

Stability of Kernel–Based Interpolation

Stefano De Marchi

*Department of Computer Science,
University of Verona (Italy)*

Robert Schaback

*Institut für Numerische und Angewandte Mathematik,
University of Göttingen (Germany)*

Abstract

It is often observed that interpolation based on translates of radial basis functions or non-radial kernels is numerically unstable due to exceedingly large condition of the kernel matrix. But if stability is assessed in function space without considering special bases, this paper proves that kernel–based interpolation is stable. Provided that the data are not too wildly scattered, the L_2 or L_∞ norms of interpolants can be bounded above by discrete ℓ_2 and ℓ_∞ norms of the data. Furthermore, Lagrange basis functions are uniformly bounded and Lebesgue constants grow at most like the square root of the number of data points. However, this analysis applies only to kernels of limited smoothness. Numerical examples support our bounds, but also show that the case of infinitely smooth kernels must lead to worse bounds in future work, while the observed Lebesgue constants for kernels with limited smoothness even seem to be independent of the sample size and the fill distance.

1 Introduction

We consider the recovery of a real–valued function $f : \Omega \rightarrow \mathbb{R}$ on some compact domain $\Omega \subseteq \mathbb{R}^d$ from its function values $f(x_j)$ on a scattered set $X = \{x_1, \dots, x_N\} \subset \Omega \subseteq \mathbb{R}^d$. Independent of how the reconstruction is done in detail, we denote the result as $s_{f,X}$ and assume that it is a linear function of the data, i.e. it takes the form

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

with certain continuous functions $u_j : \Omega \rightarrow \mathbb{R}$. To assert the stability of the recovery process $f \mapsto s_{f,X}$, we look for bounds of the form

$$\|s_{f,X}\|_{L_\infty(\Omega)} \leq C(X)\|f\|_{\ell_\infty(X)} \quad (2)$$

which imply that the map taking the data into the interpolant is continuous in the $L_\infty(\Omega)$ and $\ell_\infty(X)$ norm. Of course, one can also use $L_2(\Omega)$ and $\ell_2(X)$ norms above.

By putting (1) into (2), we see that we can bound the *stability constant* $C(X)$ below as follows

$$C(X) \geq \max_{x \in \Omega} \sum_{j=1}^N |u_j(x)| =: \Lambda_X \quad (3)$$

where Λ_X is the *Lebesgue constant* which is the maximum of the *Lebesgue function* $\lambda_X(x) := \sum_{j=1}^N |u_j(x)|$.

It is a classical problem to derive upper bounds for the stability constant in (2) and for its lower bound, the Lebesgue constant Λ_X . As well-known in recovery by polynomials, in both the univariate and the bivariate case, there exist upper bounds for the Lebesgue function. Moreover, many authors faced the problem of finding near-optimal points for polynomial interpolation. All these near-optimal sets of N points have a Lebesgue function that behaves in the one dimensional case like $\log(N)$ while as $\log^2(N)$ in the two dimensional one (cf. [2] and references therein). An important example, worth mentioning, of points suitable for polynomial interpolation in the square whose Lebesgue constant grows as $\mathcal{O}(\log^2(N))$ are the so-called *Padua-points* (see [1]).

However, stability bounds for multivariate *kernel-based* recovery processes are missing. We shall derive them as follows. Given a positive definite kernel $\Phi : \Omega \times \Omega \rightarrow \mathbb{R}$, the recovery of functions from function values $f(x_j)$ on the set $X = \{x_1, \dots, x_N\} \subset \Omega \subseteq \mathbb{R}^d$ of N different data sites can be done via interpolants of the form

$$s_{f,X} := \sum_{j=1}^N \alpha_j \Phi(\cdot, x_j) \quad (4)$$

taken from the finite-dimensional space $V_X := \text{span} \{\Phi(\cdot, x) : x \in X\}$ of translates of the kernel, and satisfying the linear system

$$A_{\Phi,X} \alpha = \mathbf{f} \quad (5)$$

where $A_{\Phi,X} := (\Phi(x_k, x_j))_{1 \leq j, k \leq N}$ is the *kernel matrix* and \mathbf{f} the vector of length N of data and α the vector of the unknown coefficients. The case of *conditionally* positive definite kernels is similar, and we suppress details here.

The interpolant of (4), as in classical polynomial interpolation, can also be written in terms of *cardinal functions* $u_j \in V_X$ such that $u_j(x_k) = \delta_{j,k}$. Then, the interpolant (4) takes the usual *Lagrangian* form (1).

The reproduction quality of kernel-based methods is governed by the *fill distance* or *mesh norm*

$$h_{X,\Omega} = \sup_{x \in \Omega} \min_{x_j \in X} \|x - x_j\|_2 \quad (6)$$

describing the geometric relation of the set X to the domain Ω . In particular, the reproduction error is small if $h_{X,\Omega}$ is small.

Unfortunately the kernel matrix $A_{\Phi,X}$ is ill-conditioned if the data locations come close, i.e. if the *separation distance*

$$q_X = \frac{1}{2} \min_{\substack{x_i, x_j \in X \\ x_i \neq x_j}} \|x_i - x_j\|. \quad (7)$$

is small. Then the coefficients of the representation (4) get very large even if the data values $f(x_k)$ are small, and simple linear solvers will fail.

As a final introductory element, we recall that the fill distance (6) and the separation distance (7) are two fundamental ingredients for standard error and stability estimates for multivariate interpolants, and they will be also of importance here. The inequality $q_X \leq h_{X,\Omega}$ will hold in most cases, but if points of X nearly coalesce, q_X can be much smaller than $h_{X,\Omega}$, causing instability of the standard solution process. Point sets X are called *quasi-uniform* with *uniformity constant* $\gamma > 1$, if the inequality

$$\frac{1}{\gamma} q_X \leq h_{X,\Omega} \leq \gamma q_X$$

holds. Later, we shall consider arbitrary sets with different cardinalities, but with uniformity constants bounded above by a fixed number. Note that $h_{X,\Omega}$ and q_X play an important role in finding good points for radial basis function interpolation, as recently studied in [9, 3, 5].

2 Main results

To generate interpolants, we allow conditionally positive definite translation-invariant kernels

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

which are reproducing in their “native” Hilbert space \mathcal{N} which we assume to be norm-equivalent to some Sobolev space $W_2^\tau(\Omega)$ with $\tau > d/2$. The kernel will then have a Fourier transform satisfying

$$0 < c(1 + \|\omega\|_2^2)^{-\tau} \leq \hat{K}(\omega) \leq C(1 + \|\omega\|_2^2)^{-\tau} \quad (8)$$

at infinity. This includes, for example, Poisson radial functions (cf. [8, 7]), Sobolev/Matérn kernels and Wendland’s compactly supported kernels (cf. e.g. [13]). It is well-known

that under the above assumptions the interpolation problem is uniquely solvable, and the space V_X is a subspace of Sobolev space $W_2^\tau(\Omega)$.

In what follows, we assume that the constants are dependent on the space dimension, the domain, and the kernel, and the assertions hold for all sets X of scattered data locations with sufficiently small fill distance $h_{X,\Omega}$.

Our main result is the following theorem.

Theorem 1 *The classical Lebesgue constant for interpolation with Φ on $N = |X|$ data locations $X = \{x_1, \dots, x_N\}$ in a bounded domain $\Omega \subseteq \mathbb{R}^d$ satisfying an outer cone condition has a bound of the form*

$$\Lambda_X \leq C\sqrt{N} \left(\frac{h_{X,\Omega}}{q_X} \right)^{\tau-d/2}.$$

For quasi-uniform sets with bounded uniformity γ , this simplifies to

$$\Lambda_X \leq C\sqrt{N}.$$

Each single cardinal function is bounded by

$$\|u_j\|_{L_\infty(\Omega)} \leq C \left(\frac{h_{X,\Omega}}{q_X} \right)^{\tau-d/2}, \quad (9)$$

which, in the quasi-uniform case, simplifies to

$$\|u_j\|_{L_\infty(\Omega)} \leq C. \quad (10)$$

For the L_2 norm,

$$\|u_j\|_{L_2(\Omega)} \leq C \left(\frac{h_{X,\Omega}}{q_X} \right)^{\tau-d/2} h_{X,\Omega}^{d/2} \quad (11)$$

while for quasi-uniform data locations they behave like

$$\|u_j\|_{L_2(\Omega)} \leq Ch_{X,\Omega}^{d/2}. \quad (12)$$

Proof. Let us start by bounding the u_j . Letting $\Psi \in \mathcal{C}^\infty$, having support in the unit ball and such that $\Psi(0) = 1$, $\|\Psi\|_{L_\infty(\Omega)} = 1$ (i.e. a "bump" function). We notice that $|u_j(x)| \leq \left| \Psi \left(\frac{x - x_j}{q_X} \right) \right| + \left| u_j(x) - \Psi \left(\frac{x - x_j}{q_X} \right) \right|$. Since the interpolant $I_X \Psi \left(\frac{\cdot - x_j}{q_X} \right)$ to $\Psi \left(\frac{x - x_j}{q_X} \right)$ on X is u_j , by using standard error estimates (cf. [15, Chap.11]), we get

$$\|u_j\|_{L_\infty(\Omega)} \leq 1 + \left\| I_X \Psi \left(\frac{\cdot - x_j}{q_X} \right) - \Psi \left(\frac{\cdot - x_j}{q_X} \right) \right\|_{L_\infty(\Omega)} \leq 1 + Ch_{X,\Omega}^{\tau-d/2} \left\| \Psi \left(\frac{\cdot}{q_X} \right) \right\|_{\mathcal{N}}. \quad (13)$$

For the L_2 norm, we obtain the inequality

$$\|u_j\|_{L_2(\Omega)} \leq q_X^{d/2} \|\Psi\|_{L_2(\Omega)} + Ch_{X,\Omega}^\tau \left\| \Psi \left(\frac{\cdot}{q_X} \right) \right\|_{\mathcal{N}}. \quad (14)$$

Hence, we simply need to estimate the native space norm of $\Psi(\frac{\cdot}{q_X})$.

$$\begin{aligned} \left\| \Psi \left(\frac{\cdot}{q_X} \right) \right\|_{\mathcal{N}}^2 &\leq C \int \left| q_X^d \hat{\Psi}(q_X \omega) \right|^2 (1 + |\omega|^2)^\tau d\omega \\ &\leq C q_X^d \int \left| \hat{\Psi}(t) \right|^2 \left(1 + \left| \frac{t}{q_X} \right|^2 \right)^\tau dt \\ &\leq C q_X^{d-\tau/2} \int \left| \hat{\Psi}(t) \right|^2 (1 + |t|^2)^\tau dt \leq C_1 q_X^{d-2\tau} \|\Psi\|_{L_2}^2. \end{aligned}$$

Thus, the estimates (9)–(12) easily follow.

Finally we give the claimed bound for the Lebesgue constant. Let $p_{f,X}(x) = \sum_{j=1}^N f(x_j) \Psi \left(\frac{x-x_j}{q_X} \right)$ be the interpolant of the function f to X written in terms of translates of the function Ψ . Then

$$\|I_X p_{f,X}\|_{L_\infty(\Omega)} \leq \|p_{f,X}\|_{L_\infty(\Omega)} + \|I_X p_{f,X} - p_{f,X}\|_{L_\infty(\Omega)}.$$

The first term is bounded by $\|f\|_{\ell_\infty(X)}$, since $p_{f,X}$ is sum of functions with nonoverlapping supports. For the second term, since $p_{f,X} \in \mathcal{N}$ we get

$$\|I_X p_{f,X} - p_{f,X}\|_{L_\infty(\Omega)} \leq Ch_{X,\Omega}^{\tau-d/2} \|p_{f,X}\|_{\mathcal{N}}.$$

Then, it remains to estimate $\|p_{f,X}\|_{\mathcal{N}}$. For $\tau \in \mathbb{N}$, we have

$$\begin{aligned} \|p_{f,X}\|_{\mathcal{N}} &\leq C \left(\sum_{|\alpha| \leq \tau} \|D^\alpha p_{f,X}\|_{L_2}^2 \right)^{1/2} \\ &\leq C \left(\sum_{|\alpha| \leq \tau} \sum_{i=1}^N |f(x_j)|^2 \left\| D^\alpha \Psi \left(\frac{x-x_j}{q_X} \right) \right\|_{L_2}^2 \right)^{1/2} \\ &\leq C \left(\sum_{|\alpha| \leq \tau} \sum_{i=1}^N |f(x_j)|^2 q_X^{d-2\tau} \|D^\alpha \Psi\|_{L_2}^2 \right)^{1/2} \\ &\leq C q_X^{d-2\tau} \|\Psi\|_{W_2^\tau} \left(\sum_{i=1}^N |f(x_j)|^2 \right)^{1/2} \leq C q_X^{d-2\tau} \|\Psi\|_{W_2^\tau} \sqrt{N} \|f\|_{\ell_\infty(X)} \end{aligned}$$

This concludes the proof. \square

But the Lebesgue constants are only upper bounds for the stability constant in function space. In fact, we can do better:

Corollary 2 *Interpolation on sufficiently many quasi-uniformly distributed data is stable in the sense of*

$$\|s_{f,X}\|_{L_\infty(\Omega)} \leq C (\|f\|_{\ell_\infty(X)} + \|f\|_{\ell_2(X)}) \quad (15)$$

and

$$\|s_{f,X}\|_{L_2(\Omega)} \leq Ch_{X,\Omega}^{d/2} \|f\|_{\ell_2(X)} \quad (16)$$

with a constant C independent of X .

Proof. The results easily follow from Theorem 1. \square

Remarks

1. Note that, in the right-hand side of the inequality (16), the ℓ_2 norm is the norm weighted by the cardinality of X , i.e. a properly scaled discrete version of the L_2 norm.
2. The assumption (8) is crucial and, as we shall show below, we are not able to extend the results to kernels with infinite smoothness, such as the Gaussian. The next section will provide examples showing that similar results are not possible for kernels with infinite smoothness.
3. All the previous results can be proved also by using a *sampling inequality* (cf. [16, Th. 2.6]), as shown in the note [4].

3 Examples

We ran a series of examples for uniform grids on $[-1, 1]^2$ and increasing numbers N of data locations. Figure 1 shows the values Λ_X of the Lebesgue constants for the Sobolev/Matern kernel $(r/c)^\nu K_\nu(r/c)$ for $\nu = 1.5$ at scale $c = 20$. In this and other examples for kernels with finite smoothness, one can see that our bounds on the Lebesgue constants are valid, but the experimental Lebesgue constants seem to be uniformly bounded. In all cases, the maximum of the Lebesgue function is attained in the interior of the domain.

Things are different for infinitely smooth kernels. Figure 2 shows the behavior for the Gaussian. The maximum of the Lebesgue function is attained near the corners for large scales, while the behavior in the interior is as stable as for kernels with limited smoothness. The Lebesgue constants do not seem to be uniformly bounded.

A second series of examples was run on 225 regular points in $[-1, 1]^2$ for different kernels at different scales using a parameter c as $\Phi_c(x) = \Phi(x/c)$.

Figures 3 to 5 show how the scaling of the Gaussian kernel influences the shape of the associated Lagrange basis functions. The limit for large scales is called the *flat limit*

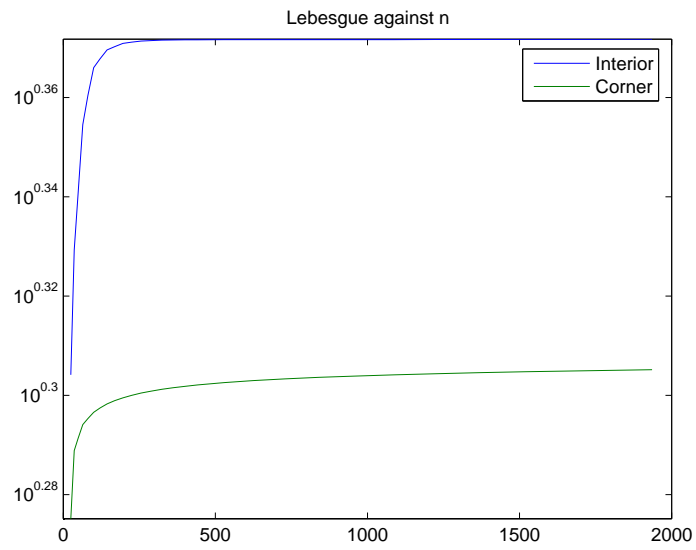


Figure 1: Lebesgue constants for the Sobolev/Matern kernel

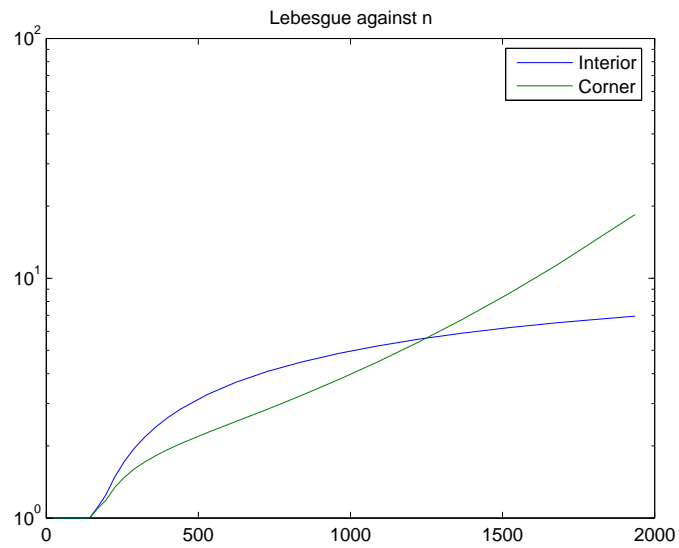


Figure 2: Lebesgue constants for the Gauss kernel

[6] which is a Lagrange basis function of the de Boor/Ron polynomial interpolation [12]. It cannot be expected that such Lagrange basis functions are uniformly bounded.

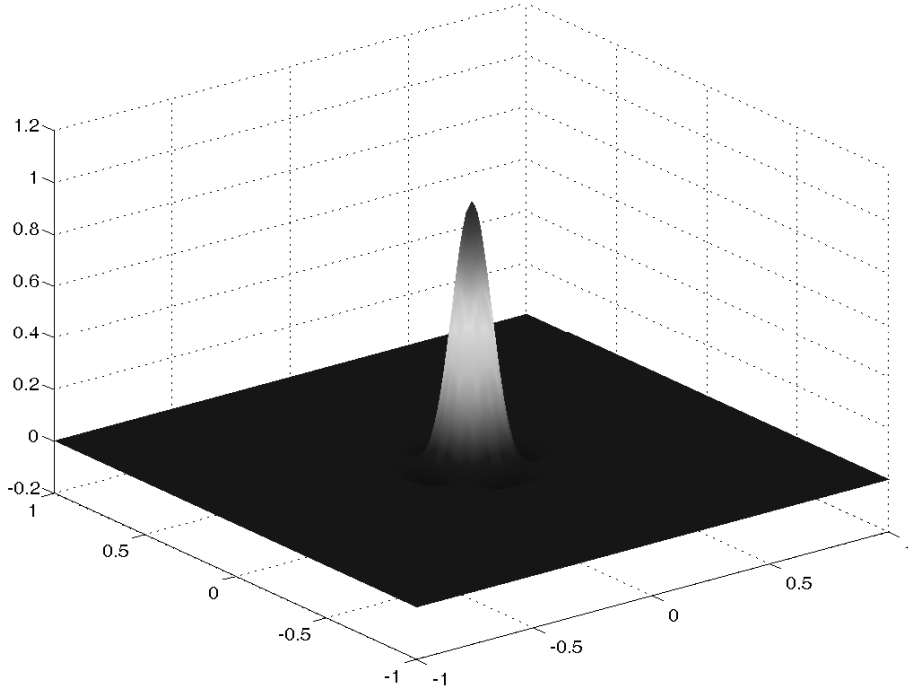


Figure 3: Lagrange basis function on 225 data points, Gaussian kernel with scale 0.1

In contrast to this, Figure 6 shows the corresponding Lagrange basis function for the Sobolev/Matern kernel at scale 320. The scales were such that the conditions of the kernel matrices were unfeasible for the double scale. Figure 7 shows the Lebesgue function in the situation of Figure 5, while Figure 8 shows the Sobolev/Matern case in the situation of Figure 6.

Figures 9 and 10 show how the same Sobolev kernel behaves on scattered data given in Figure 11. The encircled point is where the Lagrange function is taken for Figure 9. Note that the situation does not change dramatically when scattered data are used.

We also checked if the large errors in the corners of the domain in Figure 7 disappeared for domains without corners. Figures 12 and 13 show how the same Gaussian behaves on scattered data on the circle given by Figure 14. It turns out that the boundary behavior is even more dramatic here, since there are no data points on the boundary.

Using Gaussians, with other dilations, did not improve the situation. New results of a forthcoming Ph.D. thesis by Christian Rieger of the University of Göttingen, suggest that an $\mathcal{O}(h^2)$ oversampling in a strip close to the boundary should have a positive effect. To check this indirectly, we used the greedy method of [5] to determine good interpolation points by iteratively adding maxima of the power function. Figures 15 and 16 show the dramatic improvement, while the points are now distributed as in Figure

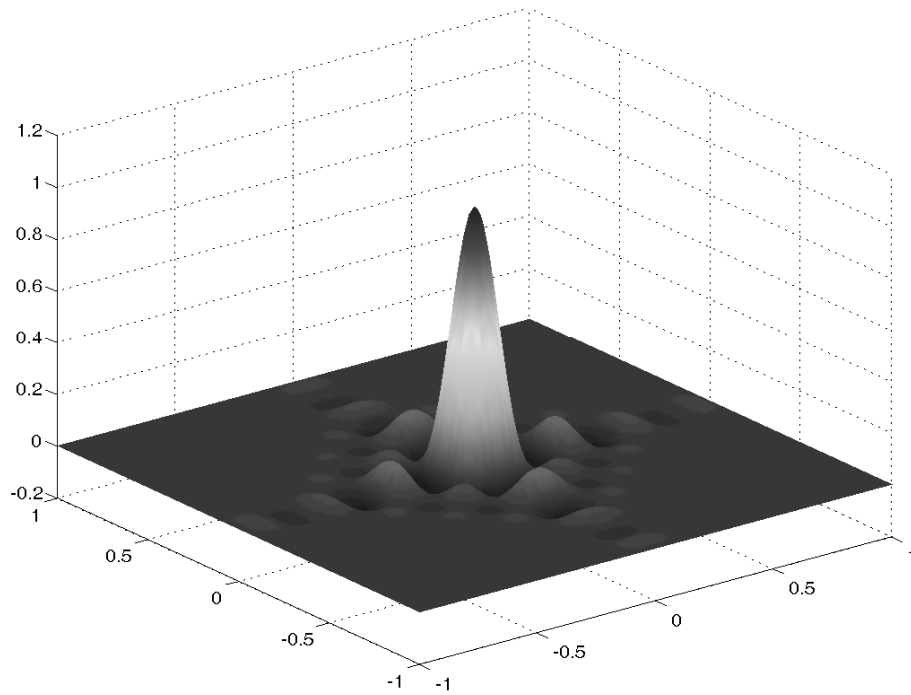


Figure 4: Lagrange basis function on 225 data points, Gaussian kernel with scale 0.2

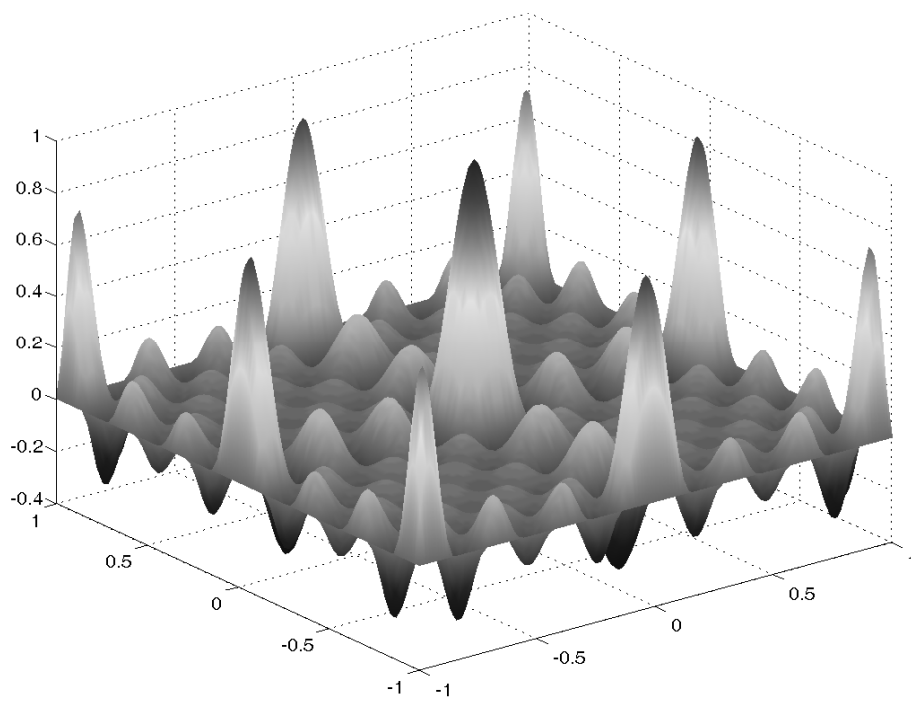


Figure 5: Lagrange basis function on 225 data points, Gaussian kernel with scale 0.4

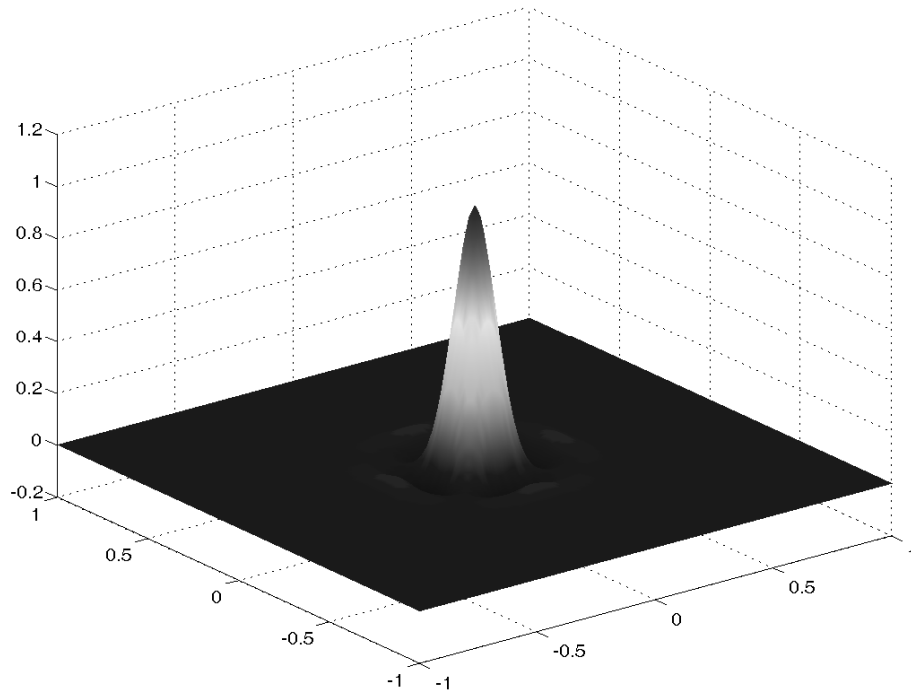


Figure 6: Lagrange basis function on 225 data points, Sobolev/Matern kernel with scale 320

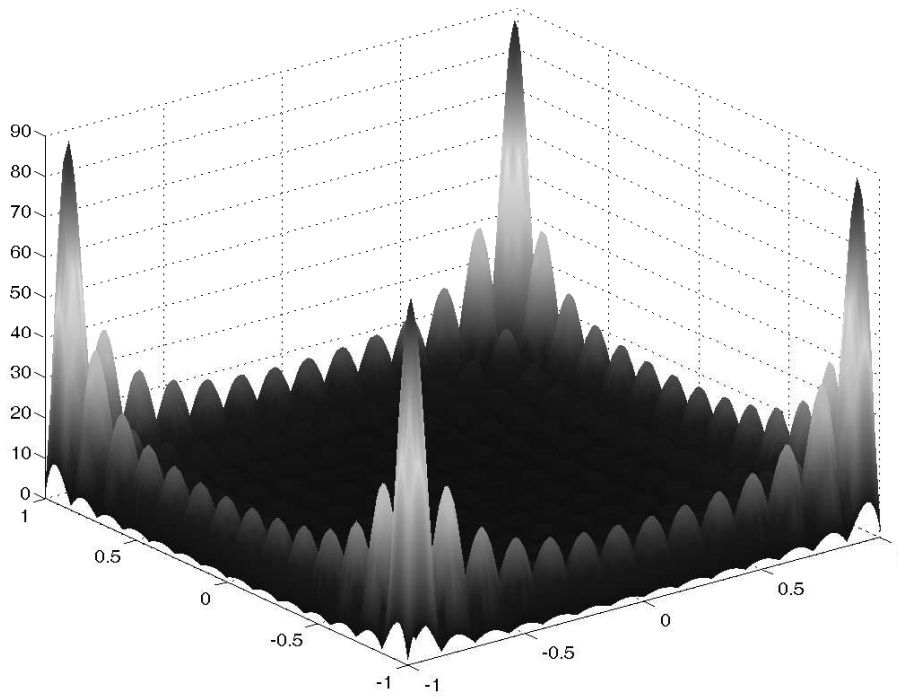


Figure 7: Lebesgue function on 225 regular data points, Gaussian kernel with scale 0.4

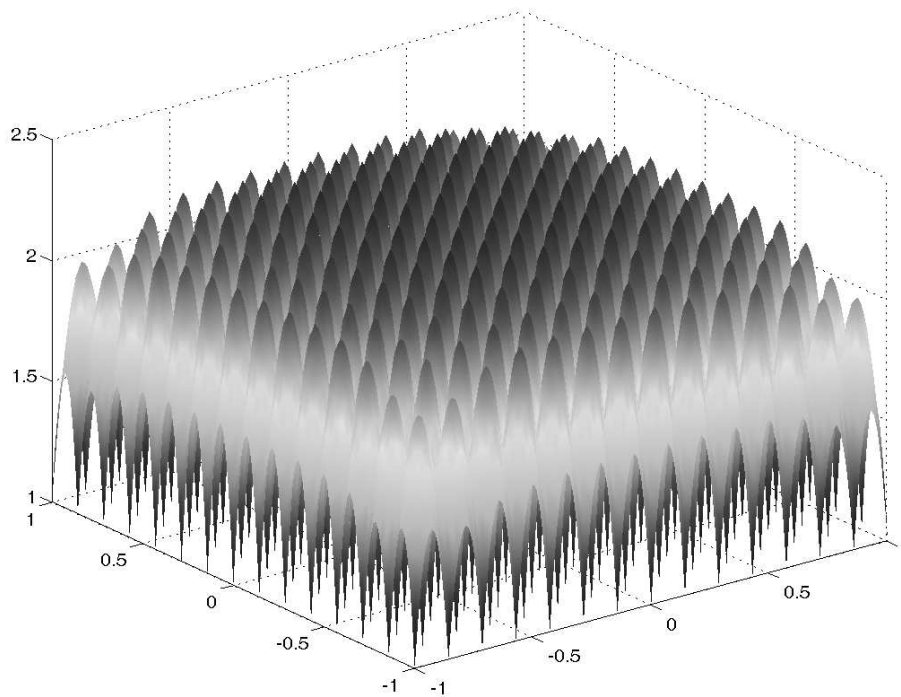


Figure 8: Lebesgue function on 225 regular data points, Sobolev/Matern kernel with scale 320

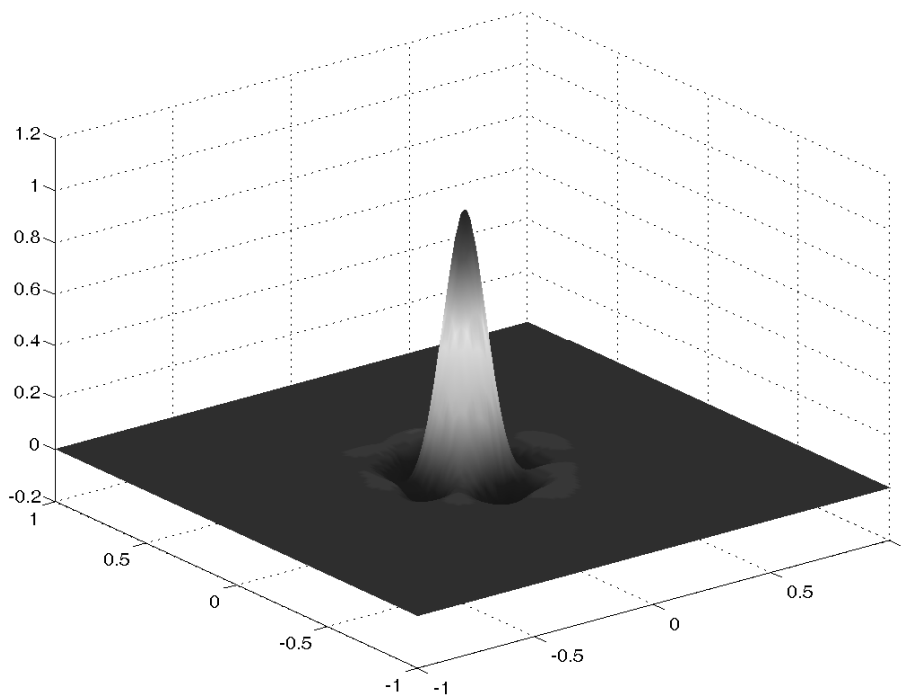


Figure 9: Lagrange basis function on 225 scattered data points, Sobolev/Matern kernel with scale 320

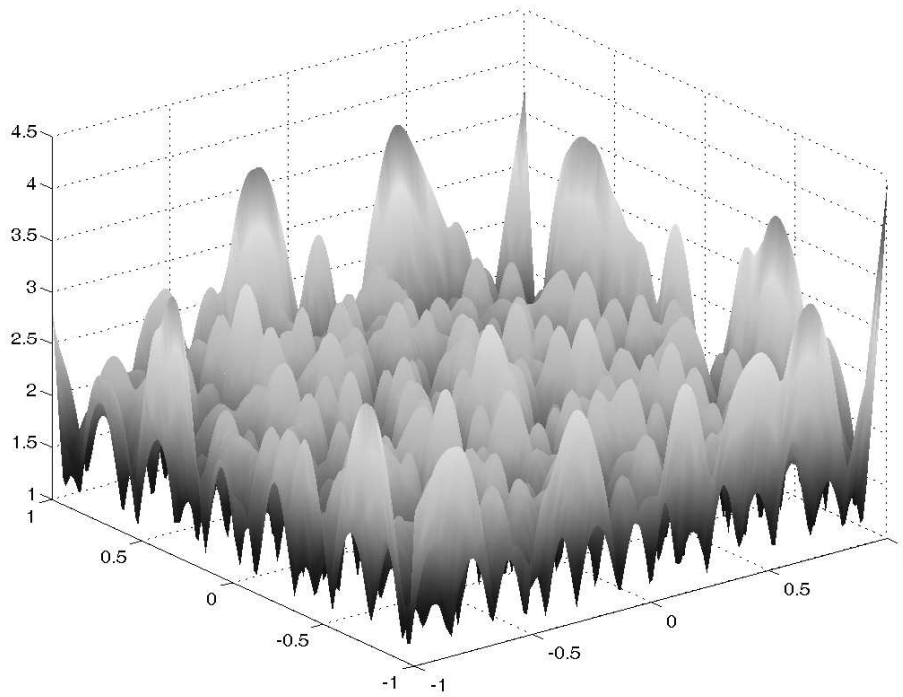


Figure 10: Lebesgue function on 225 scattered data points, Sobolev/Matern kernel with scale 320

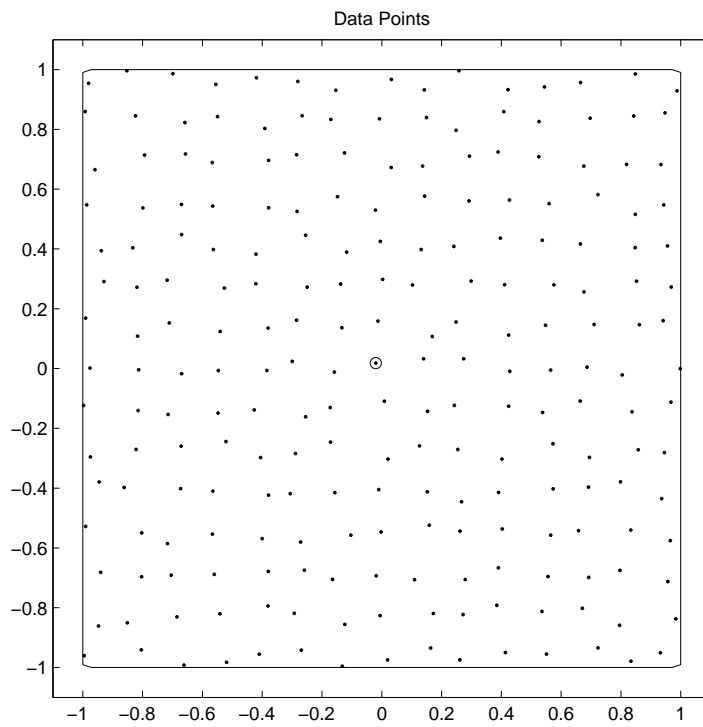


Figure 11: Data points for Figures 9 and 10

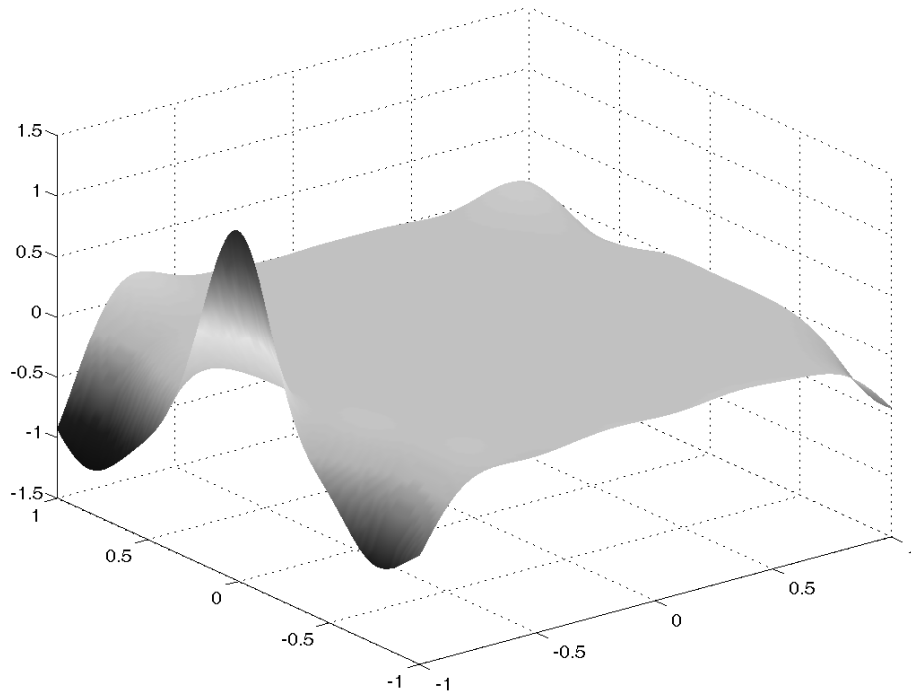


Figure 12: Lagrange basis function for 168 scattered data points on the circle, Gaussian kernel with scale 0.4

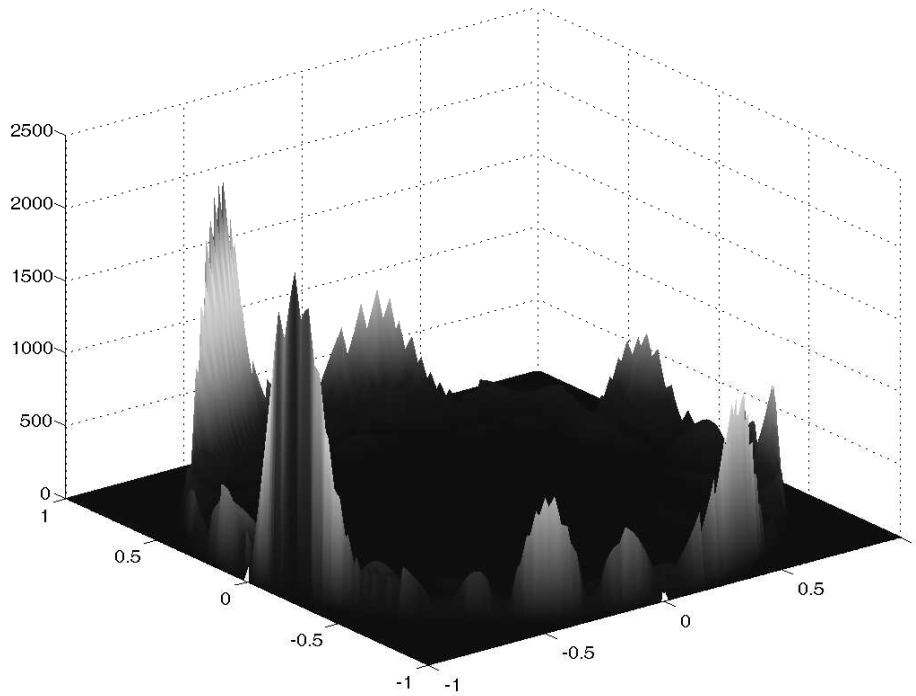


Figure 13: Lebesgue function for 168 scattered data points on the circle, Gaussian kernel with scale 0.4

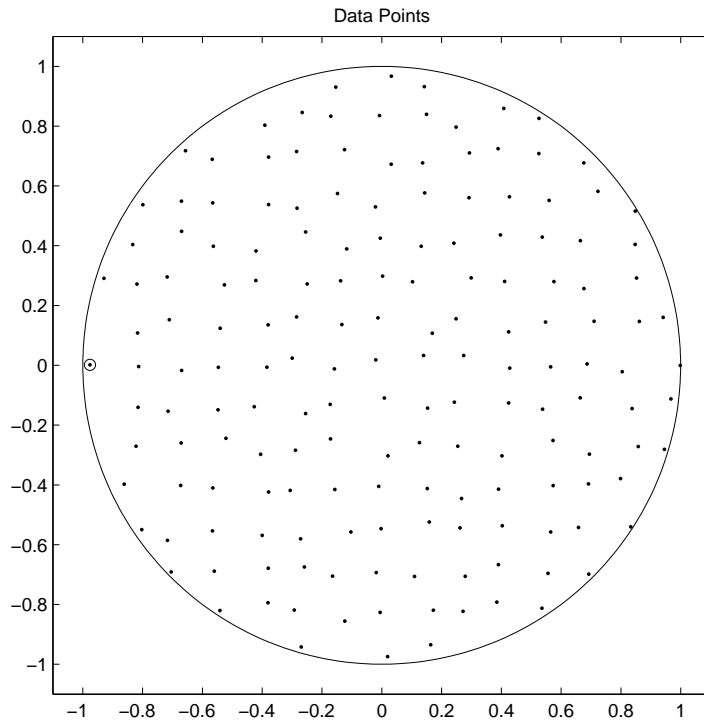


Figure 14: Data points for Figures 12 and 13

17. One could also choose new data points via the maximum of the Lebesgue function, but this strategy turned out to be inferior.

We also ran some other examples on a cardioidal domain with an incoming cusp, but the results were not much different.

The improvement by oversampling on the boundary seems to be connected to analytic kernels, since the corresponding examples for non-smooth kernels showed a much weaker effect. We add figures for the C^2 Wendland function, but we remark that Matern/Sobolev kernels behave similarly. Note that all functions are chopped at the boundary of the cardioid.

Acknowledgments. This work has been supported by the CNR-DFG exchange programme, 2006 and by the GNCS-Indam funds 2007. The authors thank the referees for some useful suggestions, including different, but only slightly shorter proofs for Theorem 1.

References

- [1] Bos, L., Caliari, M., De Marchi, S., Vianello, M. and Xu, Y., *Bivariate Lagrange interpolation at the Padua points: the generating curve approach*, J. Approx. Theory,

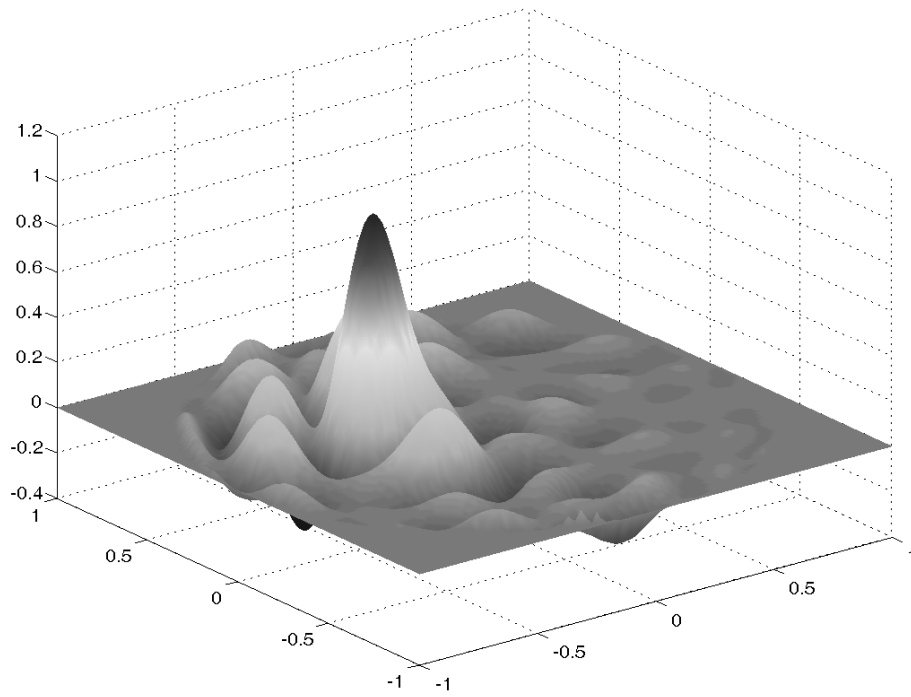


Figure 15: Lagrange basis function for 168 optimized data points on the circle, Gaussian kernel with scale 0.4

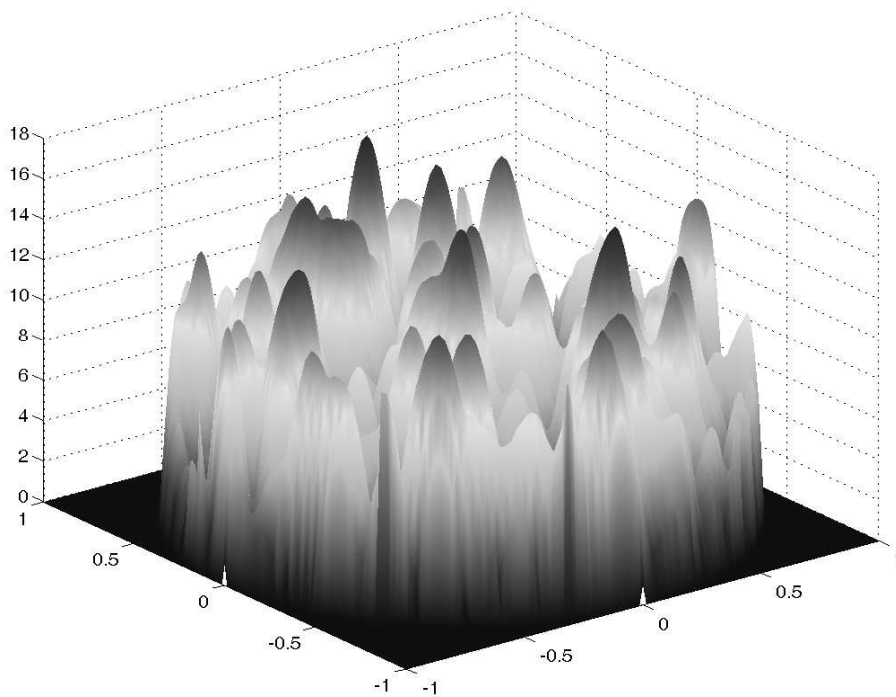


Figure 16: Lebesgue function for 168 optimized data points on the circle, Gaussian kernel with scale 0.4

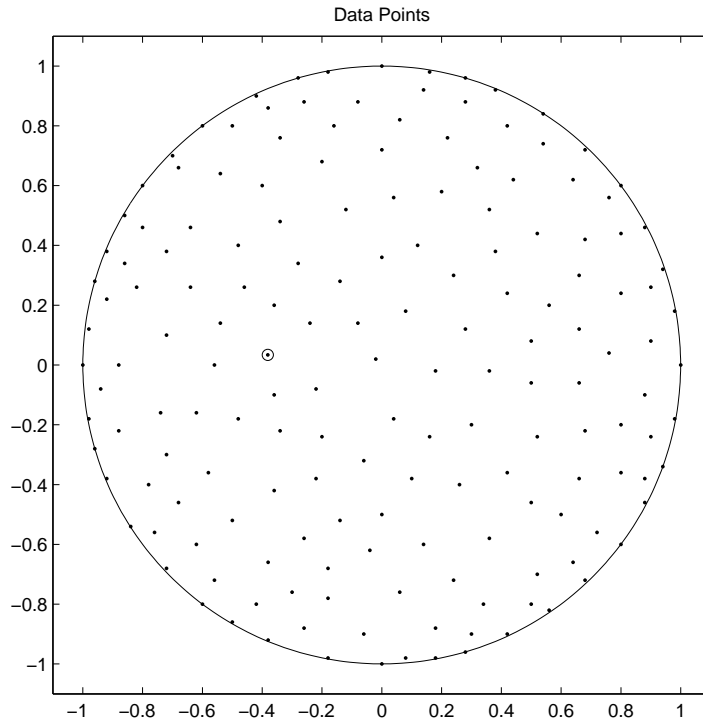


Figure 17: Data points for Figures 15 and 16

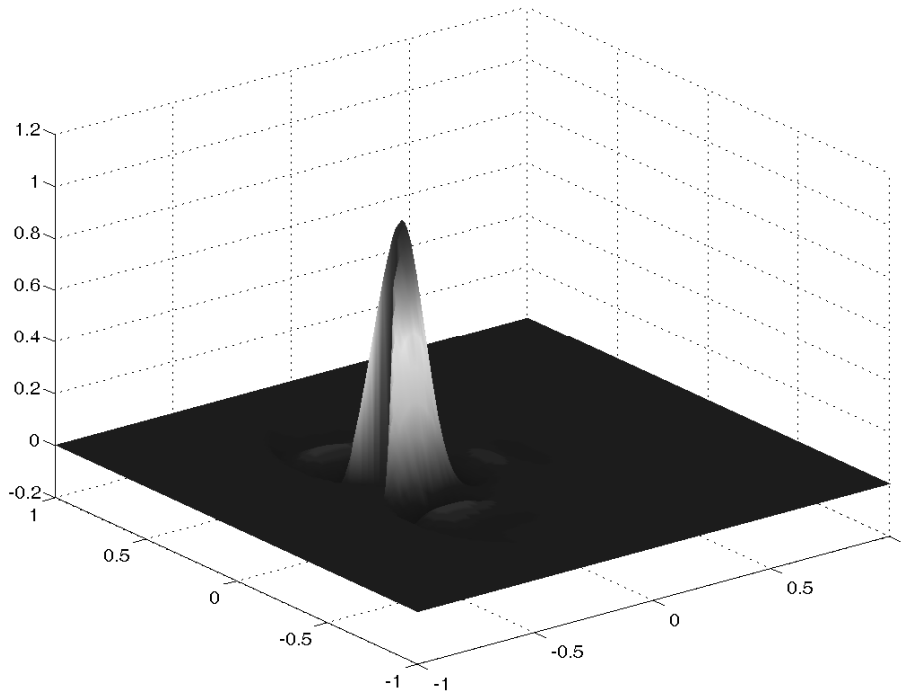


Figure 18: Lagrange basis function for 104 scattered data points on the cardioid, Wendland C^2 kernel with scale 30

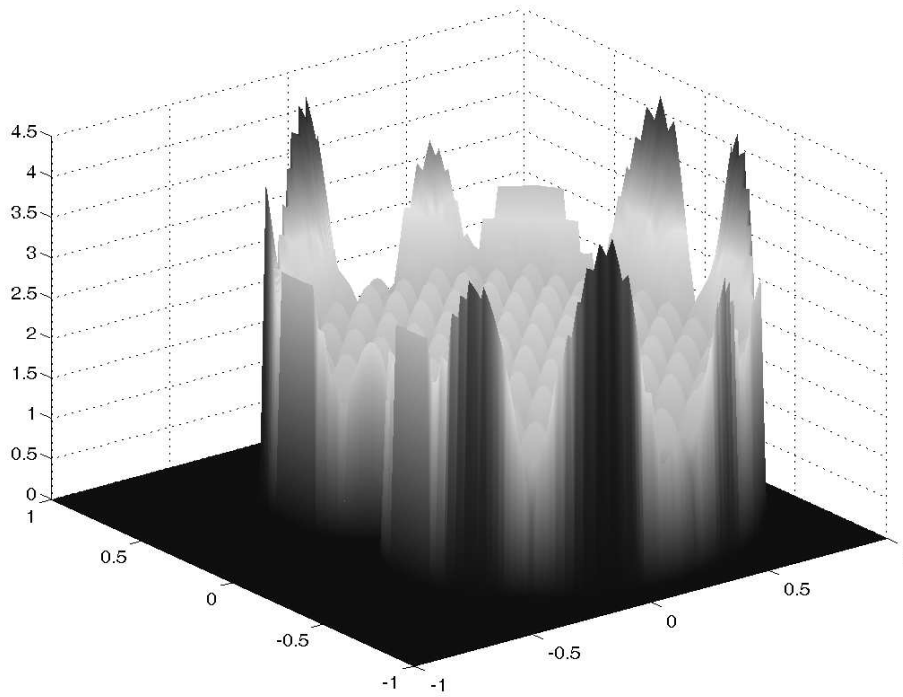


Figure 19: Lebesgue function for 104 scattered data points on the cardioid, Wendland C^2 kernel with scale 30

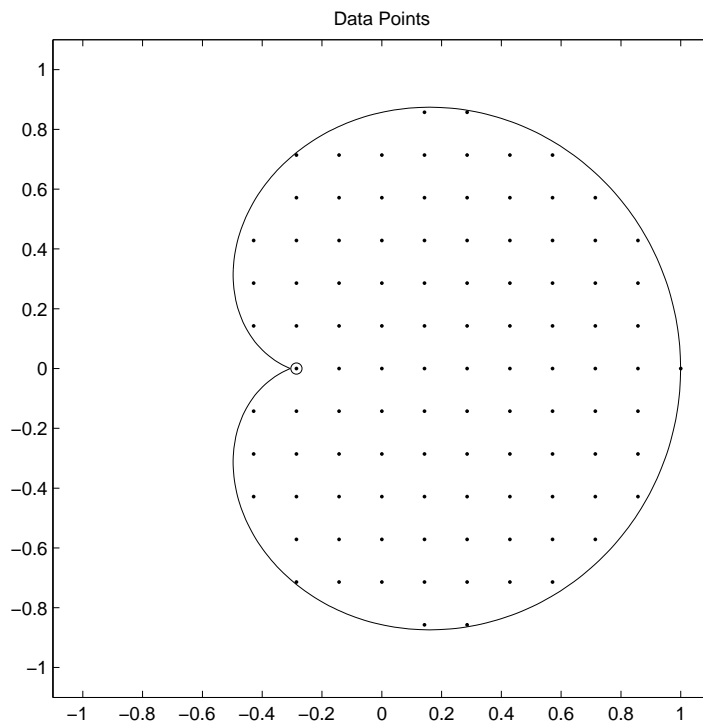


Figure 20: Data points for Figures 18 and 19

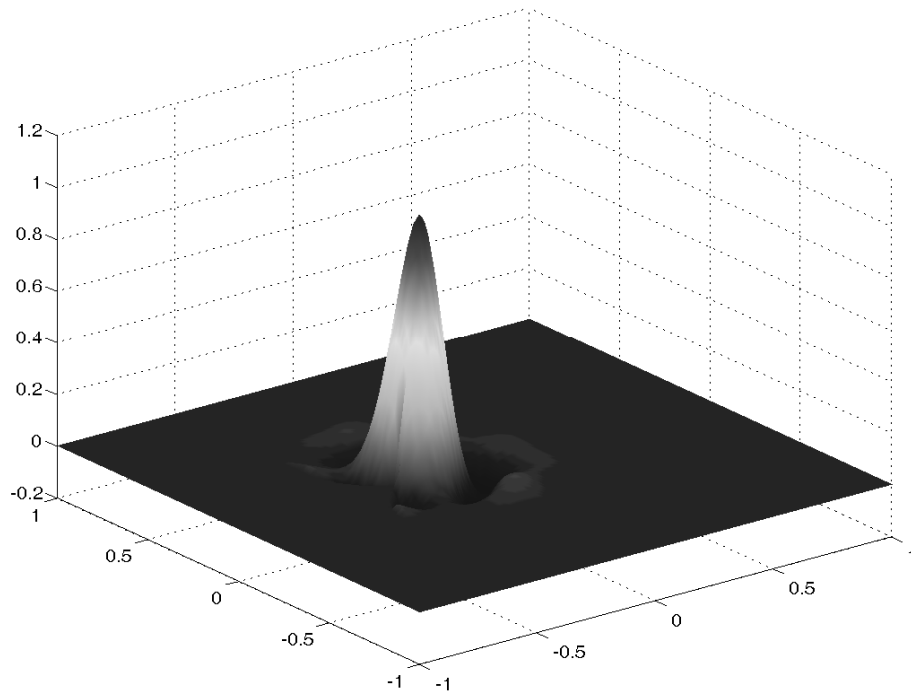


Figure 21: Lagrange basis function for 104 optimized data points on the cardioid, Wendland C^2 kernel with scale 30

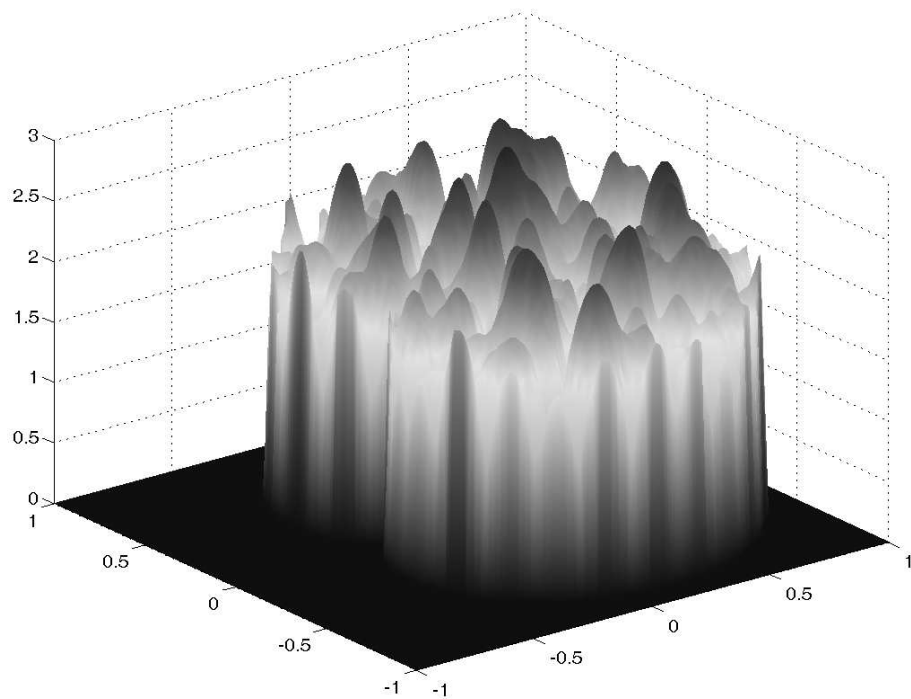


Figure 22: Lebesgue function for 104 optimized data points on the cardioid, Wendland C^2 kernel with scale 30

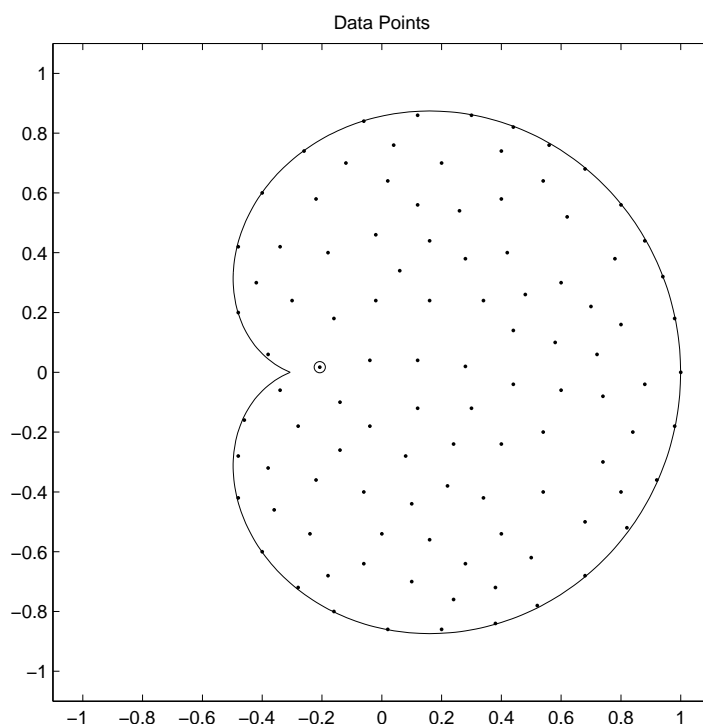


Figure 23: Data points for Figures 21 and 22

Vol. 143 (2006), 15-25.

- [2] Caliari, M., De Marchi, S. and Vianello, M. *Bivariate polynomial interpolation on the square at new nodal sets*, Appl. Math. Comput., Vol. 165(2) (2005), 261-274.
- [3] De Marchi, S. *On optimal center locations for radial basis interpolation: computational aspects*, Rend. Sem. Mat. Torino, Vol. 61(3)(2003), 343-358.
- [4] De Marchi, S. and Schaback, R. *Stability constants for kernel-based interpolation processes*, R.R. 59/08, Department of Computer Science, University of Verona, 2008.
- [5] De Marchi, S., Schaback, R. and Wendland, H. *Near-Optimal Data-independent Point Locations for Radial Basis Function Interpolation*, Adv. Comput. Math., Vol. 23(3) (2005), 317-330.
- [6] Driscoll, T.A. and Fornberg, B. *Interpolation in the limit of increasingly flat radial basis functions*, Comput. Math. Appl., Vol. 43 (2002), 413-422.
- [7] Flyer, N. *Exact polynomial reproduction for oscillatory radial basis functions on infinite lattices*, Comput. Math. Appl., Vol. 51(8) (2004), 1199-1208.
- [8] Fornberg, B., Larsson, E. and Wright, G. *A new class of oscillatory radial basis functions*, Comput. Math. Appl., Vol. 51(8) (2004), 1209-1222.

- [9] Iske, A. *Optimal distribution of centers for radial basis function methods*, Tech. Rep. M0004, Technische Universität München, 2000.
- [10] W. R. Madych, *An estimate for Multivariate interpolation II*, to appear in J. Approx. Theory **142(2)** (2006), 116–128.
- [11] R. Schaback, *Lower bounds for norms of inverses of interpolation matrices for radial basis functions*, J. Approx. Theory **79** (1994), 287–306.
- [12] R. Schaback, *Multivariate interpolation by polynomials and radial basis functions*, Constr. Approx. **21** (2005), 293–317.
- [13] R. Schaback, H. Wendland, *Characterization and construction of radial basis function*, in *Multivariate Approximation and Applications* (N. Dyn, D. Leviatan, D. Levin, and A. Pinkus, eds.), Cambridge University Press (2001), 1-24.
- [14] R. Schaback, H. Wendland, *Inverse and saturation theorems for radial basis function interpolation*, Mat. Comp. **71** (2002), 669-681.
- [15] H. Wendland, *Scattered Data Approximation*, *Cambridge Monographs on Applied and Computational Mathematics*, Vol. 17, 2005.
- [16] H. Wendland, C. Rieger, *Approximate interpolation with applications to selecting smoothing parameters*, Numer. Math. **101** (2005), 643-662.