DAMTP 1999/NA11

Krylov subspace methods for radial basis function interpolation[1]

A.C. Faul and M.J.D. Powell

Abstract: Radial basis function methods for interpolation to values of a function of several variables are particularly useful when the data points are in general positions, but hardly any sparsity occurs in the matrix of the linear system of interpolation equations. Therefore an iterative procedure for solving the system is studied. The k-th iteration calculates the element in a k-dimensional linear subspace of radial functions that is closest to the required interpolant, the subspaces being generated by a Krylov construction that employs a self-adjoint operator A. Distances between functions are measured by the semi-norm that is induced by the well-known conditional positive or negative definite properties of the matrix of the interpolation problem, and conditions on A are found that guarantee successful termination of the iterations in exact arithmetic. A particular choice of A is recommended and is tried in numerical experiments. Fortunately, the number of iterations to achieve high accuracy is much less than the number that provides termination in theory, at most ten iterations being usual in practice. Thus the given procedure solves the interpolation equations very efficiently when the number of data is large.

Department of Applied Mathematics and Theoretical Physics,
University of Cambridge,
Silver Street,
Cambridge CB3 9EW,
England.

August, 1999.

[1]Presented at the 18th Biennial Conference on Numerical Analysis (Dundee)

1. Introduction

In many applications, some values of an unknown function from $\mathcal{R}^d$ to $\mathcal{R}$ are measured or are given, and one has to choose a function, s say, that fits these data exactly. Therefore the conditions on s are the equations

$$s(\underline{x}_i) = f_i, \qquad i=1,2,\ldots,n, \tag{1.1}$$

where the points $\underline{x}_i \in \mathcal{R}^d$, $i = 1,2,\ldots,n$, and the right hand sides $f_i \in \mathcal{R}$, $i = 1,2,\ldots,n$, are the data. We allow the data points to be in general positions in $\mathcal{R}^d$, provided that they are all different, and they may have to satisfy a polynomial unisolvency condition that will be specified later. We apply a usual way of constructing s, which is to choose an n-dimensional linear space, $\mathcal{S}$ say, of functions from $\mathcal{R}^d$ to $\mathcal{R}$. Then one seeks $s\in\mathcal{S}$ that satisfies the interpolation conditions (1.1). It is elementary that this procedure defines the required s uniquely if and only if $\mathcal{S}$ has the property that $s\in\mathcal{S}$ and $s(\underline{x}_i)=0$, $i=1,2,\ldots,n$, imply that s is identically zero.

It is known that this property can fail in the multivariable case $d\geq 2$ if the choice of $\mathcal{S}$ is independent of the positions of the data points. On the other hand, in the simplest form of the radial basis function method the general element of $\mathcal{S}$ has the form

$$s(\underline{x}) = \sum_{j=1}^{n} \lambda_j\, \phi(\|\underline{x}-\underline{x}_j\|_2), \qquad \underline{x}\in\mathcal{R}^d, \tag{1.2}$$

where ϕ is a prescribed function from $\mathcal{R}^+$ to $\mathcal{R}$, and where each λ_j is a real coefficient. It follows that the required s exists and is unique if and only if the $n\times n$ symmetric matrix Φ with the elements

$$\Phi_{ij} = \phi(\|\underline{x}_i-\underline{x}_j\|_2), \qquad 1\leq i,j\leq n, \tag{1.3}$$

is nonsingular. The choice $\phi(r)=e^{-\rho r^2}$, $r\geq 0$, or $\phi(r)=(r^2+\rho^2)^{-1/2}$, $r\geq 0$, where ρ is any positive constant, causes Φ to be not only nonsingular but also positive definite (see Powell, 1992, for instance). Furthermore, when $n\geq 2$, the choice $\phi(r)=r$, $r\geq 0$, provides Φ with one positive and $n-1$ negative eigenvalues, which also implies nonsingularity.

This paper studies a useful generalization of the form (1.2), that depends on a small integer m that satisfies $m\geq -1$. The value $m=-1$ is reserved for the case (1.2) of the previous paragraph, but otherwise we let Π_m be the linear space of polynomials from $\mathcal{R}^d$ to $\mathcal{R}$ of total degree at most m, and we let p_j, $j=1,2,\ldots,\hat{m}$, be a basis of Π_m. Then the general element of $\mathcal{S}$ has the form

$$s(\underline{x}) = \sum_{j=1}^{n} \lambda_j\, \phi(\|\underline{x}-\underline{x}_j\|_2) + \sum_{j=1}^{\hat{m}} \gamma_j\, p_j(\underline{x}), \qquad \underline{x}\in\mathcal{R}^d, \tag{1.4}$$

Radial function	$\phi(r)$	Condition on m	Value of σ
Gaussian	$e^{-\rho r^2}$	$m \geq -1$	$\sigma = 1$
Inverse multiquadric	$(r^2+\rho^2)^{-1/2}$	$m \geq -1$	$\sigma = 1$
Linear	r	$m \geq 0$	$\sigma = -1$
Multiquadric	$(r^2+\rho^2)^{1/2}$	$m \geq 0$	$\sigma = -1$
Thin plate spline	$r^2 \log r$	$m \geq 1$	$\sigma = 1$
Cubic	r^3	$m \geq 1$	$\sigma = 1$

Table 1: Values of σ for several radial basis functions

where the coefficients λ_j, $j=1,2,\ldots,n$, have to satisfy the constraints

$$\sum_{i=1}^{n} \lambda_i \, p_j(\underline{x}_i) = 0, \qquad j=1,2,\ldots,\hat{m}. \tag{1.5}$$

These constraints and the interpolation conditions (1.1) provide the square system of linear equations

$$\left(\begin{array}{c|c} \Phi & P \\ \hline P^T & 0 \end{array} \right) \left(\begin{array}{c} \underline{\lambda} \\ \hline \underline{\gamma} \end{array} \right) = \left(\begin{array}{c} \underline{f} \\ \hline 0 \end{array} \right), \tag{1.6}$$

where Φ has been defined already, where P is the $n \times \hat{m}$ matrix with the elements $P_{ij} = p_j(\underline{x}_i)$, $1 \leq i \leq n$, $1 \leq j \leq \hat{m}$, where the components of $\underline{\lambda}$ and $\underline{\gamma}$ are the coefficients of the function (1.4), and where $\underline{f}$ has the components f_i, $i=1,2,\ldots,n$.

The $(n+\hat{m}) \times (n+\hat{m})$ matrix of the system (1.6) is nonsingular only if the columns of P are linearly independent. Now, for any $\underline{\gamma} \in \mathcal{R}^{\hat{m}}$, the i-th component of the product $P\underline{\gamma}$ is $p(\underline{x}_i)$, where p is the polynomial $\sum_{j=1}^{\hat{m}} \gamma_j \, p_j$. Thus the rank of P is $\hat{m}$ if and only if $p \in \Pi_m$ and $p(\underline{x}_i)=0$, $i=1,2,\ldots,n$, imply $p \equiv 0$. We assume that m is small enough to give this property for the positions of the interpolation points, which is the polynomial unisolvency condition that has been mentioned.

Now the usual choices of ϕ in the radial basis function method are such that, if m is large enough, and if $\underline{\lambda}$ is a nonzero vector in $\mathcal{R}^n$ that satisfies $P^T\underline{\lambda} = 0$, then $\underline{\lambda}^T \Phi \underline{\lambda}$ is nonzero. Further, by continuity, the sign of $\underline{\lambda}^T \Phi \underline{\lambda}$ is independent of $\underline{\lambda}$. We define $\sigma = \pm 1$ by the equation

$$|\underline{\lambda}^T \Phi \underline{\lambda}| = \sigma \, \underline{\lambda}^T \Phi \underline{\lambda}. \tag{1.7}$$

This condition on m and the value of σ for several useful radial basis functions are stated in Table 1, where ρ is still a positive constant. The $m \geq -1$ entries occur when Φ is positive definite, because then the polynomial term of expression (1.4) is optional. Proofs of the assertions in the table are given by Micchelli (1986).

From now on, we assume not only the full rank of P (polynomial unisolvency) but also the condition on m that is stated in the previous paragraph. It follows that the system (1.6) is nonsingular. We prove this assertion by deducing that, if $\underline{f}$ were zero, then $\underline{\lambda}$ and $\underline{\gamma}$ would be zero too. Indeed, by pre-multiplying equation (1.6) by $(\underline{\lambda}^T\,\underline{\gamma}^T)$ and by using $P^T\underline{\lambda}=0$, we find the identity

$$\underline{\lambda}^T\Phi\underline{\lambda} + \underline{\lambda}^T P\underline{\gamma} + \underline{\gamma}^T P^T\underline{\lambda} = \underline{\lambda}^T\Phi\underline{\lambda} = \underline{\lambda}^T\underline{f}. \tag{1.8}$$

Therefore $\underline{f}=0$ implies $\underline{\lambda}^T\Phi\underline{\lambda}=0$, and then the previous paragraph gives $\underline{\lambda}=0$. Moreover, $\Phi\underline{\lambda}+P\underline{\gamma}=\underline{f}$ with $\underline{\lambda}=0$ and $\underline{f}=0$ give $\underline{\gamma}=0$ by polynomial unisolvency, which completes the proof. Thus the interpolation conditions (1.1) are satisfied uniquely by a function of the form (1.4) subject to the constraints (1.5). Further, the set of functions s that can occur for different right hand side vectors $\underline{f}\in\mathcal{R}^n$ is an n-dimensional linear space, namely the space $\mathcal{S}$ of functions of the form (1.4) whose coefficients are subject to the constraints (1.5).

These assumptions allow the system (1.6) to be solved by a direct method, that usually gives excellent accuracy even when Φ is highly ill-conditioned, and that was proposed originally for thin plate spline radial functions by Sibson and Stone (1991). Specifically, we let $\mathcal{Z}$ be the $(n-\hat{m})$-dimensional linear subspace of $\mathcal{R}^n$ such that $\underline{\lambda}$ is in $\mathcal{Z}$ if and only if its components satisfy the conditions (1.5), and we let Z be an $n\times(n-\hat{m})$ real matrix whose columns are a basis of $\mathcal{Z}$. Then the $\underline{\lambda}$ that is defined by the system (1.6) can be written as $\underline{\lambda}=Z\underline{\mu}$ for some $\underline{\mu}$ in $\mathcal{R}^{n-\hat{m}}$. Further, because Z^TP is zero, the first part of the system (1.6) provides the relation

$$(\sigma\, Z^T\Phi Z)\,\underline{\mu} = \sigma\, Z^T\underline{f}. \tag{1.9}$$

One chooses Z so that it is not expensive to calculate all the elements of the symmetric matrix $\sigma\, Z^T\Phi Z=W$, say. It follows from the paragraph that includes equation (1.7) that W is positive definite. Therefore the direct method solves the system (1.9) for $\underline{\mu}$ by calculating the Cholesky factorization of W. Then the formula $\underline{\lambda}=Z\underline{\mu}$ gives the radial basis function coefficients of the required interpolant (1.4). Finally, the polynomial coefficients of s can be derived conveniently from the following construction. By multiplying the first part of expression (1.6) by P^T, we obtain the square system of normal equations

$$(P^TP)\,\underline{\gamma} = P^T(\underline{f}-\Phi\underline{\lambda}), \tag{1.10}$$

which defines the coefficients γ_j, $j=1,2,\ldots,\hat{m}$. Ill-conditioning can be avoided in the matrix P^TP by a careful choice of the basis p_j, $j=1,2,\ldots,\hat{m}$.

The amount of work of this procedure for solving the system (1.6) is $\mathcal{O}(n^3)$ operations, which restricts the number of data points to about one thousand. Therefore, because of the very wide range of applications of the radial basis function method (see Hardy, 1990, for instance), there has been much research on iterative techniques for calculating the coefficients of the interpolant (1.4). In particular, the procedure of Dyn, Levin and Rippa (1986) is analogous to the solution of the system (1.9) by the conjugate gradient algorithm, after picking Z in a clever way that assists the conditioning of $Z^T \Phi Z$. Another approach is suggested by Beatson and Powell (1994). It is based on the Lagrange form

$$s(\underline{x}) = \sum_{j=1}^{n} f_j \, \ell_j(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d, \tag{1.11}$$

of the solution of the interpolation equations (1.1), where ℓ_j is the s that would occur if the right hand sides had the values $f_i = \delta_{ij}$, $i=1,2,\ldots,n$. Each ℓ_j in formula (1.11) is replaced by an estimate of the Lagrange function that employs only a small subset of the interpolation points $\underline{x}_i$, $i=1,2,\ldots,n$, so the formula provides an approximation to s. The error of the approximation is corrected by iterative refinement.

Huge improvements to these methods have been developed during the last six years by Rick Beatson, George Goodsell and the authors. It is usual for the new algorithms to solve the interpolation equations (1.1) to eight decimal places of accuracy in fewer than ten iterations for general positions of the data points, even when n is large. The main task of each iteration is the calculation of the values $s(\underline{x}_i)$, $i = 1,2,\ldots,n$, for a function s of the form (1.4). Therefore fast multipole methods are highly useful when they can be applied (see Beatson and Newsam, 1992, for instance). The research in Cambridge provided a procedure of this kind about three years ago (Powell, 1997), s being the current estimate of the interpolant, because the residuals $f_i - s(\underline{x}_i)$, $i=1,2,\ldots,n$, are required. Further, many intermediate residuals were calculated. Then George Goodsell found experimentally that in most cases much work can be saved by replacing the intermediate residuals by the ones that occur at the beginning of the iteration. Occasionally that method fails to converge, so a Krylov subspace technique has been added to it. Further, we now regard the Krylov construction as the main part of the method, so the most successful development of the previous five years reduces to a particular choice of a self-adjoint operator, namely A. Thus the description of the recommended algorithm becomes easier. Therefore we are going to address the algorithm and its convergence properties from this point of view.

The Krylov subspace method requires a norm or a semi-norm for the n-dimensional linear space $\mathcal{S}$, which is the subject of Section 2. We will find there that the native semi-norm of Schaback (1993) is very suitable. Then Section 3 considers the Krylov technique when the operator A is general, including conditions on A that are sufficient for convergence, and that allow an implementation that is analogous to the conjugate gradient method. The particular A and some other details of the recommended algorithm are specified in Section 4. Finally, in Section 5, the efficiency of the given procedure is investigated numerically on some test problems, and some plans for future work are mentioned.

2. The semi-norm of $\mathcal{S}$

Let the functions (1.4) and

$$t(\underline{x}) = \sum_{j=1}^{n} \mu_j \, \phi(\|\underline{x}-\underline{x}_j\|_2) + \sum_{j=1}^{\hat{m}} \delta_j \, p_j(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d, \tag{2.1}$$

be elements of the n-dimensional linear space $\mathcal{S}$ that has been defined already. Therefore the radial basis function coefficients satisfy the constraints

$$\sum_{i=1}^{n} \lambda_i \, p_j(\underline{x}_i) = \sum_{i=1}^{n} \mu_i \, p_j(\underline{x}_i) = 0, \qquad j=1,2,\ldots,\hat{m}. \tag{2.2}$$

We introduce the bilinear form

$$\langle s,t\rangle = \sigma \, \underline{\lambda}^T \Phi \underline{\mu}, \tag{2.3}$$

where Φ is still the matrix (1.3). The symmetry of Φ implies the condition $\langle t,s\rangle = \langle s,t\rangle$. Moreover, the property (1.7) shows that $\langle s,s\rangle$ is nonnegative for every s in $\mathcal{S}$. Therefore, by defining expression (2.3) to be the scalar product between s and t, and by defining the semi-norm

$$\|s\| = \langle s,s\rangle^{1/2}, \qquad s \in \mathcal{S}, \tag{2.4}$$

the space $\mathcal{S}$ becomes a semi-Hilbert space. We will use the notation (2.3) and (2.4) extensively.

It follows from the paragraph that includes equation (1.7) that, if the function (1.4) is an element of $\mathcal{S}$, then $\|s\|$ is zero if and only if all the coefficients λ_j, $j=1,2,\ldots,n$, are zero. In other words, the semi-norm of $s \in \mathcal{S}$ is zero if and only if s is in the space Π_m of polynomials of degree at most m. Here we take the view that Π_{-1} contains just the zero function, because the semi-Hilbert space becomes a Hilbert space in the case $m=-1$.

Our algorithm requires a reformulation of expression (2.3). Specifically, equations (2.3), (1.3), (1.5) and (2.1) imply the identities

$$\begin{aligned}
\langle s,t\rangle &= \sigma \sum_{i=1}^{n} \lambda_i \Big\{ \sum_{j=1}^{n} \phi(\|\underline{x}_i-\underline{x}_j\|_2) \, \mu_j \Big\} \\
&= \sigma \sum_{i=1}^{n} \lambda_i \Big\{ \sum_{j=1}^{n} \mu_j \, \phi(\|\underline{x}_i-\underline{x}_j\|_2) + \sum_{j=1}^{\hat{m}} \delta_j \, p_j(\underline{x}_i) \Big\} \\
&= \sigma \sum_{i=1}^{n} \lambda_i \, t(\underline{x}_i) \;=\; \sigma \sum_{i=1}^{n} \mu_i \, s(\underline{x}_i),
\end{aligned} \tag{2.5}$$

the last assertion being valid due to symmetry. One advantage of this remark is that the matrix Φ will not occur in the description of the algorithm. Another advantage is that, if the function (1.4) is a general element of $\mathcal{S}$ and if $s^\star$ is the required interpolant, then the interpolation conditions provide the formula

$$\langle s, s^\star \rangle = \sigma \sum_{i=1}^{n} \lambda_i \, s^\star(\underline{x}_i) = \sigma \sum_{i=1}^{n} \lambda_i \, f_i, \qquad s \in \mathcal{S}. \tag{2.6}$$

Therefore the value of $\langle s, s^\star \rangle$ can be calculated before the coefficients of $s^\star$ are known, which is highly useful.

The semi-norm has a fundamental property that provides excellent corroboration for the claim that it is very suitable for the Krylov subspace method. The property is taken from Schaback (1993), and will not be employed by our algorithm, because it applies to a linear space of functions $\mathcal{S}$ that is larger than the one that has been defined already. Indeed, just for the remainder of this section, we let $\hat{s}$ be in $\mathcal{S}$ if it has the form

$$\hat{s}(\underline{x}) = \sum_{j=1}^{\hat{n}} \hat{\lambda}_j \, \phi(\|\underline{x} - \underline{\hat{x}}_j\|_2) + \sum_{j=1}^{\hat{m}} \hat{\gamma}_j \, p_j(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d, \tag{2.7}$$

the parameters being subject to the usual constraints

$$\sum_{i=1}^{\hat{n}} \hat{\lambda}_i \, p_j(\underline{\hat{x}}_i) = 0, \qquad j = 1, 2, \ldots, \hat{m}. \tag{2.8}$$

Here ϕ and p_j, $j = 1, 2, \ldots, \hat{m}$, are as before, but $\hat{n}$ can be much larger than n. The only restriction on the set $\{\underline{\hat{x}}_i : i = 1, 2, \ldots, \hat{n}\}$ is that it has to include all the original data points $\underline{x}_i$, $i = 1, 2, \ldots, n$. It follows that the previous $\mathcal{S}$ is an n-dimensional linear subspace of the new $\mathcal{S}$, that the polynomial unisolvency condition is preserved, and that at least one element of the new $\mathcal{S}$ interpolates the data. Further, $\hat{s} \in \mathcal{S}$ is an interpolant if and only if it can be written in the form $\hat{s} = s^\star + \hat{t}$, where $s^\star$ is still the required interpolant, and where $\hat{t} \in \mathcal{S}$ satisfies $\hat{t}(\underline{x}_i) = 0$, $i = 1, 2, \ldots, n$.

Now the analogue of expression (2.3) for the new space $\mathcal{S}$ is the bilinear form

$$\langle \hat{s}, \hat{t} \rangle = \sigma \sum_{i=1}^{\hat{n}} \sum_{j=1}^{\hat{n}} \hat{\lambda}_i \, \phi(\|\underline{\hat{x}}_i - \underline{\hat{x}}_j\|_2) \, \hat{\mu}_j, \tag{2.9}$$

where $\hat{s}$ and $\hat{t}$ are the functions (2.7) and

$$\hat{t}(\underline{x}) = \sum_{j=1}^{\hat{n}} \hat{\mu}_j \, \phi(\|\underline{x} - \underline{\hat{x}}_j\|_2) + \sum_{j=1}^{\hat{m}} \hat{\delta}_j \, p_j(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d, \tag{2.10}$$

respectively. Thus there is no change to the previous value of $\langle \hat{s}, \hat{t} \rangle$ if $\hat{s}$ and $\hat{t}$ happen to be in the original space $\mathcal{S}$. Moreover, the definition (2.4) of the semi-norm is valid for the new space $\mathcal{S}$.

The fundamental property that supports the choice (2.4) is that, if $\hat{s}$ is any function in the new space $\mathcal{S}$ that satisfies the interpolation equations (1.1), then $\langle \hat{s}, \hat{s} \rangle$ is never less than $\langle s^\star, s^\star \rangle$. We prove this assertion by writing $\hat{s} = s^\star + \hat{t}$ as before, and by considering the scalar product

$$\langle s^\star, \hat{t} \rangle = \sigma \sum_{i=1}^{\hat{n}} \hat{\lambda}_i^\star \, \hat{t}(\underline{\hat{x}}_i), \tag{2.11}$$

where $\hat{\lambda}_i^\star$, $i=1,2,\ldots,\hat{n}$, are the radial basis function coefficients of $s^\star$, and where the right hand side is taken from the identity (2.5). The coefficient $\hat{\lambda}_i^\star$ is nonzero only if $\underline{\hat{x}}_i$ is one of the given interpolation points, and then $\hat{t}(\underline{\hat{x}}_i)$ is zero. Thus expression (2.11) is zero, which gives the required bound

$$\langle \hat{s}, \hat{s} \rangle = \langle s^\star + \hat{t}, s^\star + \hat{t} \rangle = \langle s^\star, s^\star \rangle + \langle \hat{t}, \hat{t} \rangle \geq \langle s^\star, s^\star \rangle. \tag{2.12}$$

This bound is a generalization of the well-known minimum second derivative property of thin plate spline interpolation in two dimensions (Duchon, 1977), if one restricts attention to functions in the new space $\mathcal{S}$. From now on, however, we revert to the original space $\mathcal{S}$ that has dimension n.

3. The Krylov subspace method

The Krylov subspace method is iterative and depends on a linear operator A from $\mathcal{S}$ to $\mathcal{S}$ and on the semi-norm (2.4). It is similar to the GMRES algorithm of Saad and Schultz (1986), except that A need not depend on the interpolation equations and the semi-norm need not be a norm. We let k be the iteration number of the method, and we let $\mathcal{S}_k$ be the linear subspace of $\mathcal{S}$ that is spanned by the functions $A^j s^\star$, $j=1,2,\ldots,k$, where $s^\star$ is still the required interpolant. If the k-th iteration is not the final one, then its main task is to calculate an element of $\mathcal{S}_k$, say s_{k+1}, that satisfies the condition

$$\|s^\star - s_{k+1}\| \leq \|s^\star - s\|, \qquad s \in \mathcal{S}_k, \tag{3.1}$$

for the semi-norm (2.4).

An implementation of this method will be described that has the following useful properties for the choice of A that is given in Section 4. If there are no computer rounding errors, and if a tolerance parameter is set to zero, then the required interpolant $s^\star$ is obtained. The sequence of semi-norms $\|s^\star - s_j\|$, $j=1,2,\ldots,k^\star$, decreases strictly monotonically, where s_1 is the zero function and where $k^\star$ is the number of the final iteration. Condition (3.1) is achieved for every integer k in $[1, k^\star - 1]$. The coefficients of the functions $A^j s^\star$, $j=1,2,\ldots,k$, are not found explicitly, because these functions tend to provide a basis of $\mathcal{S}_k$ that is highly ill-conditioned. Instead, for $k=1,2,\ldots,k^\star$, the function s_{k+1} is calculated from s_k by applying the formula

$$s_{k+1} = s_k + \alpha_k\, d_k \tag{3.2}$$

in a way that is analogous to an iteration of the conjugate gradient method. Specifically, d_k is the search direction

$$d_k = A\,(s^\star - s_k) + \beta_k\, d_{k-1}, \tag{3.3}$$

the multiplier $\beta_k \in \mathcal{R}$ being defined by the orthogonality condition $\langle d_k, d_{k-1}\rangle = 0$, except that the $\beta_k\, d_{k-1}$ term is deleted when $k=1$, so $s_1 = 0$ implies $d_1 = As^\star$. The step-length $\alpha_k \in \mathcal{R}$ of formula (3.2) is derived from inequality (3.1). The calculation ends when the function values $s_{k+1}(\underline{x}_i)$, $i=1,2,\ldots,n$, are sufficiently close to the data f_i, $i=1,2,\ldots,n$, the termination condition being the requirement

$$|s_{k+1}(\underline{x}_i) - f_i| \le \varepsilon, \qquad i=1,2,\ldots,n, \tag{3.4}$$

where $\varepsilon \ge 0$ is the tolerance parameter that has been mentioned. Therefore every iteration, before employing the step-length that provides the property (3.1), tries to choose α_k in equation (3.2) to satisfy inequality (3.4), the calculation being completed if and only if the attempt is successful. It is necessary for A to be such that the attempt gives termination in the case $\|s^\star - s_k\| = 0$, because then the usual reduction $\|s^\star - s_{k+1}\| < \|s^\star - s_k\|$ cannot be attained. The construction of the search directions provides the orthogonality conditions

$$\langle d_k, d_j\rangle = 0, \qquad j=1,2,\ldots,k-1, \tag{3.5}$$

for every integer k in the interval $[2, k^\star - 1]$.

Instead of showing that the remarks of the previous paragraph are valid for the particular A that is specified in Section 4, we are going to establish that it is sufficient if A has the properties

$$\left.\begin{array}{llll} s\in\mathcal{S} \text{ and } s\neq 0 & \Rightarrow & As\neq 0 & \text{(nonsingularity)} \\ s\in\Pi_m & \Rightarrow & As=s & \text{(polynomial reproduction)} \\ s\in\mathcal{S} \text{ and } s\notin\Pi_m & \Rightarrow & \langle s, As\rangle > 0 & \text{(ellipticity)} \\ s\in\mathcal{S} \text{ and } t\in\mathcal{S} & \Rightarrow & \langle t, As\rangle = \langle At, s\rangle & \text{(self-adjointness)} \end{array}\right\}. \tag{3.6}$$

Then only these properties will have to be verified in Section 4.

If the required interpolant $s^\star$ is in the space Π_m, then polynomial reproduction and $s_1 = 0$ imply that the first iteration generates the direction $d_1 = As^\star = s^\star$. Thus the choice $\alpha_1 = 1$ in formula (3.2) gives $s_2 = s^\star$. It follows that the termination condition (3.4) is achieved on the first iteration, even if ε is zero.

Alternatively, in the case $s^\star \notin \Pi_m$, termination occurs on the first iteration if ε is sufficiently large, but usually s_2 is calculated by satisfying condition (3.1) for

$k=1$. Specifically, s_2 is the function $\alpha_1 d_1$, where d_1 is the direction $As^\star$, and where α_1 minimizes the scalar product

$$\langle s^\star - \alpha_1 d_1, s^\star - \alpha_1 d_1\rangle = \|s^\star - s_2\|^2, \qquad \alpha_1 \in \mathcal{R}. \tag{3.7}$$

Now $s^\star \notin \Pi_m$ implies $d_1 = As^\star \notin \Pi_m$, because otherwise A would map not only Π_m but also $s^\star$ into Π_m, which would contradict nonsingularity. Therefore $\langle d_1, d_1\rangle$ is positive, as stated in the second paragraph of Section 2. It follows that the minimization of expression (3.7) defines $\alpha_1 = \langle s^\star, d_1\rangle / \langle d_1, d_1\rangle$ uniquely. Further, α_1 is positive, because $s^\star \notin \Pi_m$, $d_1 = As^\star$ and the ellipticity condition (3.6) provide $\langle s^\star, d_1\rangle > 0$. Hence, if $k^\star \geq 2$, we deduce the strict inequality

$$\|s^\star - s_2\| < \|s^\star\| = \|s^\star - s_1\|. \tag{3.8}$$

For $k \geq 2$, we are going to prove the assertions in the second paragraph of this section by induction. Therefore we assume that they hold if k is replaced by $k-1$. Thus s_k and d_{k-1} are in the space $\mathcal{S}_{k-1}$. It follows from the definition of $\mathcal{S}_k$ and from equations (3.3) and (3.2) that, for any choice of β_k and α_k, both d_k and s_{k+1} are in the space $\mathcal{S}_k$. Further, because the inductive hypothesis includes the relations $s_k = s_{k-1} + \alpha_{k-1} d_{k-1}$ and $\|s^\star - s_k\| < \|s^\star - s_{k-1}\|$, the value of $\langle d_{k-1}, d_{k-1}\rangle$ is positive. Hence equation (3.3) and the orthogonality condition $\langle d_k, d_{k-1}\rangle = 0$ define the parameter β_k uniquely.

The possibility $\|s^\star - s_k\| = 0$ can occur. Then $s^\star - s_k$ is an element of Π_m, which is nonzero because the calculation was not terminated by the previous iteration. Moreover, the polynomial reproduction condition (3.6) shows that the search direction (3.3) is the function

$$d_k = s^\star - s_k + \beta_k d_{k-1}. \tag{3.9}$$

Now the Cauchy–Schwarz inequality

$$|\langle s^\star - s_k, d_{k-1}\rangle| \leq \|s^\star - s_k\| \, \|d_{k-1}\| \tag{3.10}$$

and $\|s^\star - s_k\| = 0$ provide $\langle s^\star - s_k, d_{k-1}\rangle = 0$, while the definition of β_k provides $\langle d_k, d_{k-1}\rangle = 0$. It follows from expression (3.9) that β_k is zero. Hence equation (3.2) takes the form $s_{k+1} = s_k + \alpha_k (s^\star - s_k)$. Thus the termination condition (3.4) is achieved by the choice $\alpha_k = 1$. It is also achieved when $\|s^\star - s_k\|$ is nonzero if the tolerance parameter ε is sufficiently large. Our analysis of the case $k = k^\star$ is complete, so the conditions $\|s^\star - s_k\| > 0$ and $2 \leq k < k^\star$ are assumed for the remainder of the inductive argument.

We recall that the step-length α_k of formula (3.2) is chosen to minimize the function

$$\|s^\star - s_{k+1}\|^2 = \langle s^\star - s_k - \alpha_k d_k, s^\star - s_k - \alpha_k d_k\rangle, \qquad \alpha_k \in \mathcal{R}. \tag{3.11}$$

We are going to deduce from the assumption $\|s^\star - s_k\| > 0$ that α_k is positive. We see that, if α_k is well-defined, then it has the value $\langle s^\star - s_k, d_k\rangle / \langle d_k, d_k\rangle$. Further, equation (3.3) gives the identity

$$\langle s^\star - s_k, d_k\rangle = \langle s^\star - s_k, A(s^\star - s_k)\rangle + \beta_k \langle s^\star - s_k, d_{k-1}\rangle. \tag{3.12}$$

Now $\|s^\star - s_k\| > 0$ is equivalent to $(s^\star - s_k) \notin \Pi_m$, so the ellipticity condition of expression (3.6) shows that $\langle s^\star - s_k, A(s^\star - s_k)\rangle$ is positive. Moreover, $d_{k-1} \in \mathcal{S}_{k-1}$ and the inductive hypothesis

$$\|s^\star - s_k\| \le \|s^\star - s\|, \qquad s \in \mathcal{S}_{k-1}, \tag{3.13}$$

imply that the least value of $\|s^\star - s_k - \theta\, d_{k-1}\|$, $\theta \in \mathcal{R}$, occurs at $\theta = 0$, so the scalar product $\langle s^\star - s_k, d_{k-1}\rangle$ is zero. Thus the identity (3.12) implies $\langle s^\star - s_k, d_k\rangle > 0$, and then the Cauchy–Schwarz inequality provides $\langle d_k, d_k\rangle > 0$. It follows that the algorithm generates a positive step-length as required. The property $\alpha_k \neq 0$ with the construction of α_k yield the strict inequality

$$\|s^\star - s_{k+1}\| < \|s^\star - s_k\|, \tag{3.14}$$

which establishes another of the assertions that are being addressed.

Equation (3.5) is satisfied when $j = k-1$, because the choice of β_k provides $\langle d_k, d_{k-1}\rangle = 0$. Therefore, if $k \ge 3$, we have to show that $\langle d_k, d_j\rangle$ is zero for $1 \le j \le k-2$. Definition (3.3) gives the identity

$$\langle d_k, d_j\rangle = \langle A(s^\star - s_k), d_j\rangle + \beta_k \langle d_{k-1}, d_j\rangle, \tag{3.15}$$

and the inductive hypothesis includes expression (3.5) when k is replaced by $k-1$. Thus, using the self-adjointness condition (3.6), we deduce the equation

$$\langle d_k, d_j\rangle = \langle s^\star - s_k, A d_j\rangle, \qquad j = 1, 2, \ldots, k-2. \tag{3.16}$$

Now $d_j \in \mathcal{S}_j$ implies $A d_j \in \mathcal{S}_{j+1}$, so $A d_j$ is an element of $\mathcal{S}_{k-1}$ for $j \le k-2$. It follows from the hypothesis (3.13) that the least value of $\|s^\star - s_k - \theta A d_j\|$, $\theta \in \mathcal{R}$, occurs at $\theta = 0$. Therefore the scalar product (3.16) is zero, which completes the proof of assertion (3.5).

We now know that the search directions $d_j \in \mathcal{S}_k$, $j = 1, 2, \ldots, k$, are mutually orthogonal with respect to the scalar product of the semi-Hilbert space $\mathcal{S}$. Further, because the Krylov subspace definition of $\mathcal{S}_k$ implies $\dim(\mathcal{S}_k) \le k$, these directions are a basis of $\mathcal{S}_k$. Therefore the least value of the right hand side of expression (3.1) is achieved when s is the function $s = \sum_{j=1}^{k} \theta_j\, d_j$, where the coefficients $\theta_j \in \mathcal{R}$, $j = 1, 2, \ldots, k$, minimize the quadratic form

$$\Big\| s^\star - \sum_{j=1}^{k} \theta_j\, d_j \Big\|^2 = \langle s^\star, s^\star\rangle - 2 \sum_{j=1}^{k} \theta_j \langle s^\star, d_j\rangle + \sum_{j=1}^{k} \theta_j^2 \langle d_j, d_j\rangle. \tag{3.17}$$

Hence the coefficients of the optimal s take the unique values

$$\theta_j^\star = \langle s^\star, d_j \rangle / \langle d_j, d_j \rangle, \qquad j=1,2,\ldots,k. \tag{3.18}$$

Similarly, the inductive hypothesis (3.13) implies that s_k is the function $\sum_{j=1}^{k-1} \theta_j^\star d_j$. It follows that $\|s^\star - s_k - \theta_k^\star d_k\|$ is the least value of $\|s^\star - s\|$, $s \in \mathcal{S}_k$. Now the given method sets $s_{k+1} = s_k + \alpha_k d_k$, where α_k minimizes $\|s^\star - s_k - \alpha_k d_k\|$, $\alpha_k \in \mathcal{R}$. Thus $s_{k+1} = s_k + \theta_k^\star d_k$ must occur, which provides the required property (3.1).

We also have to prove that the given method terminates in exact arithmetic. Therefore, for any iteration number $k < k^\star$, we consider the search directions d_j, $j=1,2,\ldots,k$, and the basis p_j, $j=1,2,\ldots,\hat{m}$, of Π_m. We pick coefficients θ_j, $j=1,2,\ldots,k$, and γ_j, $j=1,2,\ldots,\hat{m}$, such that $s = \sum_{j=1}^{k} \theta_j d_j + \sum_{j=1}^{\hat{m}} \gamma_j p_j$ is the zero element of $\mathcal{S}$. The mutual orthogonality of the search directions and the form (2.5) of the scalar product provide the equation

$$\langle s, s \rangle = \sum_{j=1}^{k} \theta_j^2 \langle d_j, d_j \rangle = 0, \tag{3.19}$$

for general values of γ_j, $j=1,2,\ldots,\hat{m}$. Because our analysis gives $\langle d_j, d_j \rangle > 0$, $j=1,2,\ldots,k$, it follows that every θ_j is zero. Then we deduce $\gamma_j = 0$, $j = 1,2,\ldots,\hat{m}$, from $s \equiv 0$, as the polynomials p_j, $j=1,2,\ldots,\hat{m}$, are a basis of Π_m. Therefore we have found $k+\hat{m}$ linearly independent elements of $\mathcal{S}$, which implies $k+\hat{m} \le \dim(\mathcal{S}) = n$ for each $k < k^\star$. In other words, assuming exact arithmetic, termination occurs on an iteration whose number $k^\star$ is at most $n-\hat{m}+1$. The justification of the remarks in the second paragraph of this section is complete.

The nonsingularity condition (3.6) on A is important, because the functions that can be calculated by the Krylov subspace method are restricted to the set $\{As : s \in \mathcal{S}\}$. Indeed, if A were singular, then the unique solution $s^\star$ of the interpolation equations would not be in this set for most values of the right hand sides f_i, $i=1,2,\ldots,n$. The polynomial reproduction condition (3.6) is also important, because it provides a way of generating the correct polynomial part of $s^\star$. This task is not done by the techniques of the method that are designed to achieve inequality (3.1), because the definition (2.4) of the semi-norm ignores any polynomial terms of s. The purpose of the ellipticity condition (3.6) is to give the method the property $s_{k+1} \neq s_k$, where s_{k+1} is defined by condition (3.1). If $s_{k+1} = s_k$ occurred, however, then formula (3.3) would be unsuitable for generating the search direction d_{k+1}, because the relation

$$A(s^\star - s_{k+1}) = A(s^\star - s_k) = d_k - \beta_k d_{k-1} \tag{3.20}$$

would put d_{k+1} in the linear space that is spanned by the first k search directions. The self-adjointness condition (3.6) implies by induction that all the orthogonality properties (3.5) are satisfied when d_k has the simple form (3.3). Only one change has to be made to the given method if this condition fails, namely that the definition (3.3) is replaced by the equation

$$d_k = A(s^\star - s_k) - \sum_{j=1}^{k-1} \langle d_j, A(s^\star - s_k)\rangle \, \langle d_j, d_j\rangle^{-1} \, d_j, \tag{3.21}$$

the sum being deleted when $k=1$. The authors have studied the use of this formula in a Krylov subspace method, because each iteration of the algorithm in Powell (1997) can be regarded as an operator A from $\mathcal{S}$ to $\mathcal{S}$, and this operator possesses only the first three of the properties (3.6).

The polynomial reproduction and ellipticity conditions imply nonsingularity. Specifically, given any nonzero $s \in \mathcal{S}$, we ask if it is in Π_m. If the answer is affirmative, then polynomial reproduction provides $As = s \neq 0$. Otherwise, ellipticity gives $\langle s, As\rangle > 0$, which shows $As \neq 0$ as required.

4. The recommended algorithm

The proof of termination in the previous section is misleading, because it suggests that the number of iterations of the Krylov subspace method may be close to n. Instead, as mentioned in Section 1, fewer than ten iterations are usual when the recommended algorithm is applied in practice, even when n becomes large. Therefore we seem to have made an excellent choice of the operator A, which will be addressed in this section. We begin by deriving this A when there are no polynomial terms ($m = -1$), so the general element of $\mathcal{S}$ has the form (1.2). This derivation will make it easy to allow $m \geq -1$ later.

When $m = -1$, we consider the choice

$$As = \sum_{j=1}^{n} \langle \psi_j, s\rangle \, \langle \psi_j, \psi_j\rangle^{-1} \, \psi_j, \qquad s \in \mathcal{S}, \tag{4.1}$$

where ψ_j, $j = 1, 2, \ldots, n$, is a basis of $\mathcal{S}$. The self-adjointness of A is shown by the symmetry of the right hand side of the equation

$$\langle t, As\rangle = \sum_{j=1}^{n} \langle \psi_j, s\rangle \, \langle \psi_j, \psi_j\rangle^{-1} \, \langle \psi_j, t\rangle. \tag{4.2}$$

There are no polynomials to be reproduced except for the trivial case of the zero function, and the following argument establishes ellipticity. It is sufficient to prove that $\langle s, As\rangle = 0$ implies $s = 0$. We make use of the remark that expression (2.4) is a norm in the case $m = -1$. We let $s = \sum_{j=1}^{n} \theta_j \, \psi_j$ be any element of $\mathcal{S}$ such that the scalar product

$$\langle s, As\rangle = \sum_{j=1}^{n} \langle \psi_j, s\rangle^2 / \, \langle \psi_j, \psi_j\rangle \tag{4.3}$$

is zero. It follows that $\langle \psi_j, s\rangle$ is zero for every j. Therefore $\|s\|^2$ has the value

$$\langle s, s\rangle = \langle \textstyle\sum_{j=1}^n \theta_j\, \psi_j, s\rangle = \displaystyle\sum_{j=1}^n \theta_j\, \langle \psi_j, s\rangle = 0, \qquad (4.4)$$

which provides $s=0$ as required. Further, we find at the end of Section 3 that the ellipticity condition implies that A is nonsingular.

We achieve good efficiency in the recommended algorithm by trying to make A close to the identity operator. Therefore it is relevant that equation (4.1) provides $A=I$, where I is the identity, if the basis functions ψ_j, $j=1,2,\ldots,n$, are mutually orthogonal with respect to the scalar product (2.3). Indeed, in this case the definition (4.1) gives $A\psi_\ell = \psi_\ell$, $\ell = 1,2,\ldots,n$, which implies $As = s$, $s\in\mathcal{S}$, because ψ_j, $j=1,2,\ldots,n$, is a basis of $\mathcal{S}$. The advantage of $A=I$ in the Krylov subspace method is that, on the first iteration, equation (3.2) takes the form $s_2 = \alpha_1\, s^\star$. Thus, because the value $\alpha_1 = 1$ is allowed, termination occurs immediately. Therefore we require a convenient way of choosing orthogonal, or nearly orthogonal, basis functions.

A method is suggested by the form (2.5) of the scalar product $\langle s,t\rangle$, where s and t are the functions (1.4) and (2.1). We see that t is orthogonal to s if all of the products $\mu_i\, s(\underline{x}_i)$, $i=1,2,\ldots,n$, are zero. Therefore, if $m=-1$ and the matrix (1.3) is positive definite, then, for every integer j in $[1,n]$, we define ψ_j to be the function of the form

$$\psi_j(\underline{x}) = \sum_{\ell=j}^n \lambda_{j\ell}\, \phi(\|\underline{x}-\underline{x}_\ell\|_2), \qquad \underline{x}\in\mathcal{R}^d, \qquad (4.5)$$

that satisfies the Lagrange equations

$$\psi_j(\underline{x}_\ell) = \delta_{\ell j}, \qquad \ell=j,j+1,\ldots,n. \qquad (4.6)$$

Thus $\langle \psi_i, \psi_j\rangle$ is zero for $j>i$, because the least index of the sum (4.5) is $\ell=j$, and all of the values $\psi_i(\underline{x}_\ell)$, $\ell=j,j+1,\ldots,n$, are zero. A development of this construction and a reformulation of equation (4.1) will provide the recommended choice of A.

A development is needed, because the calculation of the coefficients $\lambda_{j\ell}$, $j\le \ell\le n$, of expression (4.5) is comparable to the solution of the original interpolation problem when there are $n-j+1$ data, and there are many values of j. Therefore we seek a modification of the given functions ψ_j, $j=1,2,\ldots,n$, that reduces the number of coefficients greatly by allowing mild departures from the orthogonality properties $\langle \psi_i, \psi_j\rangle = 0$, $i\ne j$. Our research has shown that it is highly suitable to employ Lagrange function estimates of the kind that are proposed by Beatson and Powell (1994). Specifically, we pick a positive integer q, the value $q=30$ being typical, and then, for each j, we let ψ_j have the form

$$\psi_j(\underline{x}) = \sum_{\ell\in\mathcal{L}_j} \lambda_{j\ell}\, \phi(\|\underline{x}-\underline{x}_\ell\|_2), \qquad \underline{x}\in\mathcal{R}^d, \qquad (4.7)$$

instead of expression (4.5), where the set $\mathcal{L}_j$ contains j and at most $q-1$ of the integers $j+1, j+2, \ldots, n$. Further, the coefficients of the new ψ_j are defined by the Lagrange conditions

$$\psi_j(\underline{x}_\ell) = \delta_{\ell j}, \qquad \ell \in \mathcal{L}_j. \tag{4.8}$$

We let $\mathcal{L}_j$ be the set $\{j, j+1, \ldots, n\}$ if j satisfies $n-j+1 \le q$, but otherwise we prefer to include in $\mathcal{L}_j$ the q different integers from $[j,n]$ that minimize the Euclidean distances $\|\underline{x}_\ell - \underline{x}_j\|_2$, $\ell \in \mathcal{L}_j$. Then equations (4.1), (4.7) and (4.8) define the recommended operator A in the case $m=-1$.

In order to prove that this A has the required properties (3.6), we ask whether the new functions ψ_j, $j=1,2,\ldots,n$, are linearly independent. Let Γ be the $n \times n$ matrix with the elements $\Gamma_{j\ell}$, $1 \le j, \ell \le n$, where $\Gamma_{j\ell}$ is 0 or $\lambda_{j\ell}$ if $\ell \notin \mathcal{L}_j$ or $\ell \in \mathcal{L}_j$, respectively. It follows from equation (4.7) and the nonsingularity of Φ that linear independence occurs if and only if Γ is nonsingular. The diagonal elements of Γ are nonzero, because equations (2.5), (4.7) and (4.8) imply the strict inequality

$$0 < \langle \psi_j, \psi_j \rangle = \sigma \sum_{\ell \in \mathcal{L}_j} \lambda_{j\ell} \psi_j(\underline{x}_\ell) = \sigma \, \lambda_{jj}, \qquad j=1,2,\ldots,n. \tag{4.9}$$

Moreover, Γ is upper triangular due to the choice of the sets $\mathcal{L}_j$, $j=1,2,\ldots,n$. It follows that Γ is nonsingular. Therefore linear independence is achieved, so the remarks of the second paragraph of this section hold, which establish the conditions (3.6).

In the reformulation of the recommended A when $m=-1$, we preserve the sets $\mathcal{L}_j$, $j=1,2,\ldots,n$, and equations (4.7) and (4.8), but we write the definition (4.1) as the expression

$$As = \sum_{j=1}^{n-q} \langle \psi_j, s \rangle \, \langle \psi_j, \psi_j \rangle^{-1} \psi_j + Hs, \qquad s \in \mathcal{S}, \tag{4.10}$$

where H is the operator

$$Hs = \sum_{j=n-q+1}^{n} \langle \psi_j, s \rangle \, \langle \psi_j, \psi_j \rangle^{-1} \psi_j, \qquad s \in \mathcal{S}. \tag{4.11}$$

Thus H is a linear mapping from $\mathcal{S}$ to $\mathcal{T}$, where $\mathcal{T}$ is the q-dimensional linear subspace of $\mathcal{S}$ that contains functions of the form

$$t(\underline{x}) = \sum_{j=n-q+1}^{n} \lambda_j \, \phi(\|\underline{x} - \underline{x}_j\|_2), \qquad \underline{x} \in \mathcal{R}^d. \tag{4.12}$$

The reason for giving special attention to H is that it has the property

$$Ht = t, \qquad t \in \mathcal{T}. \tag{4.13}$$

We begin the proof of this assertion by noting that, for every j in $[n-q+1, n]$, the choice of $\mathcal{L}_j$ makes the functions (4.5) and (4.7) identical. Therefore the mutual orthogonality of the functions ψ_j, $j=n-q+1, n-q+2, \ldots, n$, is preserved. Hence the definition (4.11) implies $H\psi_\ell=\psi_\ell$, $\ell=n-q+1, n-q+2, \ldots, n$, so the required condition (4.13) follows from the remark that these functions ψ_ℓ are a basis of $\mathcal{T}$.

The property (4.13) provides a view of H that is highly useful. Indeed, we deduce from equation (4.11) and the scalar product (2.5) that the dependence of Hs on s is through the values $s(\underline{x}_i)$, $i=n-q+1, n-q+2, \ldots, n$. Therefore $Hs=Ht$ occurs, where t is the unique element of $\mathcal{T}$ that satisfies the equations

$$t(\underline{x}_i) = s(\underline{x}_i), \qquad i=n-q+1, n-q+2, \ldots, n. \tag{4.14}$$

Further, the identity (4.13) shows $Ht=t$. Thus the term Hs of expression (4.10) is the function of the form (4.12) whose coefficients are defined by the interpolation conditions (4.14).

This remark is important for $m\geq -1$, because the operator (4.11) is degenerate in the cases with $m\geq 0$. Indeed, the denominator $\langle \psi_n, \psi_n \rangle$ becomes zero, because, when $j=n$ and $\mathcal{L}_j=\{n\}$, equation (4.8) is satisfied by the constant polynomial $\psi_n(\underline{x})=1$, $\underline{x}\in\mathcal{R}^d$. Therefore, for the remainder of this section, we let H be the mapping from $\mathcal{S}$ to $\mathcal{T}$ such that Hs, $s\in\mathcal{S}$, is the element of $\mathcal{T}$ of the form

$$t(\underline{x}) = \sum_{j=n-q+1}^{n} \lambda_j \, \phi(\|\underline{x}-\underline{x}_j\|_2) + \sum_{j=1}^{\hat{m}} \gamma_j \, p_j(\underline{x}), \qquad \underline{x}\in\mathcal{R}^d, \tag{4.15}$$

that is defined by the interpolation conditions (4.14). Further, for $j=1, 2, \ldots, n$–q, we let ψ_j be the element of $\mathcal{S}$ that has the form

$$\psi_j(\underline{x}) = \sum_{\ell\in\mathcal{L}_j} \lambda_{j\ell} \, \phi(\|\underline{x}-\underline{x}_\ell\|_2) + \sum_{\ell=1}^{\hat{m}} \gamma_{j\ell} \, p_\ell(\underline{x}), \qquad \underline{x}\in\mathcal{R}^d, \tag{4.16}$$

and that satisfies the Lagrange conditions (4.8). Then expression (4.10) is the recommended choice of A for $m \geq -1$. The sets $\mathcal{L}_j$, $j = 1, 2, \ldots, n-q$, are as before, except for a new constraint on the points $\underline{x}_\ell$, $\ell \in \mathcal{L}_j$, to ensure that the denominator $\langle \psi_j, \psi_j \rangle$ of expression (4.10) is positive. In other words, the solution (4.16) of the equations (4.8) must not be an element of Π_m. Therefore the new constraint is that the polynomial unisolvency condition must hold on the set $\{\underline{x}_\ell : \ell\in\mathcal{L}_j\backslash\{j\}\}$ for every integer j in $[1, n-q]$. Thus the case $j=n-q$ ensures that H is well-defined. Our description of the recommended A is complete, but some further remarks on the choice of $\mathcal{L}_j$, $j=1, 2, \ldots, n-q$, and on the ordering of the interpolation points, will be made in Section 5.

We have to show that the conditions (3.6) hold for this A. If s is any element of Π_m, then the scalar products $\langle \psi_j, s \rangle$, $j=1, 2, \ldots, n$–q, are zero, so the definition (4.10) gives $As=Hs$. Moreover, $s\in\Pi_m$ implies $s\in\mathcal{T}$, so the unique solution of the equations (4.14) is $t=Hs=s$. Therefore the polynomial reproduction condition $As=s$, $s\in\Pi_m$, is achieved.

Our proof of the other conditions involves the scalar product $\langle Hs, H\hat{s}\rangle$, where s and $\hat{s}$ are general elements of $\mathcal{S}$. Because Hs is in $\mathcal{T}$, it has the form

$$(Hs)(\underline{x}) = \sum_{j=n-q+1}^{n} \nu_j\, \phi(\|\underline{x}-\underline{x}_j\|_2) + \text{element of } \Pi_m, \qquad \underline{x}\in\mathcal{R}^d. \tag{4.17}$$

Moreover, the definition of H provides $(H\hat{s})(\underline{x}_i)=\hat{s}(\underline{x}_i)$, $i=n-q+1, n-q+2, \ldots, n$. Therefore, by applying expression (2.5) twice, we deduce the identity

$$\langle Hs, H\hat{s}\rangle = \sigma \sum_{i=n-q+1}^{n} \nu_i\, \{(H\hat{s})(\underline{x}_i)\} = \sigma \sum_{i=n-q+1}^{n} \nu_i\, \hat{s}(\underline{x}_i) = \langle Hs, \hat{s}\rangle. \tag{4.18}$$

It follows from the symmetry of $\langle Hs, H\hat{s}\rangle$ that H has the self-adjoint property $\langle Hs, \hat{s}\rangle = \langle s, H\hat{s}\rangle$. Further, remembering our use of equation (4.2), we find that the operator (4.10) is self-adjoint, as required.

Next we let s be any element of $\mathcal{S}$ that is not in Π_m, and we consider the value of $\langle s, As\rangle$. Equations (4.10) and (4.18) provide the formula

$$\langle s, As\rangle = \sum_{j=1}^{n-q} \langle \psi_j, \psi_j\rangle^{-1} \langle \psi_j, s\rangle^2 + \|Hs\|^2, \tag{4.19}$$

and we wish to show that $\langle s, As\rangle$ is positive. We see that the ellipticity condition $\langle s, As\rangle > 0$ holds if $\|Hs\| \neq 0$, so we turn our attention to the alternative case when Hs is in Π_m. The function $\hat{s}=s-Hs$ is nonzero due to $s\notin\Pi_m$ and $Hs\in\Pi_m$. Hence there exists a greatest integer j in $[1,n]$ such that $\hat{s}(\underline{x}_j)\neq 0$. It satisfies $j\leq n-q$, because Hs interpolates s at the last q data points. Therefore $\langle\psi_j,\psi_j\rangle^{-1}\langle\psi_j, s\rangle^2$ is one of the terms in the sum of expression (4.19). Further, $Hs \in \Pi_m$ implies $\langle\psi_j, \hat{s}\rangle = \langle\psi_j, s\rangle$. It follows that $\langle s, As\rangle > 0$ is achieved if $\langle\psi_j, \hat{s}\rangle$ is nonzero. Now $\langle\psi_j,\psi_j\rangle$ is positive, and the derivation of inequality (4.9) provides $\langle\psi_j,\psi_j\rangle = \sigma\,\lambda_{jj}$ when ψ_j is the function (4.16). Thus, because the choice of j gives $\hat{s}(\underline{x}_\ell)=0$, $\ell=j+1, j+2, \ldots, n$, equations (2.5) and (4.16) with the form of $\mathcal{L}_j$ imply the required property

$$\langle\psi_j, \hat{s}\rangle = \sigma \sum_{\ell\in\mathcal{L}_j} \lambda_{j\ell}\, \hat{s}(\underline{x}_\ell) = \sigma\,\lambda_{jj}\, \hat{s}(\underline{x}_j) = \langle\psi_j,\psi_j\rangle\, \hat{s}(\underline{x}_j) \neq 0, \tag{4.20}$$

which completes the proof of the ellipticity condition (3.6). The remaining condition, namely nonsingularity, is deduced at the end of Section 3.

The preliminary work of the algorithm is the choice of $\mathcal{L}_j$, and the calculation of the coefficients of the function (4.16), for every integer j in $[1, n-q]$. Further, because the interpolation equations (4.14) are solved by the direct procedure of the paragraph that includes expressions (1.9) and (1.10), the $(q-\hat{m})\times(q-\hat{m})$ matrix

W of this procedure is formed and factorized. These operations are completed before the iterative method of Section 3 is begun. The time that they require in our numerical experiments is only of order n, because we have restricted attention to $d=2$, and then Dirichlet tessellations provide a convenient way of generating the sets $\mathcal{L}_j$ (Goodsell, 1997). Here we are assuming that a small value of q, say $q=30$, remains adequate if n becomes large.

The iterative part of the algorithm calculates the coefficients of several functions in $\mathcal{S}$. Therefore we let $\underline{\lambda}(s) \in \mathcal{R}^n$ and $\underline{\gamma}(s) \in \mathcal{R}^{\hat{m}}$ be the vectors whose components are the coefficients of the general element (1.4). At the beginning of the k-th iteration of the Krylov subspace method, the vectors $\underline{\lambda}(s_k)$ and $\underline{\gamma}(s_k)$, which define s_k, and the residuals

$$r_i^{(k)} = f_i - s_k(\underline{x}_i), \qquad i=1,2,\ldots,n, \tag{4.21}$$

are available, their values being obvious when $k=1$, because s_1 is the zero function. Further, $\underline{\lambda}(d_{k-1})$ and $\underline{\gamma}(d_{k-1})$, and the values $d_{k-1}(\underline{x}_i)$, $i=1,2,\ldots,n$, are also available when $k \geq 2$. The first task of the k-th iteration is to generate the coefficients of the term $A(s^\star - s_k)$ of the search direction (3.3), which requires only $\mathcal{O}(nq)$ operations. Indeed, we put $s=s^\star - s_k$ in the definition (4.10) of A, we write each scalar product $\langle \psi_j, s^\star - s_k \rangle$ in the form $\sigma \sum_{\ell \in \mathcal{L}_j} \lambda_{j\ell} r_\ell^{(k)}$, and we make use of the property $|\mathcal{L}_j|=q$, $j=1,2,\ldots,n-q$. When $k=1$, the required components of $\underline{\lambda}(d_k)$ and $\underline{\gamma}(d_k)$ are just the coefficients of $A(s^\star - s_k)$, because the $\beta_k d_{k-1}$ term of the search direction (3.3) is suppressed on the first iteration. Alternatively, when $k \geq 2$, the orthogonality condition $\langle d_k, d_{k-1} \rangle = 0$ gives the value

$$\beta_k = -\frac{\langle d_{k-1}, z_k \rangle}{\langle d_{k-1}, d_{k-1} \rangle} = -\frac{\sum_{i=1}^n \lambda(z_k)_i \, d_{k-1}(\underline{x}_i)}{\sum_{i=1}^n \lambda(d_{k-1})_i \, d_{k-1}(\underline{x}_i)}, \tag{4.22}$$

where $z_k = A(s^\star - s_k)$, and where $\lambda(s)_i$ denotes the i-th component of $\underline{\lambda}(s)$, $s \in \mathcal{S}$. Thus β_k is found in $\mathcal{O}(n)$ operations, and the algorithm applies the formulae

$$\underline{\lambda}(d_k) = \underline{\lambda}(z_k) + \beta_k \, \underline{\lambda}(d_{k-1}) \quad \text{and} \quad \underline{\gamma}(d_k) = \underline{\gamma}(z_k) + \beta_k \, \underline{\gamma}(d_{k-1}). \tag{4.23}$$

When n is large, the most expensive part of an iteration is the calculation of all the values $s(\underline{x}_i)$, $i=1,2,\ldots,n$, for a particular s in $\mathcal{S}$. Two attractive alternatives are $s=d_k$ and $s=s_k+d_k$. We prefer $s=d_k$, because it avoids complications if rounding errors cause substantial inconsistencies between the residuals (4.21) and the coefficients of s_k. Therefore the numbers

$$d_k(\underline{x}_i) = \sum_{j=1}^n \lambda(d_k)_j \, \phi(\|\underline{x}_i - \underline{x}_j\|_2) + \sum_{j=1}^{\hat{m}} \gamma(d_k)_j \, p_j(\underline{x}_i), \qquad i=1,2,\ldots,n, \tag{4.24}$$

are computed. One of their uses is the evaluation of new residuals, because equations (3.2) and (4.21) provide the formula

$$r_i^{(k+1)} = f_i - s_{k+1}(\underline{x}_i) = r_i^{(k)} - \alpha_k \, d_k(\underline{x}_i), \qquad i=1,2,\ldots,n. \tag{4.25}$$

Thus the algorithm seeks $\alpha_k \in \mathcal{R}$ that satisfies the termination condition (3.4). If termination does not occur, however, then we recall from Section 3 that the property (3.1) implies the step-length

$$\alpha_k = \langle s^\star - s_k, d_k \rangle \,/\, \langle d_k, d_k \rangle = \sum_{i=1}^{n} \lambda(d_k)_i \, r_i^{(k)} \Big/ \sum_{i=1}^{n} \lambda(d_k)_i \, d_k(\underline{x}_i). \qquad (4.26)$$

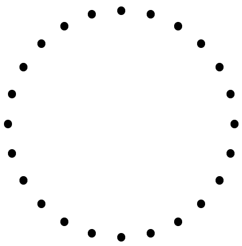
Expressions (4.25) and (4.26) give the residuals for the next iteration if $k < k^\star$. Finally, the procedure observes equation (3.2) by setting the coefficient vectors $\underline{\lambda}(s_{k+1})$ and $\underline{\gamma}(s_{k+1})$ to $\underline{\lambda}(s_k)+\alpha_k\,\underline{\lambda}(d_k)$ and $\underline{\gamma}(s_k)+\alpha_k\,\underline{\gamma}(d_k)$, respectively. Then k is increased by one if another iteration is required.

This version of the algorithm has the property that the coefficients of s_k contribute only to the coefficients of s_{k+1}. Thus any errors in these coefficients are inherited by the next iteration. Therefore, when high accuracy is required, it is advisable to generate the values $s_{k^\star}(\underline{x}_i)$, $i=1,2,\ldots,n$, using the coefficients that occur at termination. If the differences $f_i - s_{k^\star}(\underline{x}_i)$, $i=1,2,\ldots,n$, are unacceptably large, then the estimate of $s^\star$ can be improved by iterative refinement, using the Krylov subspace method again to compute a correction to $s_{k^\star}$.

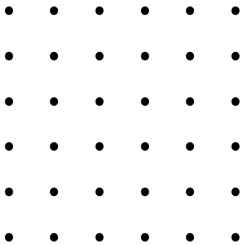
5. Numerical results and further developments

Four test problems will be considered in this section, the value of d being two in each case. The positions of the interpolation points are shown schematically in Figure 1. Specifically, in Problem 1 the points $\underline{x}_i$, $i=1,2,\ldots,n$, are equally spaced on a circle, in Problem 2 the points are the vertices of a square grid, in Problem 3 they are generated randomly on the unit disc, and in Problem 4 they are equally spaced on two concentric circular arcs, where the ratio of the distance between the arcs to the smaller radius is only 10^{-5} in practice, but a greater distance is shown in the figure. In all of the calculations, the right hand sides f_i, $i=1,2,\ldots,n$, are random numbers from the uniform distribution on $[-1,1]$.

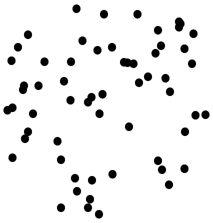
These test problems were highly useful in the development of the given algorithm. Indeed, the perfect symmetry of Problem 1 makes it easy for expression (4.16) to be a good approximation to a Lagrange function when $|\mathcal{L}_j|$ is much less than n, which provided much encouragement during the early research of Beatson and Powell (1994). Problem 2 was frustrating, however, when we were seeking estimates of Lagrange functions that have small errors at every data point, because such estimates may not exist for large n and a moderate value of $|\mathcal{L}_j|$, in the case when the true Lagrange function is one at a corner of the grid. The successful solution to this difficulty was the discovery of an algorithm that does not require small errors at the data points, provided that the scalar products $\langle \psi_i, \psi_j \rangle$, $i \neq j$, are sufficiently close to zero, but we did not know about the relevance of the scalar product of Section 2 when the new algorithm was found. That algorithm



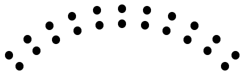
Problem 1



Problem 2



Problem 3



Problem 4

Figure 1: Positions of the data points

was developed for the square grid and could be applied for general positions of the data points. Then numerical tests with the random data points of Problem 3 gave the brilliant result that no further modifications were needed to achieve good efficiency. The thin plate spline radial function was used throughout that work, and we looked for an example of divergence or some other unacceptable behaviour of the iterative procedure. Thus the data of Problem 4 became of interest, because they caused the rate of convergence to be very slow. They also caused the failures, mentioned in Section 1, that were corrected by the Krylov subspace technique.

Tables 2 and 3 provide some numerical results for the linear radial function $\phi(r)=r$ with $m=0$ and the thin plate spline radial function $\phi(r)=r^2 \log r$ with $m=1$. The main entries of the tables are the numbers of iterations that occur when the tolerance parameter of the termination condition (3.4) is given the value $\varepsilon = 10^{-8}$. The results show overwhelmingly that the proof of termination in at most $n{-}\hat{m}$ iterations is irrelevant to practical computation. Further, they support the claim in Section 1 that it is usual to achieve eight places of accuracy in fewer than ten iterations, because $q=30$ should be preferred to $q=10$. On the other hand, our experimental work so far has addressed only interpolation problems in two variables.

If the values $\alpha_k=1$ and $\beta_k=0$ are set on every iteration, then expressions (3.2)

n	q	Problem 1	Problem 2	Problem 3	Problem 4
	10	5	15	13	5
400	30	4	8	6	5
	50	4	7	5	5
	10	5	15	14	5
900	30	4	7	7	5
	50	4	7	6	5

Table 2: Numbers of iterations when $\phi(r)=r$

n	q	Problem 1	Problem 2	Problem 3	Problem 4
	10	7	27	33	70
400	30	4	10	8	42
	50	4	8	7	29
	10	7	28	42	87
900	30	4	10	10	61
	50	4	8	8	51

Table 3: Numbers of iterations when $\phi(r)=r^2\log r$

and (3.3) imply the formula

$$s_{k+1} = s_k + A\,(s^\star - s_k), \qquad k=1,2,3,\ldots\,. \tag{5.1}$$

The test problems of Figure 1 were used to construct algorithms that satisfy this equation, and then the Krylov subspace technique was added less than a year ago. Because equation (5.1) gives the relation

$$s_{k+1} - s^\star = (I-A)\,(s_k - s^\star), \qquad k=1,2,3,\ldots, \tag{5.2}$$

those algorithms have linear rates of convergence that depend on the spectral radius of $I-A$. The addition of the Krylov subspace technique was found to be highly successful when the positions of the interpolation points $\underline{x}_i$, $i=1,2,\ldots,n$, cause the spectral radius of $I-A$ to exceed 0.5, but in other cases there is usually little change to the number of iterations, whatever the value of the tolerance parameter ε. For example, for the first three test problems of Table 3 with $q\geq 30$, the technique reduced the number of iterations by at most one third. Therefore, when the method of Section 3 requires few iterations in practice, we expect the simple procedure (5.1) to be efficient too. Further, if the right hand sides f_i,

$i=1,2,\ldots,n$, are general, then the spectral radius of $I-A$ may provide a good estimate of the number of iterations of the Krylov subspace algorithm, especially when the operator A is self-adjoint.

These remarks suggest correctly that there is little advantage in including the Krylov subspace technique when solving most radial basis function interpolation problems. On the other hand, the technique provides not only a guarantee of convergence in theory but also large gains in efficiency in difficult cases. Further, one may not be able to tell in advance if a case is difficult, and the extra work per iteration of the technique is trivial. In order to assess the extra work, we take the view that, instead of setting $\beta_k=0$ and $\alpha_k=1$ in equations (3.3) and (3.2), the technique prefers the values (4.22) and (4.26), which have to be calculated. Now both procedures require the coefficients of the function $z_k = A\,(s^\star - s_k)$ for the search direction (3.3), which takes $\mathcal{O}(nq)$ operations, but we deduce from equations (4.22), (3.3), (4.26) and (3.2) that the extra work of the Krylov subspace technique is only $\mathcal{O}(n)$ operations. Further, after generating the coefficients $\lambda(d_k)_j$, $j=1,2,\ldots,n$, and $\gamma(d_k)_j$, $j=1,2,\ldots,\hat{m}$, of d_k, both procedures apply formula (4.24), which is an $\mathcal{O}(n^2)$ calculation, except that in some cases the fast multipole method replaces formula (4.24) by an $\mathcal{O}(n\log n)$ procedure. Thus the additional computation of the Krylov subspace technique is negligible. Therefore there are strong reasons for adding it to algorithms that apply the simple formula (5.1), if the required interpolating function is in a linear space $\mathcal{S}$ that has the semi-norm of Section 2, and if A satisfies the conditions (3.6).

The relatively large number of iterations in the last column of Table 3 is a topic for future research. That inefficiency was found before the results of Table 2 were computed, so the excellent performance of the linear radial function $\phi(r)=r$ was unexpected. A major difference between the calculations of the two tables occurs in the sets $\mathcal{L}_j$, $j=1,2,\ldots,n-q$. When $\phi(r)=r$, each $\mathcal{L}_j$ contains the q different integers from $[j,n]$ that minimize the distances $\|\underline{x}_\ell-\underline{x}_j\|_2$, $\ell\in\mathcal{L}_j$, as suggested in Section 4. When $\phi(r)=r^2\log r$, however, that choice of $\mathcal{L}_j$ may be unacceptable, because the polynomial unisolvency constraint states that, for $d=2$ and $m=1$, the points x_ℓ, $\ell\in\mathcal{L}_j\backslash\{j\}$, must not be collinear. We responded to this requirement by picking three data points initially that are far from collinear, by reordering the data so that these points are $\underline{x}_{n-2}$, $\underline{x}_{n-1}$ and $\underline{x}_n$, and by forcing the integers $n-2$, $n-1$ and n to be in all of the sets $\mathcal{L}_j$, $j=1,2,\ldots,n-q$. The remaining elements of each $\mathcal{L}_j$ are selected as before by minimizing $\|\underline{x}_\ell-\underline{x}_j\|_2$, $\ell\in\mathcal{L}_j$. If future investigations show that a different choice of sets can reduce greatly the iteration counts in Table 3 for Problem 4, then we will seek a form of the new choice that can be applied for general positions of the interpolation points.

Another question for future research is the dependence of the number of iterations on changes to the ordering of the data $\underline{x}_i$, $i=1,2,\ldots,n$. A decision on the ordering was taken for the first algorithm with approximations to Lagrange functions that was tested on Problem 3, and that ordering has been retained. It is based on the assumption that, if points are removed in sequence from the beginning of the set $\underline{x}_i$, $i=1,2,\ldots,n$, then the surviving points should provide good coverage of the original set in $\mathcal{R}^d$. Therefore, in a preliminary calculation, the integer j runs through the values $1,2,\ldots,n-q$, and, for each j, $\underline{x}_j$ is exchanged with one of the points $\underline{x}_{j+1},\underline{x}_{j+2},\ldots,\underline{x}_{\hat{n}}$, if necessary, so that the nearest neighbour distance $\min\{\|\underline{x}_\ell-\underline{x}_j\|_2 : \ell=j+1,j+2,\ldots,n\}$ becomes as small as possible, where $\hat{n}=n$ is usual, but the technique of the last paragraph for thin plate splines requires $\hat{n}=n-3$. This device is included in the calculations of Tables 2 and 3. Recently, however, we tried some experiments with random orderings of the data points, using the linear radial function and the test problems of Table 2. We found no appreciable changes to the iteration counts. Therefore we intend to explore the idea of preferring an ordering that is helpful to the task of generating the sets $\mathcal{L}_j$, $j=1,2,\ldots,n-q$, automatically. This task can be done conveniently by Dirichlet tessellations in two dimensions, but some assistance from the ordering of the data may be important to efficiency for $d\geq 3$.

Most of the coding of the computer program that generated the results of Tables 2 and 3 is for the construction of Dirichlet tessellations, because it is easy to implement the iterative part of the algorithm of Section 4, if the numbers (4.24) are calculated directly instead of by the fast multipole method. The computer software will be available from the authors after attention has been given to the proposed future work that is sketched in the previous two paragraphs.

The software will be able to solve the interpolation equations using linear and multiquadric radial functions without a polynomial term, although these $m=-1$ cases are absent from Table 1. Indeed, one calculates $s^\star$ in the way that has been described with $m=0$, and also one calculates $\hat{s}^\star$, say, in this way by interpolation to the right hand sides $\phi(\|\underline{x}_i-\underline{x}_\ell\|_2)$, $i=1,2,\ldots,n$, where ℓ is now a prescribed integer from $[1,n]$. Then the function

$$\check{s}^\star(\underline{x}) = s^\star(\underline{x}) + \hat{c}\,[\hat{s}^\star(\underline{x})-\phi(\|\underline{x}-\underline{x}_\ell\|_2)], \qquad \underline{x}\in\mathcal{R}^d, \tag{5.3}$$

interpolates the data for every choice of $\hat{c}\in\mathcal{R}$, because the expression in square brackets vanishes at all the interpolation points. Further, the constant term of this expression must be nonzero, because the matrix (1.3) is nonsingular. Thus a unique value of $\hat{c}$ annihilates the constant term of $\check{s}^\star$, which forces $\check{s}^\star$ to be the required interpolant.

The low iteration counts of Tables 2 and 3 suggest that the given Krylov subspace technique will become highly useful for radial basis function interpolation, when n is so large that the $\mathcal{O}(n^3)$ cost of direct methods is prohibitive. Thus the technique may increase greatly the range of applications that can take advantage of the smoothness and accuracy properties of radial functions. Franke (1982) states that these properties compare very favourably with those of interpolating functions from other linear spaces. Further, the radial basis function method is particularly useful for interpolation on nonlinear mainfolds in $\mathcal{R}^d$, because the nonsingularity of the system (1.6) imposes hardly any (or no) restrictions on the positions of the data points $\underline{x}_i$, $i=1,2,\ldots,n$.

Acknowledgement

The work of George Goodsell on radial basis function interpolation was of major importance to the development of the given algorithm. It led to the operator A that is recommended in Section 4. Further, he wrote most of the Fortran software that has been used for numerical calculations.

References

R.K. Beatson and G.N. Newsam (1992), "Fast evaluation of radial basis functions: I", *Comput. Math. Applic.*, Vol. 24, pp. 7–19.

R.K. Beatson and M.J.D. Powell (1994), "An iterative method for thin plate spline interpolation that employs approximations to Lagrange functions", in *Numerical Analysis 1993*, eds. D.F. Griffiths and G.A. Watson, Longman Scientific & Technical (Burnt Mill), pp. 17–39.

J. Duchon (1977), "Splines minimizing rotation-invariant seminorms in Sobolev spaces", in *Constructive Theory of Functions of Several Variables, Lecture Notes in Mathematics 571*, eds. W. Schempp and K. Zeller, Springer-Verlag (Berlin), pp. 85–100.

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

R. Franke (1982), "Scattered data interpolation: tests of some methods", *Math. Comp.*, Vol. 38, pp. 181–200.

G. Goodsell (1997), "On finding p-th nearest neighbours of scattered points in two dimensions for small p", Report No. DAMTP 1997/NA01, University of Cambridge (to be published in *Computer Aided Geometric Design*).

R.L. Hardy (1990), "Theory and applications of the multiquadric-biharmonic method", *Comput. Math. Applic.*, Vol. 19, pp. 163–208.

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

M.J.D. Powell (1992), "The theory of radial basis function approximation in 1990", in *Advances in Numerical Analysis Volume II: Wavelets, Subdivision Algorithms, and Radial Basis Functions*, ed. W. Light, Clarendon Press (Oxford), pp. 105–210.

M.J.D. Powell (1997), "A new iterative method for thin plate spline interpolation in two dimensions", *Annals of Numerical Mathematics*, Vol. 4, pp. 519–527.

Y. Saad and M.H. Schultz (1986), "GMRES: A generalized minimum residual algorithm for solving nonsymmetric linear systems", *SIAM J. Sci. Statist. Comput.*, Vol. 7, pp. 856–869.

R. Schaback (1993), "Comparison of radial basis function interpolants", in *Multivariate Approximations: From CAGD to Wavelets*, eds. K. Jetter and F. Utreras, World Scientific (Singapore), pp. 293–305.

R. Sibson and G. Stone (1991), "Computation of thin plate splines", *SIAM J. Sci. Statist. Comput.*, Vol. 12, pp. 1304–1313.