Multistep scattered data interpolation using compactly supported radial basis functions

Michael S. Floater, Armin Iske*, [1]

SINTEF, Postbox 124, Blindern, N-0314 Oslo, Norway

Received 27 September 1995; revised 23 November 1995

Abstract

A hierarchical scheme is presented for smoothly interpolating scattered data with radial basis functions of compact support. A nested sequence of subsets of the data is computed efficiently using successive Delaunay triangulations. The scale of the basis function at each level is determined from the current density of the points using information from the triangulation. The method is rotationally invariant and has good reproduction properties. Moreover the solution can be calculated and evaluated in acceptable computing time.

Keywords: Hierarchical interpolation; Scattered data; Compactly supported radial basis functions

1. Introduction

During the last two decades radial basis functions have become a well established tool for multivariate interpolation of both scattered and gridded data; see [2, 7, 8, 21, 25] for some surveys. The major part

$$s(x) = \sum_{j=1}^{N} c_j \phi(\|x - x_j\|) \tag{1}$$

of the interpolant is a linear combination of translates of a prescribed continuous *basis function* $\Phi : \mathbb{R}^d \to \mathbb{R}$, $\Phi(x) = \phi(\|x\|)$, which is radial with respect to the Euclidean norm $\|\cdot\|$. Prominent examples are Duchon's thin plate splines $\phi(r) = r^2 \log(r)$ [4–6], Hardy's multiquadrics $\phi(r) = (c^2 + r^2)^{1/2}$ [13] and inverse multiquadrics $\phi(r) = (c^2 + r^2)^{-1/2}$, as well as the Gaussians

---

* Corresponding author. E-mail: armin.iske@si.sintef.no.

$\phi(r) = e^{-cr^2}$. It has been shown that interpolation based on these particular functions which we refer to as *classical radial basis functions* provide good approximation behaviour [14, 15, 32] and this confirms earlier observations in [11]. Nevertheless, although preconditioning methods are available [9], one has to face the fact that for large data sets interpolation using these classical functions requires an unacceptable amount of computation time.

In this paper we bypass this problem by restricting ϕ to have compact support and we consider interpolants purely of the form (1). Now the interpolation problem usually requires matching a given function $f: \mathbb{R}^d \to \mathbb{R}$ on a given set $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^d$ of pairwise distinct centres with respect to conditions

$$s(x_k) = f(x_k), \quad 1 \leqslant k \leqslant N. \tag{2}$$

The well-posedness of this problem depends on the nonsingularity of the interpolation matrix

$$A_{X, \phi} = (\phi(\|x_j - x_k\|))_{1 \leqslant j, k \leqslant N}.$$

Following Micchelli [16] as a sufficient condition we require that ϕ is *positive definite* on $\mathbb{R}^d$.

Definition 1. A continuous function $\phi: [0, \infty) \to \mathbb{R}$ is said to be positive definite on $\mathbb{R}^d$, $\phi \in \mathrm{PD}_d$, iff the quadratic form $c^T A_{X, \phi} c$ is strictly positive for all possible choices of $X = \{x_1, \dots, x_N\}$ and $c = (c_1, \dots, c_N) \in \mathbb{R}^N \setminus \{0\}$.

The positive definiteness of ϕ guarantees that all possible interpolation problems (2) possess a unique solution and this then justifies referring to ϕ as a basis function.

Smooth compactly supported radial basis functions have only very recently been constructed. To the best of our knowledge the first instances of such functions were provided in [27] and the details of their construction can be found in Schaback's survey paper [25]. Further developments were provided in [30, 31] who came up with a broad variety of piecewise polynomial compactly supported radial basis functions. Very recently, Wendland [29] constructed further instances of such functions. Moreover it was proven in [29] that for a specific space dimension d, these functions possess the lowest possible degree among all piecewise polynomial compactly supported radial functions which are positive definite on $\mathbb{R}^d$ and of a given order of smoothness. We find this a useful property in practice and thus provide a selection of Wendland's functions indicating the order of continuity and the relevant space dimension. Note that if $\phi \in \mathrm{PD}_d$ then $\phi \in \mathrm{PD}_k$ whenever $k \leqslant d$.

$$\phi = (1 - r)_+ \qquad\qquad C^0 \cap \mathrm{PD}_1,$$

$$\phi = (1 - r)_+^3 (3r + 1) \qquad\qquad C^2 \cap \mathrm{PD}_1,$$

$$\phi = (1 - r)_+^5 (8r^2 + 5r + 1) \qquad C^4 \cap \mathrm{PD}_1,$$

$$\phi = (1 - r)_+^2 \qquad\qquad C^0 \cap \mathrm{PD}_3,$$

$$\phi = (1 - r)_+^4 (4r + 1) \qquad\qquad C^2 \cap \mathrm{PD}_3,$$

$$\phi = (1 - r)_+^6 (35r^2 + 18r + 3) \quad C^4 \cap \mathrm{PD}_3.$$

In order to adapt the interpolation to scattered data of different densities it is clearly necessary to be able to scale the support of ϕ. So from now on we assume that the radius of support of ϕ is one and replace ϕ by

$$\phi_\alpha(\cdot) := \phi(\cdot/\alpha) \quad \text{for } \alpha > 0,$$

in (1). The radius of support of ϕ_α is then α. Given a set X of scattered centres one ends up with a symmetric interpolation matrix A_{X,ϕ_α} which is not only positive definite but also sparse provided that the scale α is adjusted to the density of X. One can then exploit this by applying an iterative method such as the conjugate gradient method [12] for the inversion of A_{X,ϕ_α}.

Unfortunately, the distribution of the centres can vary considerably, and this makes finding the right scale awkward. If α is too low, the approximation will be poor; the surface may exhibit small peaks and flat spots. On the other hand if α is too large, A_{X,ϕ_α} is no longer sparse enough to make its inversion efficient which is the main reason for choosing compactly supported functions.

Therefore, in this paper we present a hierarchical method which starts with a decomposition of X into a nested sequence

$$X_1 \subset X_2 \subset \cdots \subset X_{M-1} \subset X_M = X \tag{3}$$

M subsets

$$X_k = \{x_1^{(k)}, \ldots, x_{N_k}^{(k)}\} \subset X, \quad 1 \leqslant k \leqslant M.$$

This allows the interpolation to be broken up into M steps. The decomposition strategy makes use of Delaunay triangulations [22], and is designed so that the density of the points in each X_k is as uniform as possible and increases smoothly as k increases. The density (or rather sparsity) of X_k can be measured in various ways, but we concentrate on two which between them nicely capture this concept. The first is the *separation distance*

$$q(X_k) := \min_{1 \leqslant l < m \leqslant N_k} |x_l^{(k)} - x_m^{(k)}|/2,$$

which is half the distance between the closest pair of points in X_k. The second is the *radius of the largest inner empty sphere*

$$Q(X_k) := \max_{x \in \Omega} \min_{1 \leqslant j \leqslant N_k} |x - x_j^{(k)}|,$$

where Ω is some fixed compact region in $\mathbb{R}^d$ containing the original set X. For convenience we will assume that Ω is the closed interior of some polygon $\partial\Omega$ surrounding X which could for example be its convex hull. The measures q and Q help to provide a data-dependent criterion for choosing the scale of the basis function at each level.

We have found that regarding both computation time and approximation behaviour, the hierarchical method is able to handle small and large data sets, say $N = 2000$ or more, very well. In contrast (depending on the underlying hardware facilities), interpolation using globally supported radial basis functions is often unfeasible once N is beyond 100 or 200.

We believe therefore that the method presented in this paper will be useful in practice for interpolating general scattered data sets. The implementation only requires a routine for matrix inversion (preferably an iterative method) and code for creating and manipulating triangulations.

2. Multistep interpolation

The use of multistep techniques (also referred to as multistage or hierarchical methods) is common practice for scattered data analysis [1]. We now describe such a method in our present context. One considers a nested sequence (3) of subsets X_k of X and decomposes the interpolation of X into M steps. Starting with $k = 1$, at the kth step one matches the *error function*

$$f - (s_1 + \cdots + s_{k-1})$$

on X_k by computing the coefficients of the kth interpolant

$$s_k(x) = \sum_{j=1}^{N_k} c_j^{(k)} \phi_{\alpha_k}(\| x - x_j^{(k)} \|), \tag{4}$$

after the scale $\alpha_k > 0$ of the basis function has been specified. To be more precise, the corresponding interpolation scheme has the form

$$s_1|_{X_1} = f|_{X_1},$$

$$s_2|_{X_2} = (f - s_1)|_{X_2},$$

$$\vdots \qquad \vdots$$

$$s_M|_{X_M} = \left(f - \sum_{k=1}^{M-1} s_k \right)\bigg|_{X_M}.$$

From this one obtains the identity

$$(s_1 + \cdots + s_M)|_X = f|_X.$$

which guarantees that $s = s_1 + \cdots + s_M$ matches f at each of the centres in X. The basic idea is to set α_1 relatively large, and to let the α_k decrease as k increases. The main philosophy behind this strategy is that the basic trends and characteristics of the graph of f are captured in the first few steps, while its finer details emerge in the last ones.

Although this iterative interpolation process is easy to understand and straightforward to implement, there remain two tasks:
– setting up a decomposition of X into subsets X_k,
– making a reasonable choice for the scales α_k.

Determining the X_k and α_k is critical to the success of the method, especially in view of computation time and the quality of the interpolant s, and these will be the subjects of the next two sections.

3. Thinning algorithm

We now describe a thinning algorithm which generates from the given scattered data set X a nested sequence of subsets in such a way that the points in each subset are distributed as evenly as possible and their density increases smoothly. We shall restrict the space dimension d to two, although in principle the same algorithm could be used in arbitrary space dimensions.

The algorithm works backwards. Starting with $Y_N = X$, it recursively removes points from it, one by one, until there are only a small number K left. Some suitable subsequence of the resulting subsets

$$Y_K \subset Y_{K+1} \subset \cdots \subset Y_N = X$$

can then be chosen for the hierarchical interpolation. The criterion for point removal is that the separation distance $q(Y_{i-1})$ should if possible be greater than $q(Y_i)$. This implies that one has to locate the closest pair of points and remove one of them.

In order to study the efficacy of the algorithm with regard to the uniformity of the points in each subset, it will be useful to also keep in mind $Q(Y_i)$ as defined in Section 2. Now if $q(Y_i) = \| y_1 - y_2 \|/2$ for some $y_1, y_2 \in Y_i$ then the circle having the line segment $\overline{y_1 y_2}$ as a diameter cannot contain any other points in Y_i, and it follows that $q(Y_i) \leqslant Q(Y_i)$. If we let $\rho = q/Q$, then $0 < \rho \leqslant 1$ and we say that Y_i has *uniformity* $\rho = \rho(Y_i)$. An example of a data set with high uniformity is the set of nodes in a triangular grid made of equilateral triangles. Resorting to trigonometry, one finds that $\rho = \sqrt{3}/2$. A regular square grid has uniformity $1/\sqrt{2}$.

The closest pair and largest empty circle of a discrete set Y of points in the plane are well-known concepts in computational geometry [22], and they can both be computed relatively easily from the Voronoi diagram $\mathscr{V}_Y$ of Y. It is shown in [22] that the closest pair of points in Y are Voronoi neighbours in $\mathscr{V}_Y$. It also follows from observations in [22] that the centre of the largest inner empty circle must be either a node of $\mathscr{V}_Y$, or an intersection between an edge of $\mathscr{V}_Y$ and $\partial\Omega$, or a vertex of $\partial\Omega$. This is due to the fact that the function $\min_{y \in Y} | \cdot - y |$ is convex in each Voronoi polygon.

In practice, however, we find it easier to make computations with the Delaunay triangulation $\mathscr{T}_Y$ of the points in Y, which as a planar graph is the dual of $\mathscr{V}_Y$ [22, 28]. The closest pair are neighbouring nodes in $\mathscr{T}_Y$ from which one computes $q(Y)$. Regarding $Q(Y)$, observe that each node of $\mathscr{V}_Y$ is the circumcentre of a triangle in $\mathscr{T}_Y$ and the corresponding circumcircle contains no other points in Y (the Delaunay criterion). Thus, by ignoring circles whose centres lie on $\partial\Omega$, we obtain a very close approximation to $Q(Y)$ by computing the radius of the largest circumcircle of the triangles in $\mathscr{T}_Y$ whose centres lie in Ω. Since this is considerably easier to calculate in practice, we have approximated $Q(Y)$, in all of our numerical examples, by

$$\max_{\substack{c(T) \in \Omega \\ T \in \mathscr{T}_1}} \min_{y \in Y} \| c(T) - y \|, \tag{5}$$

where $c(T)$ denotes the circumcentre of triangle T. Note that the quantity in (5) is bounded above by $Q(Y)$. Their difference is generally small and only depends on the distribution of points near the boundary of Ω.

Bearing these facts in mind, we compute at each step of the thinning algorithm the Delaunay triangulation of the current subset. Now let $\mathscr{T}$ be a Delaunay triangulation. Those nodes in $\mathscr{T}$ which lie on the boundary will be called *boundary nodes* while the rest will be called *interior nodes*. Boundary nodes are treated a little differently in order to prevent boundary erosion in the thinning algorithm.

Definition 2. Suppose the neighbouring nodes (interior or boundary) of any interior node $x \in \mathscr{T}$ are $x_1, \ldots, x_k$. Then define

$$d_{\min}(x) = \min_{i=1,\ldots,k} \| x_i - x \|, \qquad d_{\max}(x) = \max_{i=1,\ldots,k} \| x_i - x \|.$$
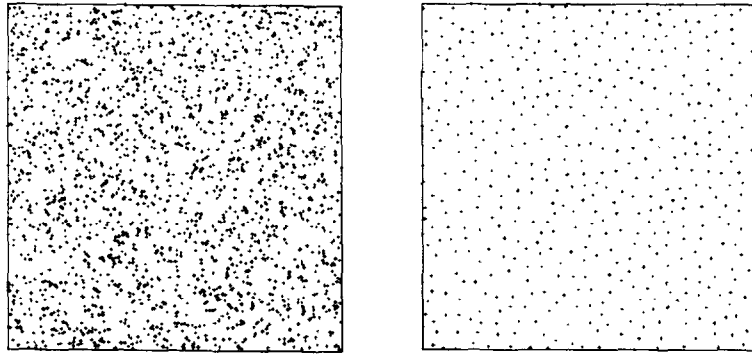
Figs. 1 and 2. Original data set $X = X_4 = Y_{2000}$ and $X_3 = Y_{500}$.
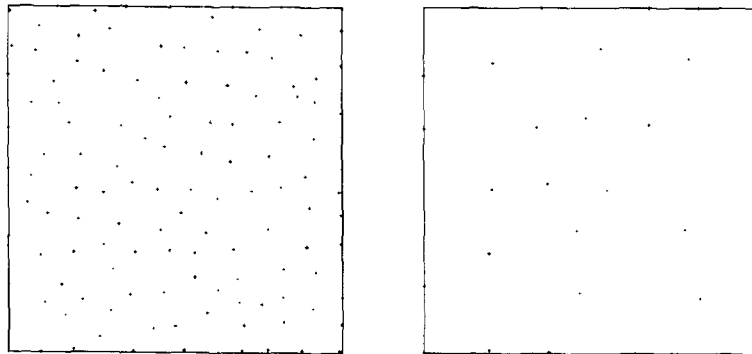
Figs. 3 and 4. $X_2 = Y_{125}$ and $X_1 = Y_{31}$.

If $x \in \mathcal{T}$ is a boundary node it has precisely two boundary neighbours x_1, x_2. In this case define

$$d_{\min}(x) = \min_{i=1,2} \| x_i - x \|, \qquad d_{\max}(x) = \max_{i=1,2} \| x_i - x \|.$$

**Definition 3.** We say that a node $x \in \mathcal{T}$ is removable if
  (i) $d_{\min}(x) \leqslant d_{\min}(y)$ for all $y \in \mathcal{T}$, and
  (ii) $d_{\max}(x) \leqslant d_{\max}(y)$ whenever $d_{\min}(x) = d_{\min}(y)$.

Notice that if any closest pair of points are either both boundary or both interior nodes, then $d_{\min}$ will attain its minimum value at both of these points. Thus in most cases both criterions (i) and (ii) will be used to choose the point for removal. However, there still may be more than one removable point (this will certainly be the case when the nodes form a regular grid).

Now recall that $X = \{x_1, \ldots, x_N\}$. Choose some small $K \geqslant 3$ with $K \ll N$.

Thinning algorithm.
  (1) *Let* $Y_N = X$ *and compute* $\mathcal{T}_N = \mathcal{T}_{Y_N}$.
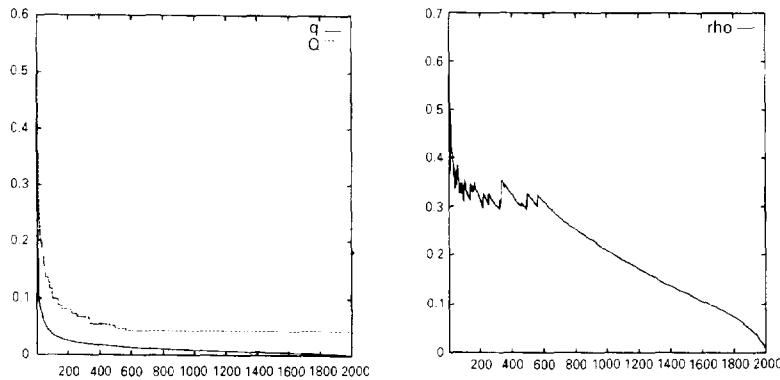
Figs. 5 and 6. q, Q and their ratio ρ at each step.

(2) *For decreasing* $i = N, \ldots, K + 1$:
   (2a) *Choose any removable node* $x \in \mathcal{T}_i$,
   (2b) *Let* $y_i = x$ *and* $Y_{i-1} = Y_i \setminus \{x\}$,
   (2c) *Compute* $\mathcal{T}_{i-1} = \mathcal{T}_{Y_{i-1}}$ *from* $\mathcal{T}_i$.
(3) *For* $i = 1, \ldots, K$, *let* $y_i = x_i$ *where* $(x_1, \ldots, x_K)$ *is any ordering of* Y_K.

Since $\mathcal{T}_{i-1}$ differs from $\mathcal{T}_i$ in only a few edges in the vicinity of the deleted node, $\mathcal{T}_{i-1}$ can be computed quickly from the other. The same property has been exploited before in the context of point insertion in Delaunay triangulations [3].

When the algorithm has finished, the sequence $y_1, \ldots, y_N$ is an ordering of the points in X. Indeed it merely remains to choose M and $N_2, \ldots, N_{M-1}$ such that $K = N_1 < N_2 < \cdots < N_M = N$ and let $X_k = Y_{N_k} = \{y_1, \ldots, y_{N_k}\}$. The sequence $(X_k)_k$ will then be the desired decomposition. In analogy with hierarchical methods based on regular rectangular grids we propose choosing $(N_k)_k$ to be a geometric sequence and we will come back to this in Section 5. Figs. 1–4 show an example where $X (= Y_{2000})$ contains 2000 randomly generated points in the unit square, while the subsets Y_{31}, Y_{125}, and Y_{500} were generated by the thinning algorithm.

If we calculate and store in Step 2, $q(i) = q(Y_i)$ and $Q(i) = Q(Y_i)$, for $i \in \{K, \ldots, N\}$, this information can be used to decide on the scales α_k in (4). Figs. 5 and 6 show the graphs of $q(i)$, $Q(i)$ and $\rho(i) = q(i)/Q(i)$ for $i = 3, \ldots, N$.

4. Scale of the support

In this section we concentrate on the choice of the scale α_k of the support of the basis function at the kth step of the method described in Section 2.

The discussion depends on the set X_k, the separation distance $q(X_k)$ and the radius $Q(X_k)$ of the largest inner empty sphere. For reasons of notational simplicity, we shall for the moment drop the index k, so that $X = X_k$, $q = q_k$, $Q = Q_k$, $\alpha = \alpha_k$.

In order to make a reasonable choice for α one has to be aware of its effect on the corresponding interpolation with respect to both

– approximation behaviour and
– stability.

First note that the identity $A_{X,\phi_\alpha} = A_{X/\alpha,\phi}$ holds. In other words, interpolation at locations from X using ϕ_α can be understood as interpolation with ϕ on X/α. Therefore we shall fix ϕ and vary X/α, so that q is then to be replaced by $q_\alpha = q/\alpha$ and Q by $Q_\alpha = Q/\alpha$.

In view of the approximation behaviour, it is intuitively clear that a higher density of points yields a better approximation and from this point of view one desires as large as possible a value for α.

However, for large values of α one gets small quantities q_α which makes roughly speaking at least two rows (columns) of the collocation matrix nearly coincide. In these cases one has to expect a rather unstable interpolation process. In view of this one would prefer a small α.

Consequently, when selecting the scale one cannot expect to achieve both best possible stability and best possible approximation. One has to balance α in a useful manner according to the requirements of the particular application.

This phenomenon has already been observed in a more general context by Schaback who refers to it as an *uncertainty relation* in his survey paper [26]. Indeed, it is shown there that (radial) basis functions with good approximation behaviour necessarily provide bad upper bands for the spectral norm of the inverse of the corresponding collocation matrix.

Following along the lines of [26], pointwise error estimates of the interpolation method are usually to be obtained in terms of a local density measure around each of the points x in Ω. To be more precise, for some positive r this pointwise density measure is given by

$$h_r(x) := \sup_{y \in B_r(x)} \min_{1 \le j \le N} \| y - x_j \|$$

and is subject to $h_r(x) \le h$ for all $x \in \Omega$ with some positive constant h which does not depend on x.
Our definition of

$$Q := \max_{x \in \Omega} \min_{1 \le j \le N} \| x - x_j \|$$

establishes an appropriate relation to h_r since

$$Q = \lim_{r \to 0} \max_{x \in \Omega} h_r(x).$$

On the other hand [26], the stability of the interpolation process is intimately related to the size of the smallest eigenvalue of the corresponding collocation matrix $A_{X,\phi}$. In the setting of Schaback's survey, which is in these parts based on earlier papers [17–19] and [23, 24] the latter depends on the separation distance

$$q = \min_{1 \le j \le k \le N} \| x_j - x_k \| / 2$$

of the set X.

In view of our particular situation, this motivates for each possible α to use q_α as a measure for the sensitivity of the stability of the interpolation process, while we relate the magnitude of Q_α to the quality of its approximation behaviour.

If one wishes to emphasize quality of approximation throughout the hierarchical scheme, one should bound Q_α above by some fixed, and preferably small, number $\gamma > 0$ which will depend on the size of the domain Ω. This leads to

$$\alpha \geqslant Q/\gamma \tag{6}$$

as a lower bound for α.

Or, alternatively one may prefer to guarantee stability of the inversion of the collocation matrix by bounding q_α below by some sufficiently large fixed positive number δ, in which case one obtains an upper bound for α, namely

$$\alpha \leqslant q/\delta. \tag{7}$$

If one combines (6) and (7) one finds that

$$Q/\gamma \leqslant \alpha \leqslant q/\delta,$$

provided $Q/\gamma \leqslant q/\delta$, which in particular yields the lower bound

$$\rho = q/Q \geqslant \delta/\gamma.$$

Consequently, since we would ideally prefer to choose γ as low as possible and δ as high as possible, this suggests that the best balance can be achieved for high values of ρ. In consequence, in order to improve the effectiveness of the hierarchical method one could consider choosing the N_k to coincide with a selection of the peaks in the graph of ρ.

Finally, if one decides either to guarantee good approximation or stability throughout the hierarchical scheme then, based on the above observations, one obtains two possible strategies for choosing α_k in (4):

(i) fix $\gamma > 0$ and set $\alpha_k = Q_k/\gamma$ for all k, or
(ii) fix $\delta > 0$ and set $\alpha_k = q_k/\delta$ for all k.

Alternatively, one could even compromise the two strategies, and set

$$\alpha_k = (\lambda Q_k/\gamma + (1 - \lambda)q_k/\delta)$$

for some $\lambda \in (0, 1)$.

5. Numerical results

In analogy with hierarchical methods based on regular rectangular grids we chose $(N_k)_k$ to be a geometric sequence: $N_1 = 31, N_2 = 125, N_3 = 500, N_4 = 2000$. In our example we used Franke's function (Fig. 7)

$$f(x, y) = 0.75e^{-0.25(9x-2)^2-0.25(9y-2)^2} + 0.75e^{-(9x-2)^2/49-(9y-2)^2/10}$$
$$+ 0.5e^{-0.25(9x-7)^2-0.25(9y-3)^2} - 0.2e^{-(9x-4)^2-(9y-7)^2},$$

a well known standard function used in the investigation of methods for interpolating scattered data; see [10, 11]. The function was sampled on a randomly generated set of 2000 points in the unit square $\Omega = [0, 1] \times [0, 1]$ where its range is approximately $[-0.15, 1.50]$.
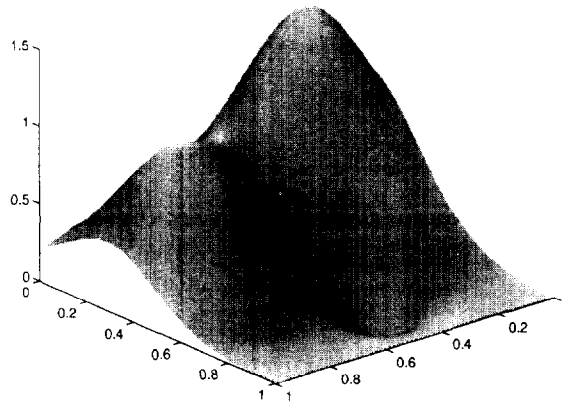
Fig. 7. Franke's function f.

Table 1

k	N_k	q_k	Q_k	ρ_k	α_k	e_k	u_k
1	31	0.0731	0.1957	0.373	1.957	0.1288	0.02
2	125	0.0316	0.0993	0.318	0.993	0.0219	0.53
3	500	0.0148	0.0456	0.325	0.456	0.0012	9.96
4	2000	0.0001	0.0423	0.002	0.100	0.0011	17.91

Figs. 1–4 show the points $X = X_4 = Y_{2000}$ and three subsets $X_3 = Y_{500}$, $X_2 = Y_{125}$, and $X_1 = Y_{31}$. Figs. 5 and 6 show q, Q and ρ in the thinning algorithm. Notice that X_1, X_2 and X_3 are much more uniform than X_4 and this is confirmed by the values of ρ in Table 1.

For the interpolation we chose the basis function $\phi(r) = (1 - r)^4_+ (4r + 1)$, and took γ to be 0.1 so that $\alpha_k = Q_k/\gamma = 10 \cdot Q_k$ for $k = 1, 2, 3$. Observe from Table 1 that the sequence $(\alpha_k)_k$ satisfies approximately $\alpha_k \approx 2\alpha_{k-1}$ except for $k = 4$ where we decided to reduce the radius α_k a little since ρ_4 is very small. The maximum error $e_k := |f - \sum_{j=1}^{k} s_j|_{\infty,\Omega}$ was found approximately by sampling on a 101×101 uniform rectangular grid in the unit square. The accumulated time in CPU units required for the inversion of the collocation matrices up to and including the kth step is denoted by u_k. For this purpose we have implemented the conjugate gradient method.

Figs. 8–11 show in sequence the resulting approximations s_1; $s_1 + s_2$; $s_1 + s_2 + s_3$; and $s = s_1 + s_2 + s_3 + s_4$. Figs. 12–15 show in sequence the approximation errors $f - s_1$; $f - (s_1 + s_2)$; $f - (s_1 + s_2 + s_3)$; and $f - s$.

In order to demonstrate the utility of the hierarchical method we have generated several test examples for the special case $M = 1$ where Franke's function was interpolated on X with different choices of the scale α between 0.065 and 0.225. Figs. 16 and 17 reflect our results. In Fig. 16 one can see the graph $u(\alpha)$ for the computational costs in CPU units as a function of the scale α and Fig. 17 shows the plot of the maximum error $e(\alpha)$ on Ω. Observe that there is no choice of α which achieves the balance between speed and reproduction performance offered by the hierarchical method. As
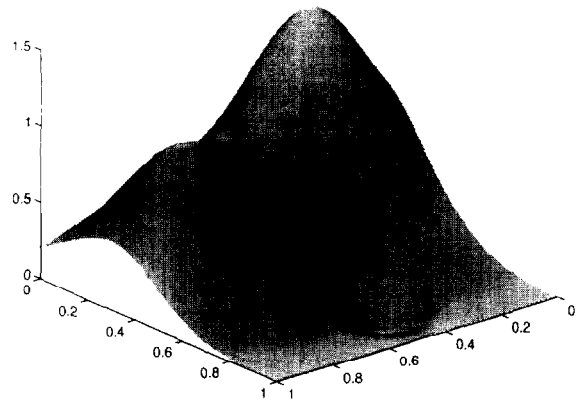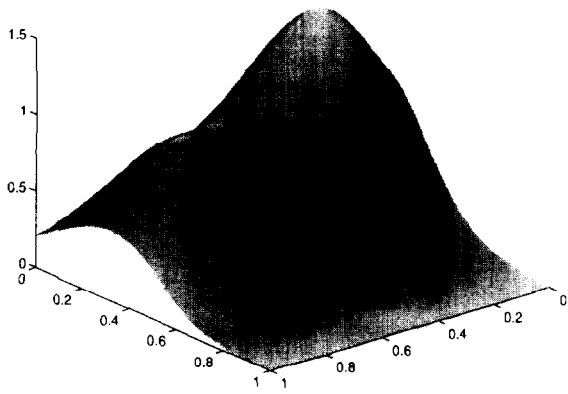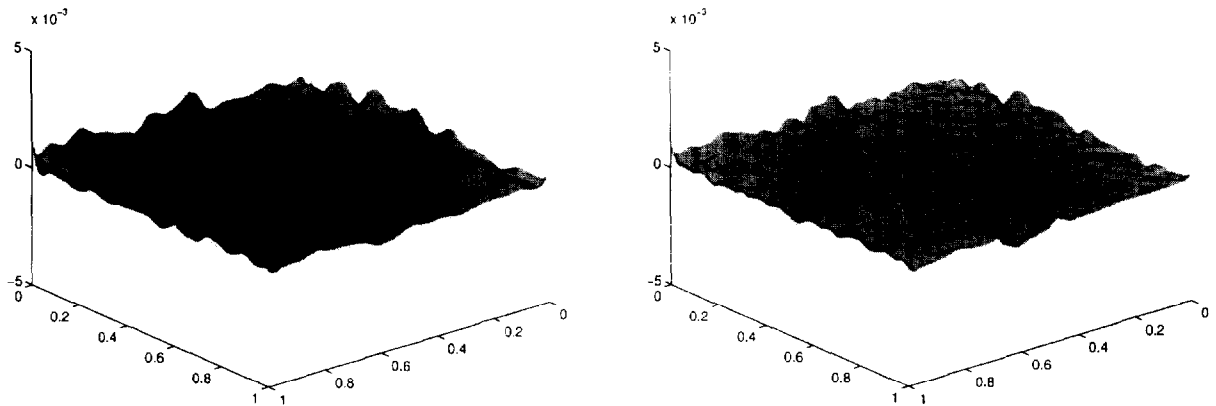
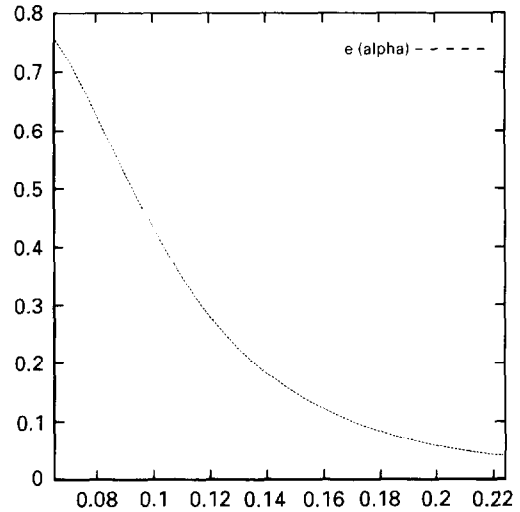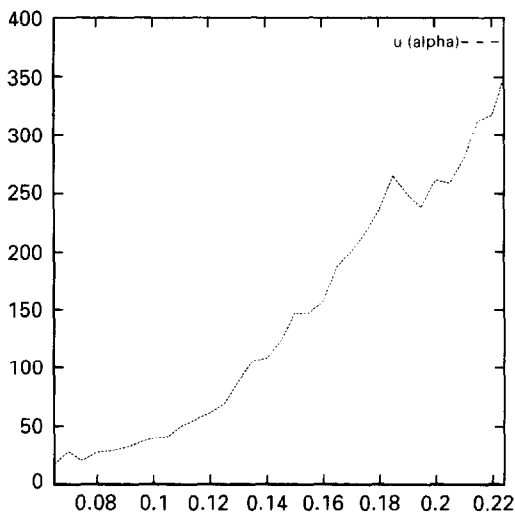Figs. 8 and 9. Approximation to f by s_1 and $s_1 + s_2$.
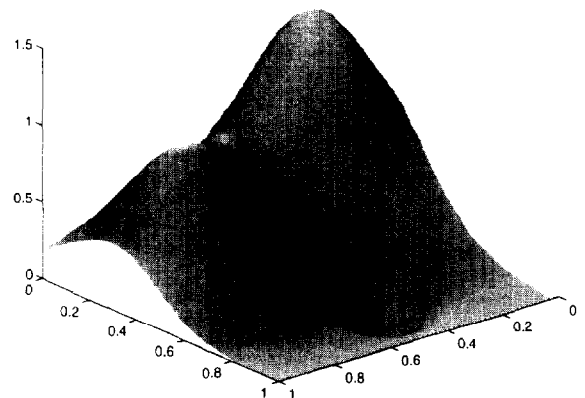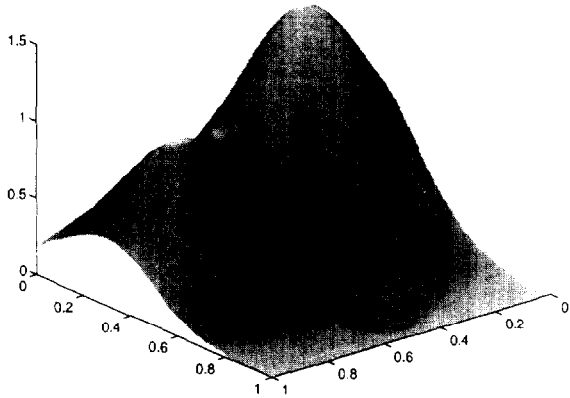
Figs. 10 and 11. Approximation to f by $s_1 + s_2 + s_3$ and $s = s_1 + s_2 + s_3 + s_4$.
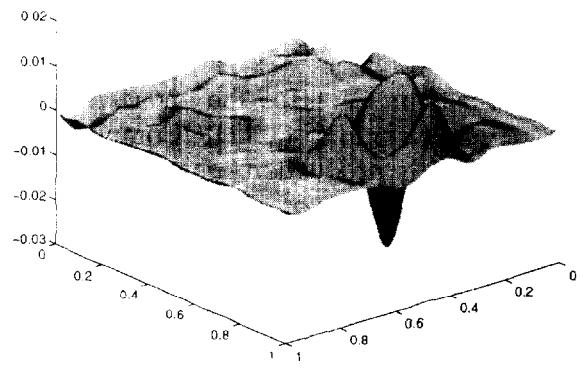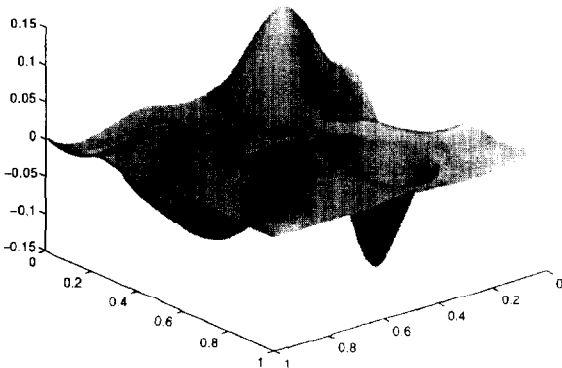
Figs. 12 and 13. Error functions $f - s_1$ and $f - (s_1 + s_2)$.

Figs. 14 and 15. Error functions $f - (s_1 + s_2 + s_3)$ and $f - (s_1 + s_2 + s_3 + s_4)$.

Figs. 16 and 17. CPU time $u(\alpha)$ and maximum error $e(\alpha)$ for the one-step method.

an example, for $\alpha = 0.165$ the one-step method requires $u(\alpha) = 187.49$ units of CPU time with a maximum error $e(\alpha) = 0.1117$. Compared to the hierarchical method this is more than ten times slower with a reproduction behaviour which is more than hundred times worse.

For the purpose of comparison we interpolated Franke's function on X_2 using thin plate splines. The maximum error on Ω was found to be 0.0296 attained at $(x, y) = (0.46, 0.8)$. This is approximately equivalent to the performance of the hierarchical method after step two. However, note that the thin plate spline (and any other classical radial basis function technique) will have stability problems on X due to the small separation distance $q(X)$.

Finally, we note that, in general, the computational complexity of the thinning algorithm is $\mathcal{O}(N \log N)$ for constructing the Delaunay triangulation and with our implementation, $\mathcal{O}(N^2)$ for

making the nested sequence $Y_K, \ldots, Y_N$, because finding a removable point at each step requires $\mathscr{C}(N)$ operations. The computational time for the multistep approximation depends on the method used for inverting the collocation matrices. Since we applied the conjugate gradient method on sparse matrices the costs are asymptotically $\mathscr{C}(N)$ only. The computational time could probably be reduced by further research along the lines of multilevel preconditioning strategies, which are typically used in the numerical solution of partial differential equations. For an up to date survey and further references we recommend [20].

6. Conclusions and final remarks

A hierarchical scheme, based on a thinning algorithm, has been presented for interpolating scattered data with radial basis functions of compact support.

One nice feature of the thinning algorithm is that the representation of the hierarchy by the sequence $(Y_i)_i$ can be stored simply as the sequence of points $y_1, \ldots, y_N$, which is just a reordering of the original set X. This requires relatively little storage space, and we found this convenient when experimenting with different basis functions and/or scales on the same data set.

The thinning algorithm depends on the scattered data points x_i and not on the values $f(x_i)$. One could also, especially when gradients in f are large, consider making the method data dependent.

Moreover, the hierarchical method could be regarded as a type of data reduction strategy. Indeed after several numerical observations, it has often been found that the graph of $\rho(i)$ is typically roughly constant for i up to some threshold i_s, after which it is monotonically decreasing. In the example of Fig. 6, $i_s = 563$. This suggests that one could consider only interpolating on the first i_s points, while ignoring the rest and still obtain a very good approximation. Indeed this is confirmed by the fact that there is very little change when going from the penultimate error $f - (s_1 + s_2 + s_3)$ (Fig. 14) to the last one $f - s$ (Fig. 15).

Acknowledgements

References

[1] F.P. Agterberg, *Geomathematics. Mathematical Background and Geo-Sciences Applications* (Elsevier, New York, 1974).

[2] M.D. Buhmann, New developments in the theory of radial basis function interpolation, in: K. Jetter and F. I. Utreras, Eds., *Multivariate Approximation: From CAGD to Wavelets* (World Scientific, Singapore, 1993) 35–75.

[3] L. De Floriani, A pyramidal data structure for triangle-based surface description, IEEE Comp. Graph. & App. (March 1989) 67–78.

[4] J. Duchon, Interpolation des Fonctions de deux variables suivant le principe de la flexion des plaques minces, *RAIRO Analyse Numeriques* **10** (1976) 5–12.

[5] J. Duchon, Splines minimizing rotation-invariant semi-norms in Sobolev spaces, in: W. Schempp and K. Zeller, Eds., *Constructive Theory of Functions of Several Variables* (Springer, Berlin, 1977) 85–100.

[6] J. Duchon, Sur L'erreur d'interpolation des fonctions de plusiers variables par les D^m-splines, *RAIRO Analyse Numeriques* **12** (1978) 325–334.

[7] N. Dyn, Interpolation of scattered data by radial functions, in C.K. Chui, L.L. Schumaker and F.I. Utreras, Eds., *Topics in Multivariate Approximation* (Academic Press, New York, 1987) 47–61.

[8] N. Dyn, Interpolation and approximation by radial and related functions, in C.K. Chui, L.L. Schumaker and J.D. Ward, Eds., *Approximation Theory VI: Vol. 1* (Academic Press, 1989) 211–234.

[9] N. Dyn, D. Levin and S. Rippa, Numerical procedures for global surface fitting of scattered data by radial functions, *SIAM J. Sci. Statist. Comput.* **7** (1986) 639–659.

[10] T.A. Foley, Interpolation and approximation of 3-D and 4-D scattered data, *Comp. Math. Appl.* **13** (1987) 711–740.

[11] R. Franke, Scattered data interpolation: test of some methods, *Math. Comp.* **38** (1982), 181–200.

[12] G.H. Golub and C.F. Van Loan, *Matrix Computation* (John Hopkins University Press, Baltimore, 2nd ed., 1989).

[13] R.L. Hardy, Multiquadric equations of topography and other irregular surfaces, *J. Geophys. Res.* **76** (1971) 1905–1915.

[14] W.R. Madych and S.A. Nelson, Multivariate interpolation and conditionally positive definite functions, *Approx. Theory Appl.* **4** (4) (1988) 77–89.

[15] W.R. Madych and S.A. Nelson, Multivariate interpolation and conditionally positive definite functions II, *Math. Comput.* **54** (189) (1990) 211–230.

[16] C.A. Micchelli, Interpolation of scattered data: distance matrices and conditionally positive definite functions, *Constr. Approx.* **2** (1986) 11–22.

[17] F.J. Narcowich, N. Sivakumar and J.D. Ward, On condition numbers associated with radial-function interpolation, Tech. Report No. 287, Department of Mathematics, Texas A&M University, 1993.

[18] F.J. Narcowich and J.D. Ward, Norms of inverses and condition numbers for matrices associated with scattered data, *J. Approx. Theory* **64** (1991) 69–94.

[19] F.J. Narcowich and J.D. Ward, Norm estimates for the inverses of a general class of scattered-data radial-function interpolation matrices, *J. Approx. Theory* **69** (1992) 84–109.

[20] P. Oswald, *Multilevel Finite Element Approximation* (Teubner, Stuttgart, 1994).

[21] M.J.D. Powell, The theory of radial basis function approximation in 1990, in: W.A. Light, Ed., *Advances in Numerical Analysis II: Wavelets, Subdivision Algorithms, and Radial Basis Functions* (Oxford University Press, Oxford, 1992) 105–210.

[22] F.P. Preparata and M.I. Shamos, *Computational Geometry* (Springer, New York, 1985).

[23] R. Schaback, Lower bounds for norms of inverses of interpolation matrices for radial basis functions, *J. Approx. Theory* **79** (1994) 287–306.

[24] R. Schaback, Error estimates and condition numbers for radial basis function interpolation, *Adv. Comput. Math.* **3** (1995) 251–264.

[25] R. Schaback, Creating surfaces from scattered data using radial basis functions, in: M. Dæhlen, T. Lyche and L. Schumaker, Eds., (Vanderbilt University Press, Nashville, 1995) 477–496.

[26] R. Schaback, Multivariate interpolation and approximation by translates of a basis function, in: C.K. Chui and L.L. Schumaker, Eds., (World Scientific, Singapore, 1995).

[27] R. Schaback and H. Wendland, Special case of compactly supported radial basis functions, manuscript, Göttingen, 1993.

[28] L.L. Schumaker, Triangulation methods, in: C.K. Chui, L.L. Schumaker and F. Utreras, Eds., *Topics in Multivariate Approximation* (Academic Press, New York, 1987) 219–232.

[29] H. Wendland, Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree, *Adv. Comput. Math.* **4** (1995) 389–396.

[30] Z. Wu, Characterization of positive definite radial functions, in: M. Dæhlen, T. Lyche and L. Schumaker, Eds., *Mathematical Methods for Curve and Surfaces* (Vanderbilt University Press, Nashville, 1995).

[31] Z. Wu, Multivariate compactly supported positive definite radial functions, *Adv. Comput. Math.* **4** (1995) 283–292.

[32] Z. Wu and R. Schaback, Local error estimates for radial basis function interpolation, *IMA J. Numer. Anal.* **13** (1993) 13–27.