we have $M_\alpha \to \infty$ and we don't get $h_{X,\Omega}^{\beta-\frac{d}{2}} \leq C\epsilon$.

$(\mathbf{b})$ Φ is the Gaussian (cfr. Paley-Wiener theory).

Now, **assuming that X is already** *quasi-uniform*, i.e. $h_{X,\Omega} \approx q_X$, we can define $f_y = \Phi(\cdot, y) - \sum_{j=1}^N u_j(y)\Phi(\cdot, x_j)$ for every $y \in \Omega$. For this function we have

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

i.e. there is equality in (2). Hence, the assumption on the approximation properties of the set X gives $P_{\Phi,X}(y) \leq \epsilon$ and the desired results follow from lower bounds on the power function .

The Greedy Method (G.M.)

Idea: we generate larger and larger data sets by adding the maxima of the Power function w.r.t. preceeding set. This method produces well-distributed point sets.

Greedy Algorithm (G.A.)

- **starting step:** $X_1 = \{x_1\}, \ \ x_i \in \Omega, \ arbitrary$.
- **iteration step:** $X_j = X_{j-1} \cup \{x_j\}$ with $P_{\Phi, X_{j-1}}(x_j) = \|P_{\Phi, X_{j-1}}\|_{L_\infty(\Omega)}$.

Convergence: we hope that $\|P_{\Phi, X_j}\|_{L_\infty(\Omega)} \to 0$ as $j \to \infty$ when Ω convex, $\Phi \in \mathcal{C}^2(\Omega \times \Omega)$ or $\Phi \in \mathcal{C}^2(\Omega_1 \times \Omega_1)$, $\Omega \subseteq \Omega_1$ convex.

The greedy algorithm converges

Theorem 2 *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** *converges at least like*

$$\|P_j\|_{L_\infty(\Omega)} \le C\, j^{-\frac{1}{d}}$$

with a constant $C > 0$.

Remark: $P_j := P_{X_j,\Phi}$.

Geometric Greedy Method (G.G.M.)

Notice: Practical experiments show that the greedy minimization algorithm of the power function, fills the currently largest hole in the data point close to the center of the hole.

Geometric Greedy Algorithm (G.G.A.)

- **starting step:** $X_0 = \emptyset$ and define $\mathrm{dist}(x, \emptyset) := A,\ A > \mathrm{diam}(\Omega)$.
- **iteration step:** given $X_n \in \Omega, |X_n| = n$ pick $x_{n+1} \in \Omega \setminus X_n$ s.t. $x_{n+1} = \max_{x \in \Omega \setminus X_n} \mathrm{dist}(x, X_n)$. Then, form $X_{n+1} := X_n \cup \{x_{n+1}\}$.

Remark: the algorithm works very well for subsets X_n of Ω, with **small** fill-distance $h_{X,\Omega}$ and **large** separation distance q_X.

Convergence of the G.G.A.

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

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

Lemma 1 *The G.G.A. produces point sets which are quasi-uniform. To be more precise,*

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

Remarks

If Ω is a bounded region in $\mathbb{R}^d$, the G.G.A. constructs asymptotically uniformly distributed data sets that cover Ω in asymptotically optimal way since $B(X_n, h_n)$ cover Ω while $B(X_n, q_n)$ are disjoint. With

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

we find

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

$v_1 = \text{vol}\left(B_1(\mathbb{R}^d)\right)$, showing that both h_n and q_n decay asymptotically like $n^{-1/d}$.

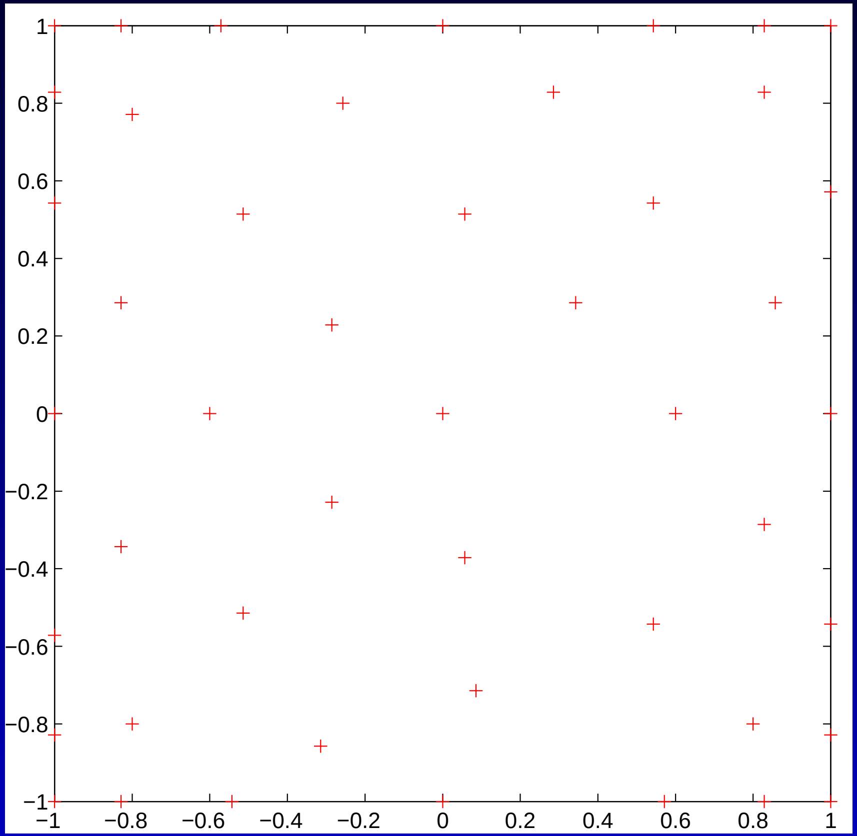
Examples

$\Omega = [-1,1] \times [-1,1]$ discretized on a regular grid of $5041 = 71 \times 71$ pts.

The kernels are: the Gaussian (with scale 1) and Wendland's function (with scale 15).

- **Greedy method (G.M.)**. Executed untill $\|P_N\|^2_{L_\infty(\Omega)} \le 2 \cdot 10^{-5}$.
- **Greedy geometric method (G.G.M.)**. The sets X_n are computed by the G.G.A. while the error is evaluated on this point set.

G.M. and G.G.M. : Gaussian I



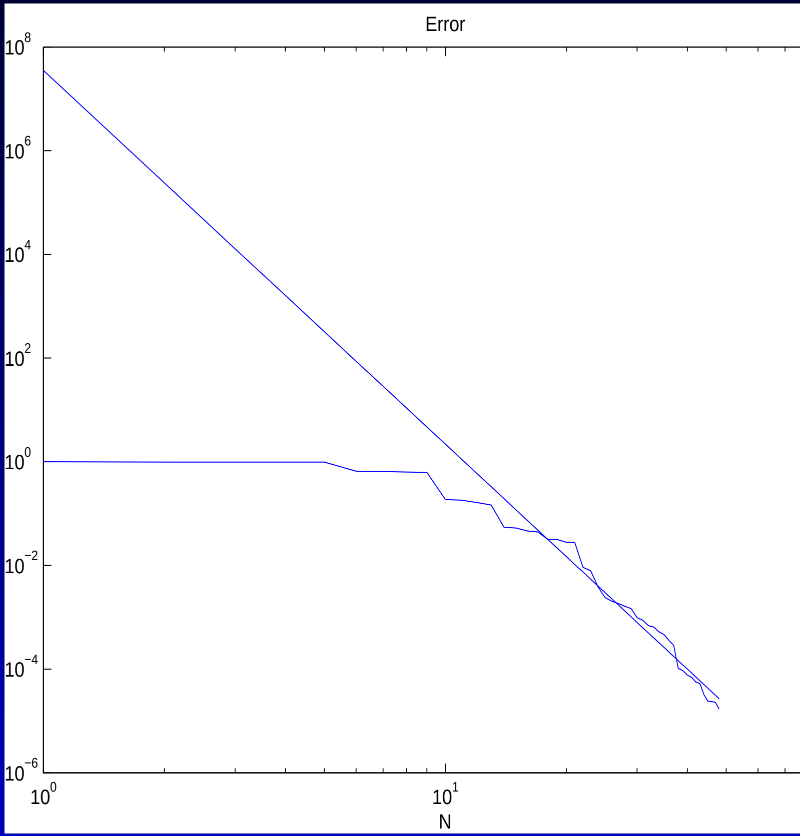


Figure 1: Gaussian: (left) the N=48 optimal points, (right) the error as function of N, decays as $N^{-7.2}$

G.M. and G.G.M. : Gaussian II

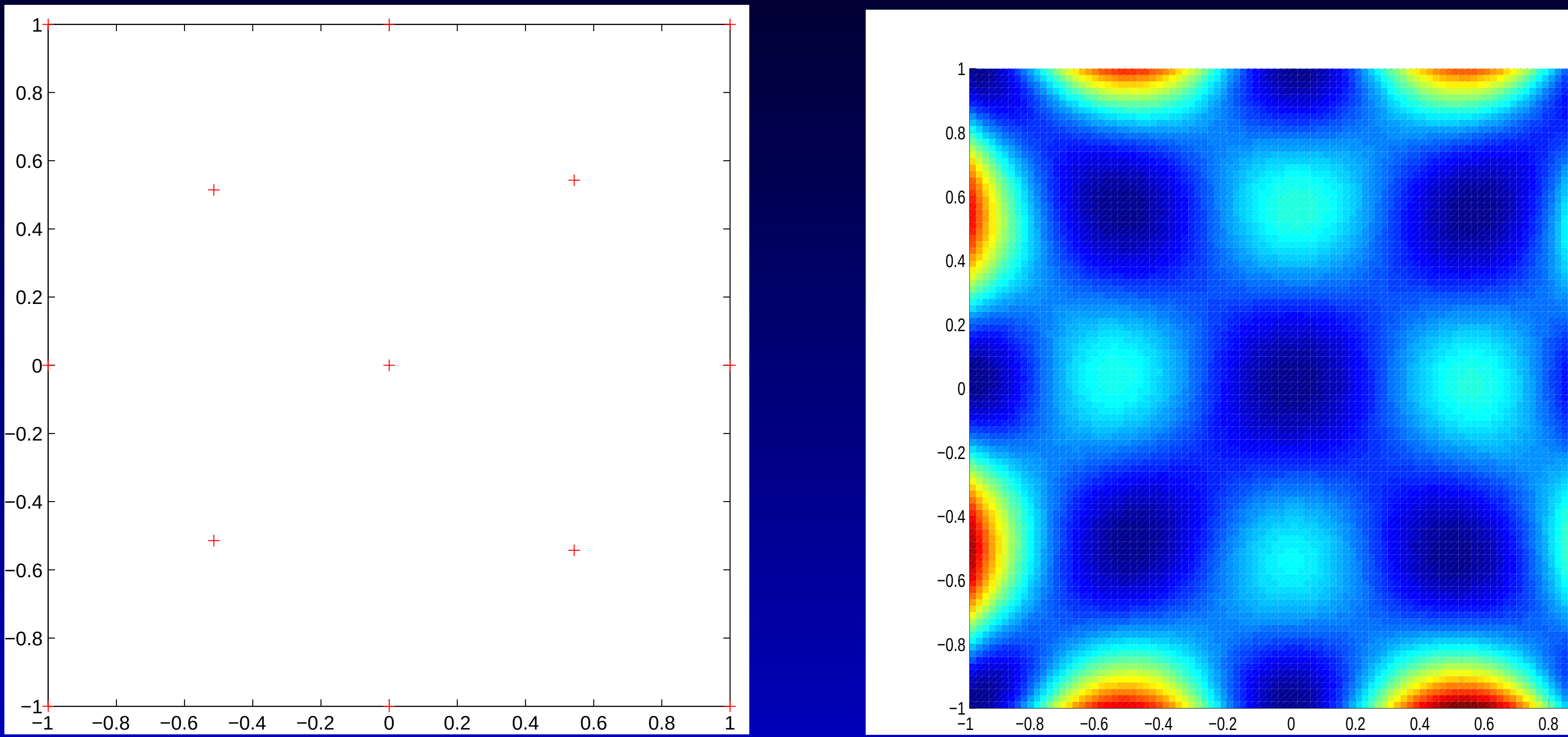
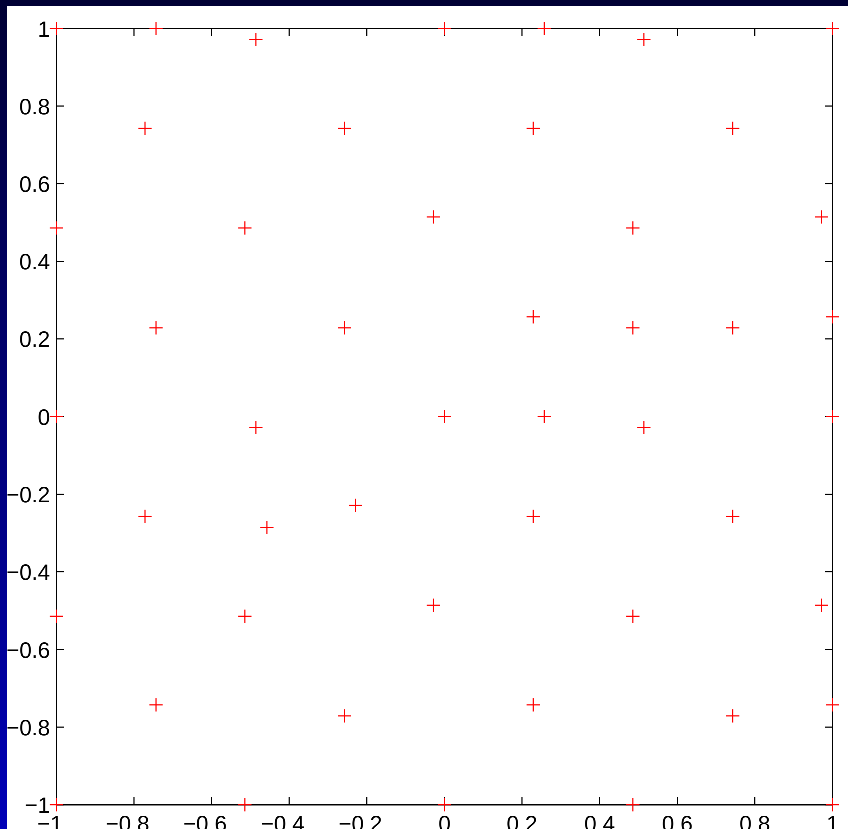


Figure 2: Gaussian: (left) the N=13 optimal points when $\|P_N\|^2_{L_\infty(\Omega)} \leq 0.1$, (right) the power function where the maxima are taken.

G.M. and G.G.M. : Gaussian III



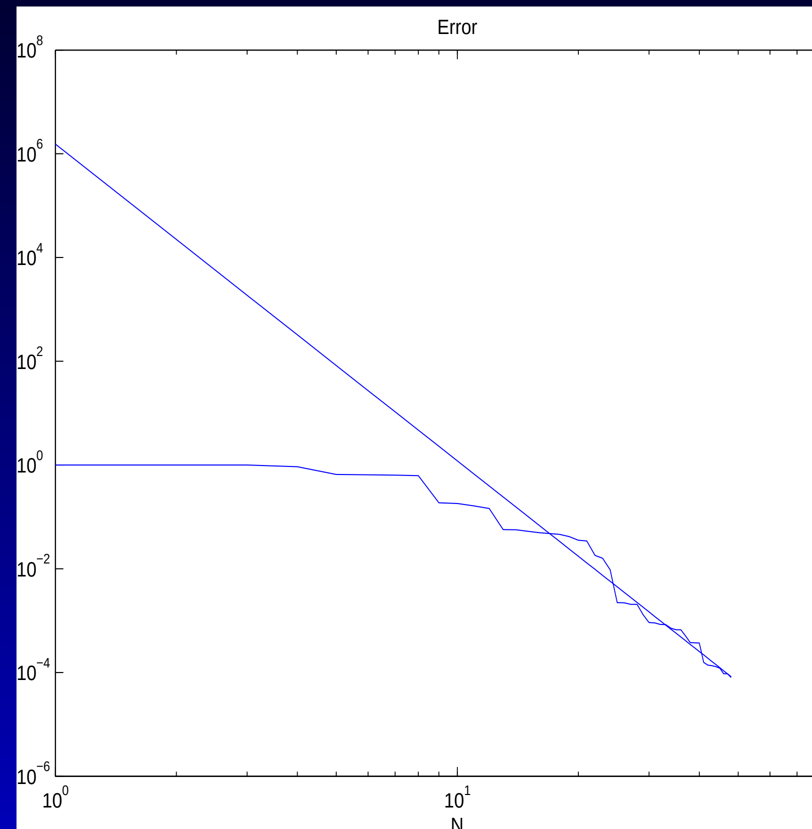
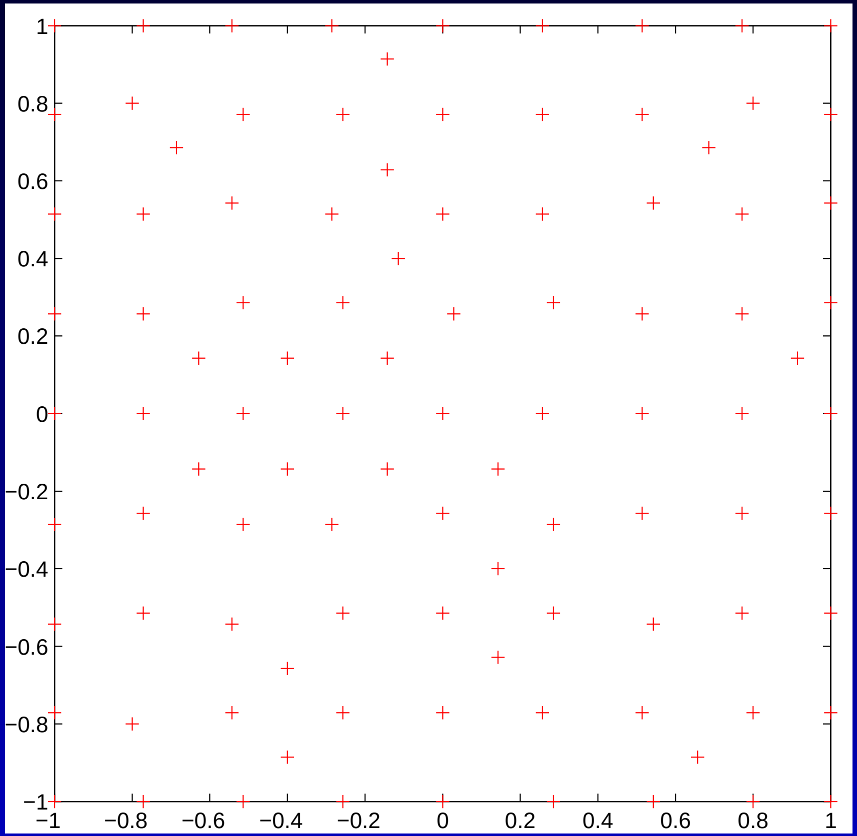


Figure 3: Gaussian, (left) geometric greedy data X_{48}, (right) the error is larger by a factor 4 and decays as $n^{-6.1}$.

G.M. and G.G.M.: Wendland's function I



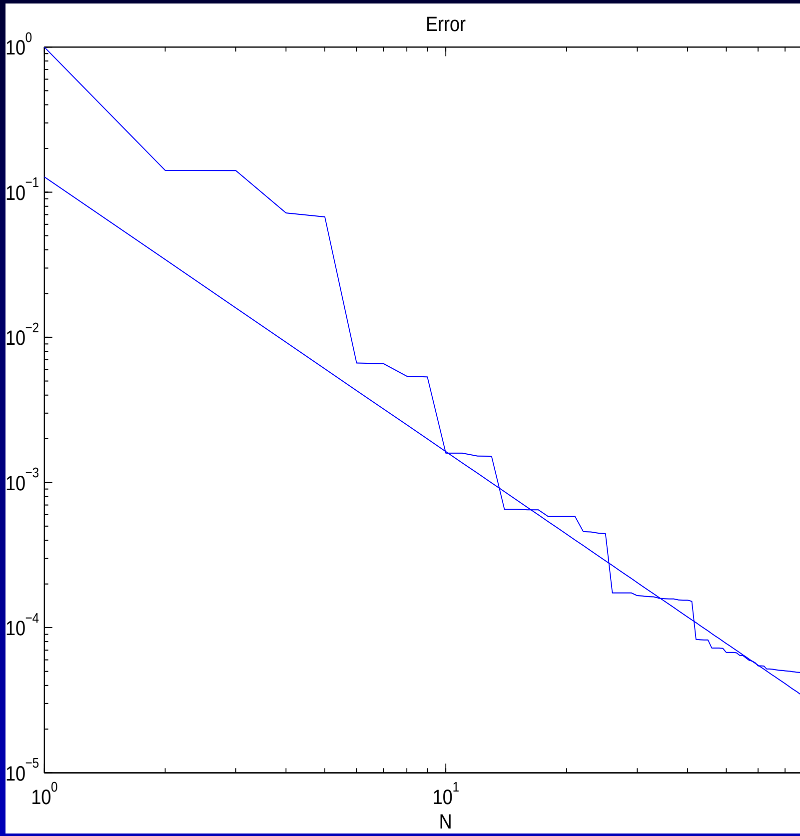
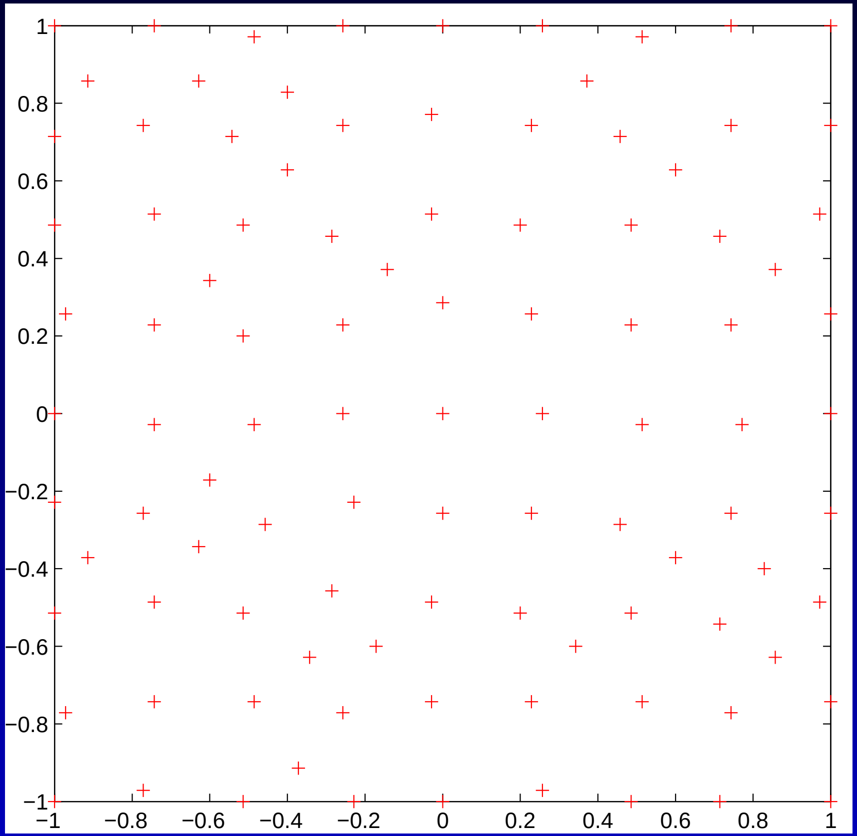


Figure 4: Wendland's fnc.: (left) the N=100 optimal points, (right) the error as function of N that decays as $N^{-1.9}$.

G.M. and G.G.M.: Wendland's function II



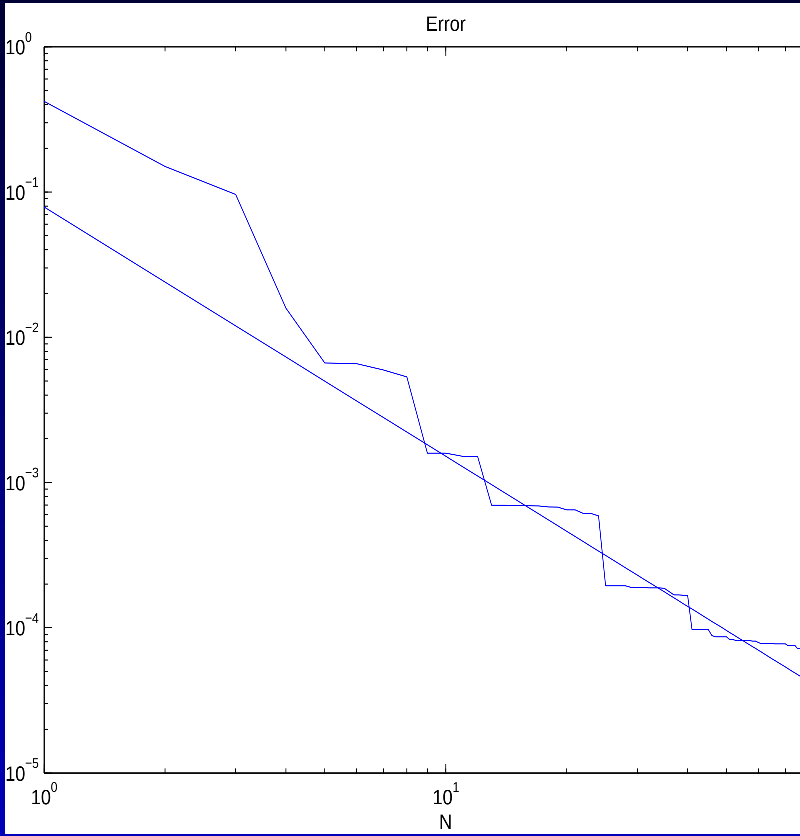


Figure 5: Wendland's fnc., (left) geometric greedy data X_{100}, (right) the error is larger by a factor 1.4 and decays as $n^{-1.72}$.

Distances

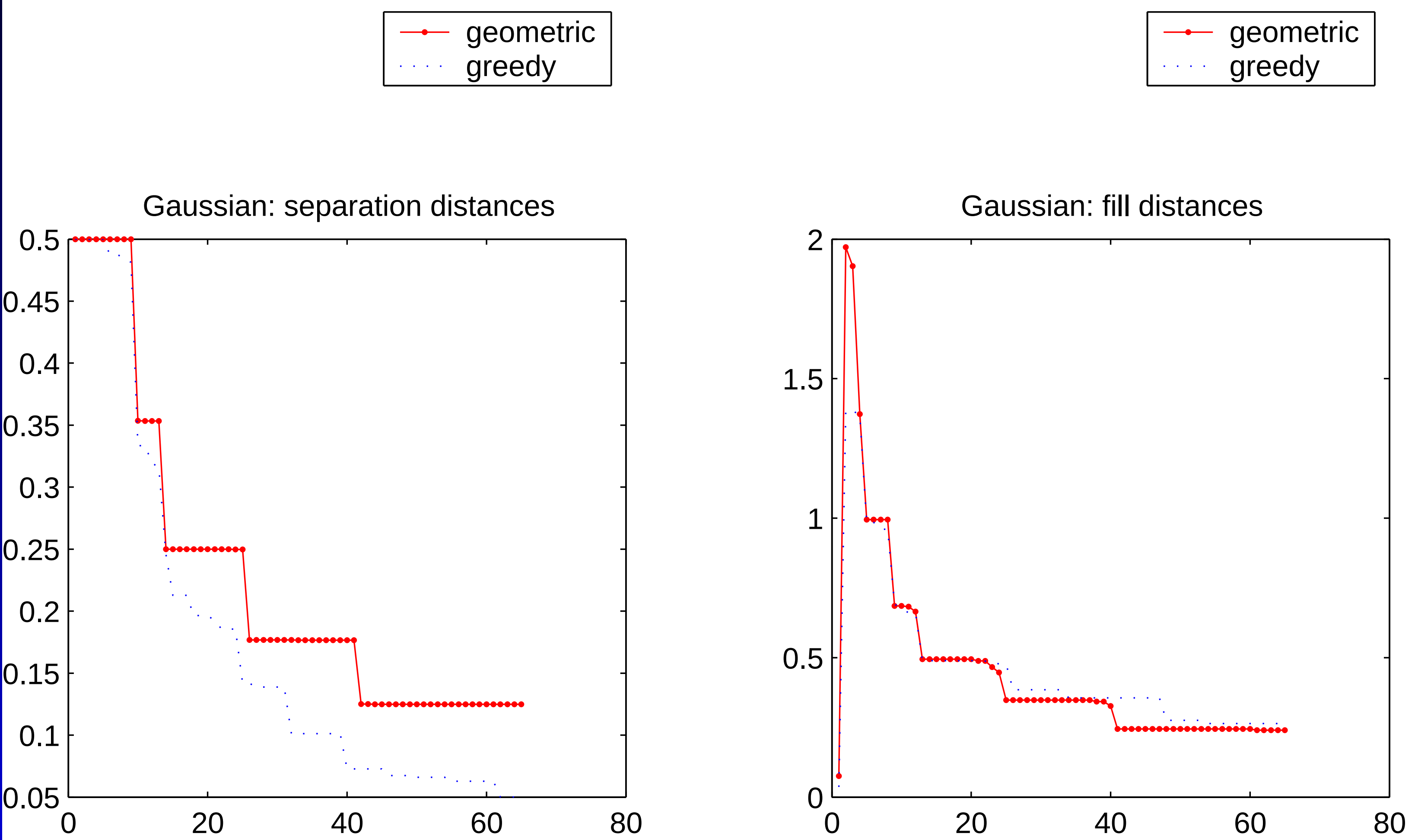


Figure 6: Gaussian.

Distances

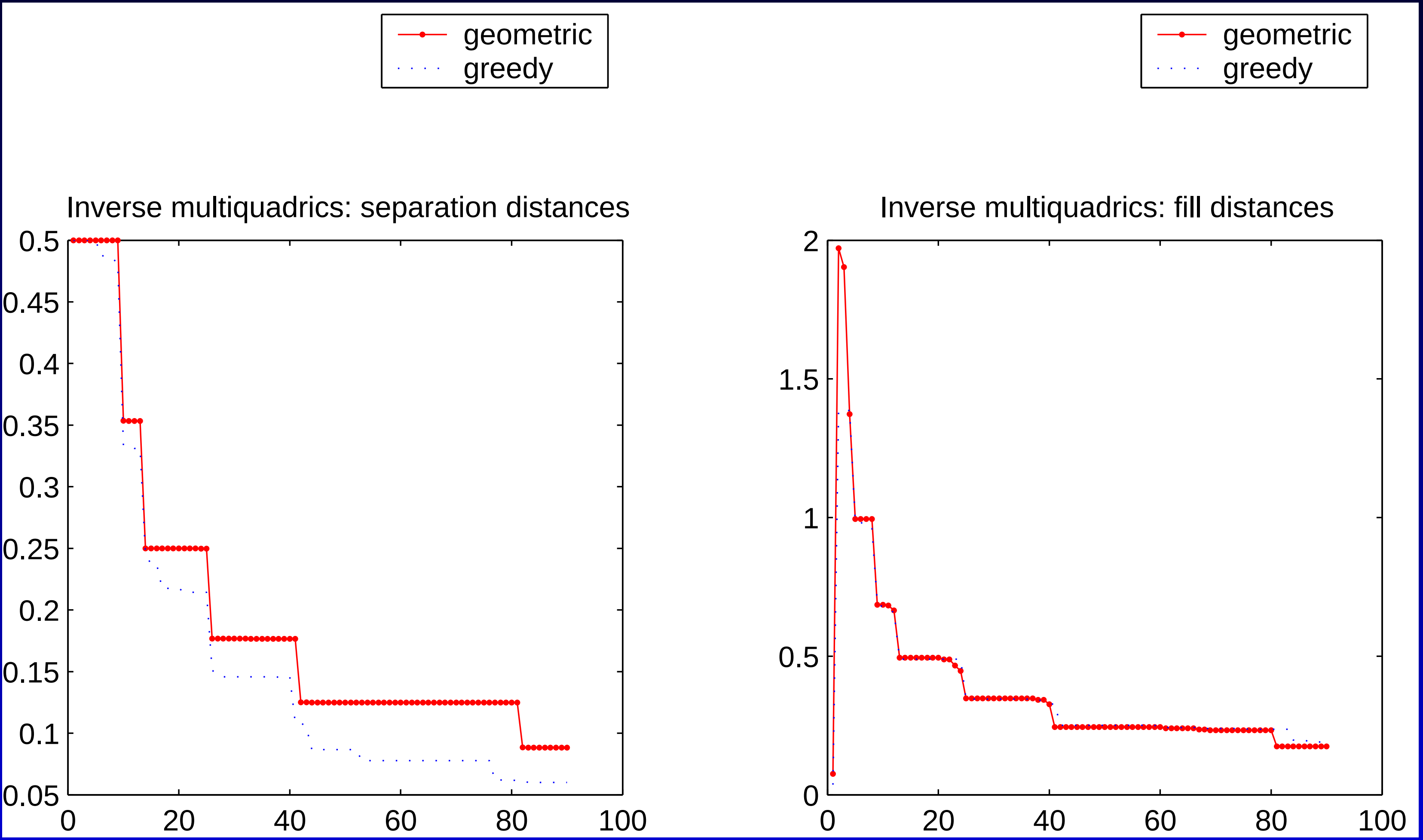


Figure 7: Inverse Multiquadrics function.

Distances

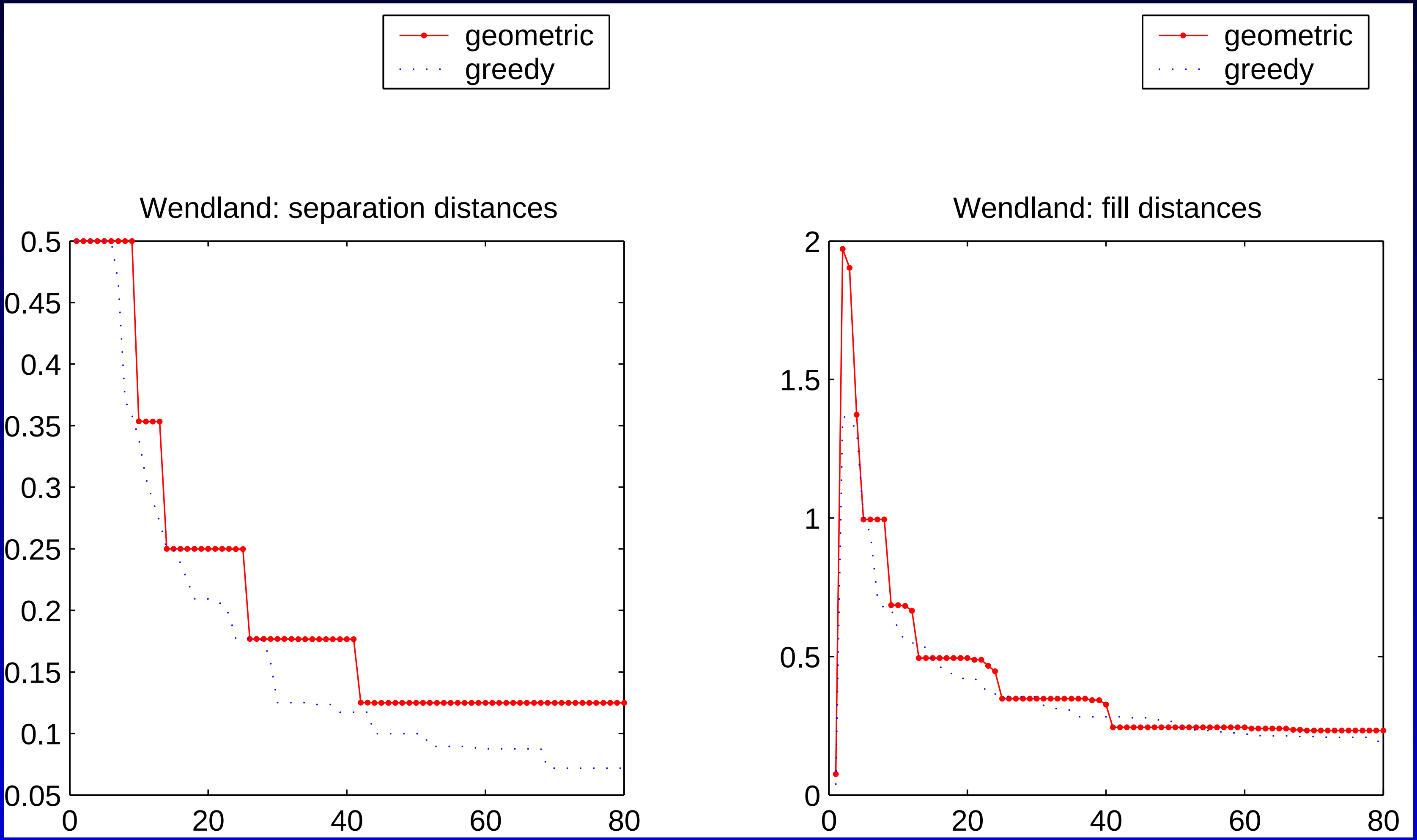


Figure 8: Wendland's function.

Final remarks

- The G.G.A. is independent on the kernel and we proved that generates asymptotically optimal sequences. It still inferior to the G.A. that takes maxima of the power function.
- So far, we have no proof of the fact the G.G.A. generates a sequence with $h_n \leq C n^{-1/d}$, as required by asymptotic optimality.