Numerical Techniques Based on Radial Basis Functions

Robert Schaback and Holger Wendland

Abstract. Radial basis functions are tools for reconstruction of multivariate functions from scattered data. This includes, for instance, reconstruction of surfaces from large sets of measurements, and solving partial differential equations by collocation. The resulting very large linear $N \times N$ systems require efficient techniques for their solution, preferably of $\mathcal{O}(N)$ or $\mathcal{O}(N \log N)$ computational complexity. This contribution describes some special lines of research towards this future goal. Theoretical results are accompanied by numerical examples, and various open problems are pointed out.

§1. Introduction

Many problems of numerical analysis take the form of a generalized interpolation in spaces of multivariate functions [21]. Due to the Mairhuber-Curtis theorem [12], such spaces cannot be fixed beforehand, but must necessarily depend on the given data. For a plain multivariate interpolation problem on a finite set $X = \{x_1, \ldots, x_N\}$ of pairwise different points in a domain $\Omega \subseteq \mathbb{R}^d$, there is an easy possibility to generate a data-dependent space via linear combinations of something that depends on a free variable $x \in \Omega \subseteq \mathbb{R}^d$ and the data locations x_j, namely

$$S_{X,\Phi} := \text{span } \{\Phi(x, x_j) \, : \, 1 \leq j \leq N\} \tag{1}$$

with a fixed function $\Phi \, : \, \Omega \times \Omega \to \mathbb{R}$. The numerical generation of the space can be simplified considerably in the special situations

1) $\Phi(x, y) = \phi(x - y)$ with $\phi \, : \, \mathbb{R}^d \to \mathbb{R}$ (translation invariance),
2) $\Phi(x, y) = \phi(\|x - y\|_2)$ with $\phi \, : \, [0, \infty) \to \mathbb{R}$ (radiality),

Curve and Surface Fitting: Saint-Malo 1999
Albert Cohen, Christophe Rabut, and Larry L. Schumaker (eds.), pp. 359–374.

ISBN 0-8265-1357-3.

and this is how the notion of a radial basis function came up. To assure that the interpolation in the points of $X = \{x_1, \ldots, x_N\}$ is uniquely defined, the matrix

$$A_{\Phi,X} := (\Phi(x_j, x_k))_{1 \le j,k \le N} \tag{2}$$

must be nonsingular. By definition, positive definite functions Φ even make this matrix symmetric and positive definite, and the positive definite radial functions

$$\phi(r) = \exp(-r^2) \text{ on } \mathbb{R}^d \text{ for all } d \text{ (Gaussian)},$$
$$\phi(r) = (1-r)_+^4(1+4r) \text{ on } \mathbb{R}^d, d \le 3, \text{ see [25]},$$

are typical examples. See the review articles [4,15,18,16] for details. Though the above functions are scalar, the positive definiteness property of the second depends on the dimension d of the space containing x and y when forming the scalar argument $r = \|x - y\|_2$.

These two examples already show that the matrix $A_{\Phi,X}$ in (2) can be sparse or have a strong off-diagonal decay. It will be very large if we consider real-world problems with many data, arising e.g. from inverse engineering or terrain modelling. If the data points are rather densely scattered over the domain Ω, the approximation power of the space (1) will be very good, but the matrix in (2) will have lots of similar rows and columns, yielding a bad condition number. The connection between these phenomena is described in some detail in [17].

This paper concentrates on the numerical solution of large symmetric positive definite systems with matrices of the form (2). There are some additional goals:

1) $\mathcal{O}(N)$ complexity of solving the $N \times N$ system,
2) $\mathcal{O}(1)$ complexity of evaluating an element of (1) with N terms and
3) getting away with $n << N$ terms at a tolerable loss of accuracy, when interpolating N data.

We shall describe greedy algorithms as recently studied by deVore [5] and Temlyakov [24], but we shall omit multiple scales as proposed by Floater and Iske [9] and continued later in [13] and [6]. The techniques will be partially based on Krylov subspaces as recently and independently studied by Faul and Powell [8].

§2. Splitting the Native Space Energy

Dropping Φ in the notation, we can write functions from (1) in the form

$$s_{c,X} := \sum_{j=1}^{N} c_j \Phi(\cdot, x_j) \tag{3}$$

with $c \in \mathbb{R}^N$, $X = \{x_1, \ldots, x_N\} \subseteq \Omega \subseteq \mathbb{R}^d$ and arbitrary N. Our main tool will be the natural inner-product

$$(s_{c,X}, s_{d,Y})_\Phi = \left(\sum_j c_j \Phi(\cdot, x_j), \sum_k d_k \Phi(\cdot, y_k) \right)_\Phi := \sum_j \sum_k c_j d_k \Phi(x_j, y_k) \tag{4}$$

introduced by a positive definite function Φ on the union of spaces (1) for arbitrary data sets X. We note in passing that one can form the corresponding Hilbert space closure to get the native Hilbert space of Φ, but we refer the reader to [11] and [19] for details. If we fix the set $X = \{x_1, \ldots, x_N\}$ and the corresponding positive definite matrix $A_{\Phi,X}$, we get an inner-product

$$(c, d)_{A_{\Phi,X}} := c^T A_{\Phi,X} d \tag{5}$$

on $\mathbb{R}^N$ which is familiar from the theory of the conjugate gradient method based on Krylov subspaces.

Note that the inner-product (4) is zero if the function $s_{c,X}$ vanishes on Y or if $s_{d,Y}$ vanishes on X. If we assume $Y \subseteq X$ and make sure that $s_{d,Y}$ agrees with $s_{c,X}$ on Y (e.g. by interpolation on Y), then

$$\begin{aligned} (s_{d,Y}, s_{c,X} - s_{d,Y})_\Phi &= 0 \\ \|s_{d,Y}\|_\Phi^2 + \|s_{c,X} - s_{d,Y}\|_\Phi^2 &= \|s_{c,X}\|_\Phi^2. \end{aligned} \tag{6}$$

The second identity can be viewed as a split of the energy of the function $s_{c,X}$ into the energy of its interpolant $s_{d,Y}$ on a subset Y of X and the residual $s_{c,X} - s_{d,Y}$. We shall use this energy split over and over again.

§3. Interpolation in Native Spaces

At this point, we digress a little and study the interpolation of an *arbitrary* real-valued function f on a domain $\Omega \subseteq \mathbb{R}^d$. On each fixed finite subset $X \subset \Omega$ we can interpolate the values of f by a function $s_{f,X}$ from (1). Due to (6), the energy $\|s_{f,X}\|_\Phi^2$ is a monotonic function of X with respect to addition of points, and it can be easily evaluated using (4). The energy is bounded independent of f if and only if [22] the function f lies in the native space of Φ, and in this case we have $\|f\|_\Phi = \sup_X \|s_{f,X}\|_\Phi$.

This observation has some consequences for applications. If the user does not have any a-priori information on f, the proper choice of Φ is a problem. But if the behaviour of $\|s_{f,X}\|_\Phi^2$ with respect to X is monitored for larger and larger sets X, the user can switch to a less smooth Φ if the energy values grow dramatically with X. By (6) the behaviour of $\|s_{f,X}\|_\Phi^2$ is related to the energy $\|f - s_{f,X}\|_\Phi^2$ of the residual, and further study of this as a function of X is needed, especially for f with additional smoothness properties. Current starting points are in [20] and [10], and readers are encouraged to proceed from there.

§4. Iteration on Residuals

We now fix a positive definite function Φ and a function f_0 from the native space of Φ or, at least, from some space of the form (1) for a rather large finite set X. Our goal is to reconstruct f_0 by an iterative process. Note that this solves a large system with the matrix from (2), if we start from $f_0 := s_{f,X}$ interpolating some given function f on X. In this case, we do not reconstruct f but rather its interpolant. In both cases we need not worry about the existence of $\|f_0\|_\Phi$.

If we pick some numerically manageable finite set Y_0 and interpolate f_0 on Y_0, we can define $f_1 := f_0 - s_{f_0,Y_0}$ and proceed iteratively by

$$f_{j+1} := f_j - s_{f_j,Y_j}, \quad Y_j \subset \Omega, \quad j = 0, 1, \ldots \tag{7}$$

using finite sets Y_j that we shall have to deal with later. Anyway, the energy splitting (6) yields

$$\begin{aligned} \|s_{f_j,Y_j}\|_\Phi^2 + \|f_j - s_{f_j,Y_j}\|_\Phi^2 &= \|f_j\|_\Phi^2 \\ \|f_j - f_{j+1}\|_\Phi^2 + \|f_{j+1}\|_\Phi^2 &= \|f_j\|_\Phi^2, \end{aligned} \tag{8}$$

and by summation we get the telescoping sum

$$\sum_{j=0}^{k} \|s_{f_j,Y_j}\|_\Phi^2 = \sum_{j=0}^{k} \|f_j - f_{j+1}\|_\Phi^2 = \|f_0\|_\Phi^2 - \|f_{k+1}\|_\Phi^2 \le \|f_0\|_\Phi^2 \tag{9}$$

which must necessarily converge for $k \to \infty$ even if the choice of Y_j is bad, e.g. $Y_j = Y_0$ for all j. By standard Hilbert space argumentation via Cauchy sequences, the functions

$$g_{k+1} := \sum_{j=0}^{k} s_{f_j,Y_j} = \sum_{j=0}^{k} (f_j - f_{j+1}) = f_0 - f_{k+1}$$

must converge in the norm $\|.\|_\Phi$ to some element g in the native space, but we do not want to use this fact. Our goal is to prove that the residuals f_k converge to zero, and this would imply that the functions g_k converge to f_0, yielding the desired reconstruction.

Of course we shall need some additional assumptions on the sets Y_j to be successful. Equations (8) and (9) suggest that we should let s_{f_j,Y_j} take up as much energy from f_j as possible, and this will be our guideline for the convergence analysis in the following sections.

§5. Conditions for Linear Convergence

For simplicity, let us first assume that s_{f_j,Y_j} picks up at least a fixed percentage of the energy of f_j, i.e.

$$\|s_{f_j,Y_j}\|_\Phi^2 \ge \gamma \|f_j\|_\Phi^2 \tag{10}$$

with some fixed $\gamma \in (0, 1]$. This is a disguised hypothesis on the proper choice of Y_j, and we have to prove later how to satisfy this assumption. From (8) and (10) we conclude linear convergence of f_j to zero via $\|f_{j+1}\|_\Phi^2 \le (1-\gamma)\|f_j\|_\Phi^2$. This proves

Theorem 1. *If the choice of sets Y_j satisfies (10), the residual iteration (7) converges linearly in the native space norm, and there is an error bound*

$$\|f_0 - g_k\|_\Phi^2 = \|f_k\|_\Phi^2 \leq (1-\gamma)^k \|f_0\|_\Phi^2.$$

Assumption (10) is not easy to handle, because the norm involves Φ and the value of the right-hand side is not explicitly known. But in case of $f_0 := s_{f,X}$ for some large finite set $X \subset \Omega \subseteq \mathbb{R}^d$ we can restrict ourselves to sets $Y_j \subseteq X$, and all functions f_j will stay in the finite-dimensional space (1). On this space, we can pick any norm $\|.\|_X$, for instance any discrete L_p norm of functions on X, and make use of the norm equivalence

$$c\|s\|_X \leq \|s\|_\Phi \leq C\|s\|_X \tag{11}$$

for all functions s from the space (1), where the constants satisfy $0 < c \leq C$. Then we can try to get away with

$$\|s_{f_j,Y_j}\|_X^2 \geq \delta \|f_j\|_X^2 \tag{12}$$

with some $\delta \in (0,1]$ instead of (10). But since this equation implies (10) with $\gamma \geq \delta c^2/C^2 > 0$, we get

Theorem 2. *If the choice of sets Y_j satisfies (12) for some norm $\|.\|_X$ of functions on X, the residual iteration (7) for reconstruction of $f_0 := s_{f,X}$ converges linearly in $\|.\|_X$ and there is an error bound*

$$\|f - g_k\|_X \leq \frac{C}{c}\left(1 - \delta\frac{c^2}{C^2}\right)^{k/2} \|f\|_X.$$

§6. Maximizing Energy of Interpolants

Our argument at the end of Section 4 leads to the problem of finding a finite set Y such that the energy $\|s_{f,Y}\|_\Phi^2$ of the interpolant of some function (or residual) f is large. If $f_Y \in \mathbb{R}^{|Y|}$ is the vector of values of f on Y, the interpolant $s_{f,Y}$ solves a system with a matrix $A_{\Phi,Y}$ defined as in (2), and the energy is given by the quadratic form

$$\|s_{f,Y}\|_\Phi^2 = f_Y^T {A_{\Phi,Y}}^{-1} f_Y \geq \|f_Y\|_2^2 \lambda_{min}\left({A_{\Phi,Y}}^{-1}\right) = \|f_Y\|_2^2 / \lambda_{max}\left(A_{\Phi,Y}\right)$$

as a function of f and Y. The maximal eigenvalue $\lambda_{max}\left(A_{\Phi,Y}\right)$ is hard to discuss in general (see Narcowich and Ward [14] for results), and we simply view this quantity as a factor that depends on the geometry of Y and the number $|Y|$ of points in Y. It is an interesting open problem to design some Remes-type algorithm based on exchanges of points to arrive at the best choice of a set Y with a prescribed number of points.

In the special case $|Y| = 1$, $Y = \{y\}$ things are easy. We get

$$\|s_{f,Y}\|_\Phi^2 = f(y)^2 \Phi(y,y) \geq f(y)^2 \min_{z \in \Omega} \Phi(z,z)$$

and the maximum of f^2 will be the best choice, especially if Φ is translation-invariant or radial.

If we have $Y \subseteq X$ for a large finite set X, we can invoke Courant's minimum-maximum principle to get $\lambda_{max}(A_{\Phi,Y}) \leq \lambda_{max}(A_{\Phi,X})$ as an upper bound that does not depend on the choice of Y. A reasonable strategy for maximizing $\|s_{f,Y}\|_\Phi^2$ then is to pick the $|Y|$ points of X where f takes its largest absolute values.

In the general situation, we have to face the fact that coalescing points are not allowed. A reasonable strategy is to mimic the previous situation, i.e. to take some large set X of well-distributed points and pick the points of X where f is largest in absolute value.

Another possible strategy is the iterative greedy collection of more and more points, forming a recursive Cholesky factorization. Since this possibility does not seem to be familiar to researchers in this area, we outline the process here. Assume that an interpolant to f on Y is available together with the inverse B of $A_{\Phi,Y}$. We now want to add another point $z \in \Omega \setminus Y$ to Y, thus enlarging the energy of the interpolant. A naive choice of z is via the maximum of the absolute value of $f - s_{f,Y}$, but since we have

$$\|s_{f,Y\cup\{z\}}\|_\Phi^2 = \|s_{f,Y}\|_\Phi^2 + 2f(z) \sum_{y \in Y} f(y)\Phi(z,y) + f^2(z)\Phi(z,z), \qquad (13)$$

the best choice of z for fixed Y is obtained by maximizing the right-hand side of this equation. Having found z, one has to update B in a suitable way. First, calculate the vector $v \in \mathbb{R}^{|Y|}$ with components $\Phi(z,y)$, $y \in Y$ and form $w := Bv$. The number

$$1/\alpha := \Phi(z,z) - v^T w = \Phi(z,z) - v^T B v$$

can be shown to be positive, because Φ is positive definite and z does not belong to Y. Then form $u := -\alpha w$ and $C := B + u^T u/\alpha$. The matrix

$$\begin{pmatrix} C & u \\ u^T & \alpha \end{pmatrix}$$

then is the inverse of $A_{\Phi,Y\cup\{z\}}$ needed for the interpolation on $Y \cup \{z\}$. Unfortunately, there is no numerical experience in this direction so far, especially for the maximization of (13). A more careful calculation of the numerical complexity reveals that we have nothing else here than a special formulation of the partial Cholesky algorithm with pivoting. The choice of pivots, however, is adapted to the setup of the problem as an interpolation.

Altogether, this section was intended to motivate readers to look at the problem of finding good finite sets Y for improvement of interpolants.

§7. A Simple Greedy Algorithm

Among other things, the previous section showed how to work on subsets Y consisting of a single point y each. The best possible choice is to take the point where f takes its maximum absolute value, and the interpolant is $s_{f,\{y\}} = f(y)\Phi(y,\cdot)/\Phi(y,y)$. We now do this iteratively in the sense of (7) by picking $Y_j := \{y_j\}$ with $|f_j(y_j)| = \|f_j\|_\infty$. In the "discrete" case $f_0 := s_{f,X}$ we take the Chebyshev norm on X, while in the "continuous" case $f_0 := f \in C(\Omega)$ with Ω being a compact subset of $\mathbb{R}^d$, we take the Chebyshev norm on Ω.

Due to Theorem 2, the discrete case leads to linear convergence towards $f_0 := s_{f,X}$, because (12) is satisfied with $\delta = 1$. From a theoretical viewpoint, this is much better than the non-quantitative convergence result of Faul and Powell in [7]. On the other hand, there always is the conjugate gradient method as a competitor, and it has linear convergence, too. But it needs to form matrix-vector products, while our greedy algorithm does not even store the matrix. It simply needs two arrays of length $|X|$ for the residuals and the coefficients, and in each cycle it updates one cofficient and runs once over the residuals to update them and find the maximum for the next iteration. This extreme numerical simplicity must come at a price, and the price is very slow convergence after some good progress in the first few iterations. We report on numerical experiments and adaptive multiscale improvements in [23], but at this point we want to direct the reader's attention to extend the above strategy, e.g. via some suitable preconditioning.

Before we look more closely at the greedy algorithm in the discrete case, let us digress a little into the continuous case.

Theorem 3. *If Φ is a continuous translation-invariant positive definite function on a compact domain $\Omega \subset \mathbb{R}^d$, the greedy algorithm for interpolation of a function f from the native space of Φ converges uniformly.*

Proof: We have

$$\|s_{f_j,Y_j}\|_\Phi^2 = f_j^2(y_j)\phi(0) = \|f_j\|_\infty^2\phi(0)$$

and (9) shows that the quantities $\|f_j\|_\infty^2$ are summable. Consequently, the residuals f_j converge uniformly to zero on the compact set Ω. □

Corollary 4. *Under the assumptions on Φ as in Theorem 3, the native space norm is expressible via a series*

$$\|f\|_\Phi^2 = \phi(0)\sum_{j=0}^{\infty} f_j^2(y_j) = \phi(0)\sum_{j=0}^{\infty}\|f_j\|_\infty^2,$$

where

$$f_0 := f,\ |f_j(y_j)| = \|f_j\|_\infty,\ f_{j+1} := f_j - f_j(y_j)\phi(\cdot - y_j)/\phi(0).$$

This result may look complicated at first sight, but it should be compared to other definitions of the native space norm, e.g. via Fourier transforms, by

abstract completion of a pre-Hilbert space, or by the supremum of the action of certain functionals.

We do not want to go into details here (see [11], for instance), but prefer to give an illustrative example. If specialized to Sobolev space $W_2^k(\Omega)$ with $k > d/2$, one has to take $\phi(x) = \|x\|_2^{k-d/2} K_{k-d/2}(\|x\|_2)$ in order to recover Sobolev space as a native space for $\Phi(x,y) = \phi(\|x-y\|_2)$. Now by Corollary 4 one gets the Sobolev norm of a function as a series containing just function values on the domain in a numerically accessible way, using neither derivatives nor integration (but, of course, maximization). It should be pointed out that this technique provides some means to assess the Sobolev smoothness of a given function numerically. Readers are encouraged to proceed from here.

§8. Dual Techniques

Another possible approach to solving a large $N \times N$ system with a large symmetric and positive definite coefficient matrix $A_{\Phi,X}$ via smaller finite subproblems is to define certain finite-dimensional subspaces S_j of the native space and to approximate the exact solution on $X = \{x_1, \ldots, x_N\}$ by approximation in the native space norm. More precisely, the iteration starts like in Section 4 with some function f_0 and $j := 0$, and iterates like (7) according to

$$\begin{aligned} \|f_j - s_j\|_\Phi &:= \inf_{s \in S_j} \|f_j - s\|_\Phi \\ f_{j+1} &:= f_j - s_j. \end{aligned} \tag{14}$$

By standard arguments, this iteration also satisfies (8) and the rest of Section 4, including the summability condition (9). Note that if a space $S_j = S_{Y_j,\Phi}$ has the form (1) for some finite set Y_j, then the best approximation solution s_j in (14) coincides with s_{f_j,Y_j} and we are back to the method in Section 4. This observation follows from Theorem 7 in Section 9.

But there are other possible choices for the spaces S_j. In particular, Faul and Powell [7] pick certain one-dimensional spaces $S_j = \operatorname{span}\{u_j\}$ for all $j \geq 0$. Then $s_j := \alpha_j u_j$ with $\alpha_j := (u_j, f_j)_\Phi / (u_j, u_j)_\Phi$ solves the approximation problem, and we have the summability condition

$$\sum_{j=0}^{k} \|s_j\|_\Phi^2 = \sum_{j=0}^{k} \|u_j\|_\Phi^2 \alpha_j^2 = \sum_{j=0}^{k} \left(\frac{u_j}{\|u_j\|_\Phi}, f_j\right)_\Phi^2 = \|f_0\|_\Phi^2 - \|f_{k+1}\|_\Phi^2 \leq \|f_0\|_\Phi^2. \tag{15}$$

§9. Cyclic and Greedy Dual Strategies

In [7], Faul and Powell fix N such functions u_j by a certain precalculation that we shall discuss later. These functions are used periodically, i.e. u_j is used in step $j + kN$ for all $k \geq 0$. The periodic reuse has the advantage that one can precalculate and store the u_j, if their construction is somewhat involved. We start with a generalized and simplified version of the convergence result in [7]:

Theorem 5. *If f_0 is in the span of the functions u_j for $1 \le j \le N$, then the cyclic dual method of Faul and Powell converges to f_0.*

Proof: Since everything takes place in a finite-dimensional space, and since the technique involves an energy split, the functions $g_j = f_0 - f_j$ converge to some function g in the span of the u_j, and the f_j converge to $f_0 - g$. But as (15) implies

$$\lim_{k\to\infty} (\frac{u_j}{\|u_j\|_\Phi}, f_{j+kN})_\Phi = 0 = (\frac{u_j}{\|u_j\|_\Phi}, f_0 - g)_\Phi$$

for all j, the functions f_0 and g must coincide. □

There are lots of choices of u_j that satisfy the hypothesis of Theorem 5. Conjugate directions and $u_j := \Phi(\cdot, x_j)$ would do the job. The latter strategy coincides with the greedy method, if the cyclic choice is given up in favour of picking the point where the residual is maximal in absolute value. A linear convergence result is possible, if such a modification is made in general:

Theorem 6. *If f_0 is in the span of the functions u_j for $1 \le j \le N$, then the iteration (14) with*

$$\begin{aligned} (f_j, u_{k_j})_\Phi^2 &:= \max_k (f_j, u_k)_\Phi^2 \\ S_j &:= \text{span}\,\{u_{k_j}\} \end{aligned} \tag{16}$$

converges linearly to f_0.

Proof: We can proceed as in Section 5, using

$$\|u\|_X^2 := \max_j (u, u_j)_\Phi^2$$

for all functions u in the span U of all u_j. The assumption (12) is satisfied for s_j instead of s_{f_j,Y_j} due to

$$\begin{aligned} s_j &= \frac{(f_j, u_{k_j})_\Phi}{(u_{k_j}, u_{k_j})_\Phi} u_{k_j} \\ \|s_j\|_X^2 &\ge (s_j, u_{k_j})_\Phi^2 = (f_j, u_{k_j})_\Phi^2 = \|f_j\|_X^2, \end{aligned}$$

and the rest follows easily. □

The inner-products in (16) can be evaluated explicitly, if we work in the space (1) and use (4) and (5) in the form

$$(s_{c,X}, s_{d,X})_\Phi = \sum_k c_k s_{d,X}(x_k). \tag{17}$$

This is particularly efficient if the functions u_j have only a small number of nonzero coefficients in their representation of the form $u_j = s_{c^j,X}$. Another possibility, exploiting the dual nature of the algorithm, is to store and update

the inner-products $(f_j, u_k)_\Phi$ instead of the values $f_j(x_k)$. So far, there is no numerical experience with dual greedy algorithms, unfortunately.

One has a lot of leeway for picking suitable functions u_j, especially when preconditioning arguments come into play. Faul and Powell use local Lagrange functions u_j based on relatively small subsets Y_j of X that contain x_j. In particular, $u_j \in S_{Y_j,\Phi}$ is defined by the interpolation conditions

$$\begin{aligned} &u_j(x_j) = 1, \\ &u_j(x_k) = 0 \text{ for all } x_k \in Y_j,\ k \neq j, \end{aligned} \tag{18}$$

and is expressible in the form $u_j = s_{c^j,Y_j} = s_{d^j,X}$ with at most $|Y_j|$ nonzero coefficients. The precalculation involves the solution of N systems with $|Y_j| \times |Y_j|$ matrices A_{Φ,Y_j},and it can be kept at $\mathcal{O}(N)$, if the values $|Y_j|$ are bounded independent of N. Our arguments in Section 10 will show how this technique can be interpreted as preconditioning the matrix $A_{\Phi,X}$. For a fixed accuracy to be obtained, and for their special choice of the sets Y_j, Faul and Powell then observe that they need only a small fixed number of cycles of the dual algorithm. Each cycle has N one-dimensional subproblems, but there are techniques to keep each subproblem at a reasonable complexity, provided that techniques like multipole expansions [1] or compactly supported radial basis functions [26, 25] are used.

The selection of functions u_j is particularly good if there are orthogonality or conjugacy relations among them. Let us look at an inner-product $(u_j, u_k)_\Phi$ in case of (18), using (17) and $u_j = s_{c^j,X}$. We get

$$(u_j, u_k)_\Phi = \sum_{m:x_m \in Y_j} c^j_m u_k(x_m),$$

and this quantity vanishes if $Y_j \subseteq Y_k \setminus \{x_k\}$.

This can be seen as a motivation for choosing

$$x_j \in Y_j \subseteq \{x_j, x_{j+1}, \ldots, x_N\} \tag{19}$$

as done by Faul and Powell. Even if the functions u_j are in general not mutually orthogonal they are at least linear independent as needed for Theorems 5 and 6. To see this note that the matrix $C = (c^j_i)$ which describes the transition from the basis $(\Phi(\cdot, x_j), 1 \leq j \leq N)$ to $(u_j, 1 \leq j \leq N)$ is an upper triangle matrix and thus invertible if $c^i_i \neq 0$ for $1 \leq i \leq N$. This is indeed the case because of

$$0 \neq \|u^i_i\|^2_\Phi = \sum_{m:x_m \in Y_j} c^i_m u_i(x_m) = c^i_i.$$

We finish this section by pointing out how to make optimal use of solving N systems with $|Y_j| \times |Y_j|$ matrices A_{Φ,Y_j} for subsets Y_j in a precalculation. If the full inverses of the A_{Φ,Y_j} are stored instead of the coefficients of u_j, one can use the cyclic dual algorithm with $S_j := S_{Y_j,\Phi}$. The energy split at each step of the algorithm will then be better or equal to the split obtained by the dual cyclic algorithm using a single $u_j \in S_{Y_j,\Phi}$ like the one used by Faul and Powell. This is clear from (14), and the following theorem, which is well known since the advent of splines, shows that we end up with a cyclic interpolatory method of the form (7).

Theorem 7. *If Y is a finite subset of Ω, the approximation problem*

$$\inf_{s \in S_{Y,\Phi}} \|f - s\|_\Phi$$

for any f in the native space of Φ is solved by the interpolant $s_{f,Y}$.

Proof: Equations (6) generalize via continuous transition to the Hilbert space completion to

$$(s_{f,Y}, f - s_{f,Y})_\Phi = 0$$
$$\|s_{f,Y}\|_\Phi^2 + \|f - s_{f,Y}\|_\Phi^2 = \|f\|_\Phi^2$$

for all f in the native space, and the assertion follows. □

Consequently, algorithms using interpolants on finite subsets make optimal use of the information contained in the space $S_{Y,\Phi}$. This links the dual techniques back to the interpolatory methods in Section 4. Numerical results concerning the above cyclic interpolatory method, e.g. using the sets Y_j of Faul and Powell, are still missing. The progress must be better due to Theorem 7, but at the expense of much more storage. And, an incorporation of greedy selections using the good preconditioning power of the Faul-Powell approach seems worth investigating.

§10. Quasi-Interpolation

There is a hidden link between the Faul-Powell technique, preconditioning of $A_{\Phi,X}$, and certain quasi-interpolation methods using local Lagrange functions, as investigated by Beatson, Powell and their coworkers (see for example [2]). If we write the interpolant $s_{f,X}$ to some function f in Lagrange representation

$$s_{f,X} = \sum_{j=1}^{N} f(x_j) v_j \tag{20}$$

with N Lagrange basis functions $v_k \in S_{X,\Phi}$ satisfying $v_j(x_k) = \delta_{jk}$, we can relax (20) to a quasi-interpolation formula

$$s_{f,u,X} := \sum_{j=1}^{N} f(x_j) u_j \tag{21}$$

for any other choice of functions u_j that approximate the global Lagrange basis functions v_j. The choice (18) for certain subsets Y_j is quite natural, because one can often [3] observe that local Lagrange functions based on a set Y_j of neighbouring points to $x_j \in Y_j$ decay quickly away from x_j. Assuming (18) (but not (19)) from now on, the representation (21) can be rewritten in terms of $u_j = s_{c^j,X}$ and (3) as

$$s_{f,u,X} = \sum_{j=1}^{N} f(x_j) \sum_{k:x_k \in Y_j} c_k^j \Phi(\cdot, x_k) = \sum_{k=1}^{N} \Phi(\cdot, x_k) \sum_{j:x_k \in Y_j} f(x_j) c_k^j.$$

The coefficients of the second representation can be evaluated locally, and the computational advantage is particularly evident in case of compactly supported radial basis functions.

We now want to look at the quality of such quasi-interpolants on the discrete set X itself. The operator that maps the vector

$$f_{|X} := (f(x_1), \ldots, f(x_N))^T \in \mathbb{R}^N$$

to $s_{f,u,X_{|X}} \in \mathbb{R}^N$ can be written as the matrix product $A_{\Phi,X} \cdot C$, where $C = (c_k^j)$ is the nonsymmetric $N \times N$ matrix with row index k and column index j containing the coefficients c_k^j of the u_j columnwise. The operator that generates the residuals on X then is $E_N - A_{\Phi,X} \cdot C$ with the $N \times N$ identity matrix E_N. In case of $Y_j = X$ for all j we have $C = {A_{\Phi,X}}^{-1}$, and there are good reasons to expect that there are numerically interesting cases where some matrix norm of $E_N - A_{\Phi,X} \cdot C$ is smaller than one. In such cases one can solve the problem on X by successive quasi-interpolation via a Neumann series. In terms of vectors f^j and s^j containing the values of residuals f_j and quasi-interpolants to f_j on X, we have the linearly convergent iteration

$$\begin{aligned} f^0 &:= f_{0|X} \\ s^j &:= A_{\Phi,X} \cdot C f^j \\ f^{j+1} &:= f^j - s^j = (E_N - A_{\Phi,X} \cdot C)^{j+1} f^0 \end{aligned}$$

calculating the interpolant to the data of f_0 on X as the sum over the s^j. Note that we cannot use the energy split here, because we have left the context of interpolation and approximation. Note further that C acts as a (nonsymmetric!) preconditioner or an approximate inverse to $A_{\Phi,X}$.

§11. Experiments Concerning Quasi-Interpolation

To calculate the norm of $E_N - A_{\Phi,X} \cdot C$ numerically, we observe that the matrix $A_{\Phi,X} \cdot C$ has the entries $u^j(x_i)$, where i is the row index. Thus the entry at (i,j) of $E_N - A_{\Phi,X} \cdot C$ vanishes for $x_i \in Y_j$, and the column-sum norm of $E_N - A_{\Phi,X} \cdot C$ can be written as

$$\max_j \sum_{i:x_i \notin Y_j} |u^j(x_i)|. \tag{22}$$

Again it turns out that the decay of local Lagrange functions is essential.

In case of data on the uniform grid $(h\mathbb{Z})^2$, a radial basis function ϕ_c with support in $[0,c]$, and sets $Y_j := \{y \in (h\mathbb{Z})^2 \ : \ \|x_j - y\|_2 \leq R\}$ of neighbours to x_j within a radius R, the norm in (22) can be evaluated by looking at the local Lagrange function u_0 with respect to the origin and the set $Y_0 := \{y \in (h\mathbb{Z})^2 \ : \ \|y\|_2 \leq R\}$ of local interpolation points. Since both Y_0 and the support of ϕ are bounded, the function u_0 is zero on integer grid

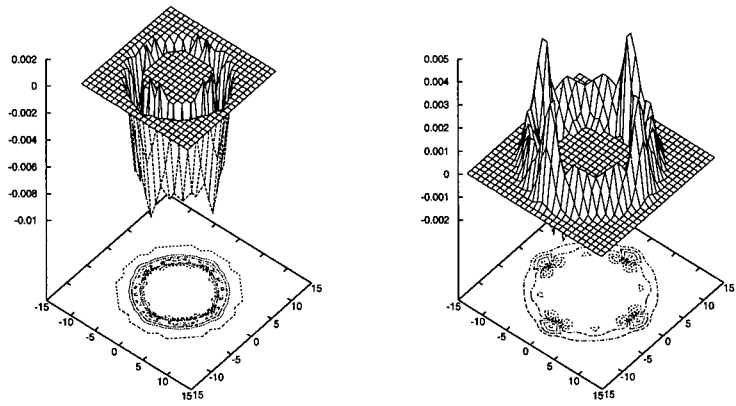

Fig. 1. $C = 4, R = 6$, norm $= 0.48$, and $C = 5, R = 8$, norm $=0.29$.

N	5	9	13	21	25	29	37	45	49	57	61
$M_{0.9}$	5	5	5	9	29	49	69	81	97	145	145
$M_{0.1}$	5	5	21	29	109	137	149				

Tab. 1. Point numbers M_p required for norm $\leq p$ and N points in support of ϕ.

points outside the disk around zero with radius $R + c$. Omitting the value 1 at the origin for scaling reasons, Figure 1 shows the behaviour of u_0 on integer gridpoints around the origin. We picked two cases for the C^2 function $\phi_c(r) := (1 - r/c)_+^4(1 + 4r/c)$ from [25] where the norm of $E_N - A_{\Phi,X} \cdot c$ is smaller than one, and the corresponding numbers of local interpolation points in Y_0 are 113 and 197, respectively.

For applications, it is necessary to know how large R must be for fixed c and h in order to make the norm of $E_N - A_{\Phi,X} \cdot c$ smaller than 0.9 or 0.1, say. Since R and c scale with h, the numbers M and N of points in Y_0 and the support of ϕ depend on R/h and C/h, respectively. Given a support radius c and a maximal meshwidth h such that the support of ϕ_c contains $N = 1, 5, 9, 13, \ldots$ points, we provide in Table 1 the minimal number M_p of points in Y_0 that are necessary to keep the norm of $E_N - A_{\Phi,X} \cdot c$ below p. Another way of reading Table 1 is that if the matrix $A_{\Phi,X}$ for interpolation by $\Phi(x, y) := \phi_c(\|x - y\|_2)$ on a regular grid has bandwidth N, then it has an approximate inverse with bandwidth M_p that leads to a residual matrix of norm p. The quasi-interpolant is to be calculated via subproblems with $M_p \times M_p$ matrices. It is an interesting challenge to provide sparse approximate inverses for sparse symmetric positive definite matrices, because normally the exact inverses will not be sparse.

§12. Conclusions

At first sight, our results on linear convergence look promising, but they still are too weak to provide a convergence rate that is independent of N, since no preconditioning techniques are involved. Improvements should thus focus on preconditioning, e.g. along the lines of Faul and Powell. Greedy methods for fixed Φ are limited to quick-and-dirty approximations with few nonzero coefficients and need extension to multiscale techniques. The adaptive greedy method in [23] is a first step, but the results shown there imply that it has to be stopped before it runs into scales that are too small. A possible continuation at small scales is provided by quasi-interpolation as outlined here. A combination of both techniques generates approximations which consist first of $K << N$ global terms obtained by an adaptive greedy method, followed by N local terms constructed by quasi-interpolation. The overall complexity can thus be kept at $\mathcal{O}(N)$.

Acknowledgments. The authors are grateful to Fabien Hinault for detecting an error in the first version of the paper.

References

1. Beatson, R. K. and G. N. Newsam, Fast evaluation of radial basis functions. I, Comput. Math. Appl. **24** (1992), 7–19.
2. Beatson, R. K., G. Goodsell, and M. J. D. Powell, On multigrid techniques for thin plate spline interpolation in two dimensions, in *Lectures in Applied Mathematics*, Vol. 32, 77–97.
3. Buhmann, M. D., Multivariate cardinal interpolation with radial-basis functions, Constr. Approx. **6** (1990), 225–255.
4. Buhmann, M. D., New developments in the theory of radial basis function interpolation, in *Multivariate Approximation: From CAGD to Wavelets*, K. Jetter and F. I. Utreras (eds.), World Scientific, Singapore, 1993, 35–75.
5. DeVore, R. A. and V. N. Temlyakov, Some remarks on greedy algorithms, Advances in Comp. Math. **5** (1996), 173–187.
6. Fasshauer, G. and J. Jerome, Multistep approximation algorithms: improved convergence rates through postconditioning with smoothing kernels, Advances in Comp. Math. **10** (1999), 1–27.
7. Faul, A. C. and M. J. D. Powell, Proof of convergence of an iterative technique for thin plate spline interpolation in two dimensions, Advances in Comp. Math. **11** (1999), 183–192.
8. Faul, A. C. and M. J. D. Powell, Krylov subspace methods for radial basis function interpolation, preprint.
9. Floater, M.S. and A. Iske: Multistep scattered data interpolation using compactly supported radial basis functions, J. Comput. Appl. Math. **73** (1996), 65–78.

10. Johnson, M., Overcoming the boundary effects in suface spline interpolation, preprint.

11. Luh, L.-T., Characterizations of native spaces, PhD. Dissertation, Göttingen, 1998.

12. Mairhuber, J. C., On Haar's theorem concerning Chebychev approximation problems having unique solutions, Proc. Amer. Math. Soc. **7** (1956), 609–615.

13. Narcowich F. J., R. Schaback, and J. D. Ward, Multilevel interpolation and approximation, Applied and Computational Harmonic Analysis, to appear.

14. Narcowich, F. J., N. Sivakumar, and J. D. Ward, On condition numbers associated with radial-function interpolation, J. Math. Anal. Appl. **186** (1994), 457–485.

15. Powell, M. J. D., The theory of radial basis function approximation in 1990, in *Advances in Numerical Analysis Vol. 2,* W. Light (ed.), Clarendon Press, Oxford, 1992, 105–210.

16. Schaback, R., Multivariate interpolation and approximation by translates of a basis function, in *Approximation Theory VIII, Vol. 1: Approximation and Interpolation*, Charles K. Chui and Larry L. Schumaker (eds.), World Scientific Publishing Co., Inc., Singapore, 1995, 491–514.

17. Schaback, R., Error estimates and condition numbers for radial basis function interpolation, Advances in Comp. Math. **3** (1995), 251–264.

18. Schaback, R., Creating surfaces from scattered data using radial basis functions, *Surface Fitting and Multiresolution Methods*, A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press, Nashville, 1997, 477–496.

19. Schaback, R., Native spaces for radial basis functions I, in *New Developments in Approximation Theory*, M. W. Müller, M. D. Buhmann, D. H. Mache, and M. Felten (eds.), Birkhäuser, Basel, 1999, 255–282.

20. Schaback, R., Improved error bounds for scattered data interpolation by radial basis functions, Math. Comp. **68** (1999), 201–216.

21. Schaback, R. and H. Wendland, Using compactly supported radial basis functions to solve partial differential equations, in *Boundary Element Technology XIII*, C. Chen, C. A. Brebbia, and D. W. Pepper (eds.), Wit-Press, Southampton, Boston, 1999, 311–324

22. Schaback, R. and H. Wendland, Inverse and saturation theorems for radial basis function interpolation, preprint.

23. Schaback, R. and H. Wendland, Adaptive greedy techniques for approximate solution of large RBF systems, Numerical Algorithms, to appear.

24. Temlyakov, V. N. The best m-term approximation and greedy algorithms, Advances in Comp. Math. **8** (1998), 249–265.

25. Wendland, H., Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree, Advances in Comp. Math. **4** (1995), 389–396.
26. Wu, Z., Multivariate compactly supported positive definite radial functions, Advances in Comp. Math. **4**, (1995), 283–292.

Robert Schaback & Holger Wendland
Institut für Numerische und Angewandte Mathematik
Universität Göttingen, Lotzestraße 16-18
D-37083 Göttingen, Germany
schaback@math.uni-goettingen.de
wendland@math.uni-goettingen.de