DAMTP 1998/NA05

Recent research at Cambridge on radial basis functions[1]

M.J.D. Powell

Abstract: Much of the research at Cambridge on radial basis functions during the last four years has addressed the solution of the thin plate spline interpolation equations in two dimensions when the number of interpolation points, n say, is very large. It has provided some techniques that will be surveyed because they allow values of n up to 10^5, even when the positions of the points are general. A close relation between these techniques and Newton's interpolation method is explained. Another subject of current research is a new way of calculating the global minimum of a function of several variables. It is described briefly, because it employs a semi-norm of a large space of radial basis functions. Further, it is shown that radial basis function interpolation minimizes this semi-norm in a way that is a generalisation of the well-known variational property of thin plate spline interpolation in two dimensions. The final subject is the deterioration in accuracy of thin plate spline interpolation near the edges of finite grids. Several numerical experiments in the one-dimensional case are reported that suggest some interesting conjectures that are still under investigation.

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

June, 1998.

[1]Presented at the 2nd International Dortmund Meeting on Approximation Theory.

1. Introduction

Let the values $f_i = f(\underline{x}_i)$, $i=1,2,\ldots,n$, of a real valued function $f(\underline{x})$, $\underline{x} \in \mathcal{R}^d$, be given, where the points $\underline{x}_i \in \mathcal{R}^d$, $i=1,2,\ldots,n$, are all different. A radial basis function interpolant to the data has the form

$$s(\underline{x}) = \sum_{j=1}^{n} \lambda_j \, \phi(\|\underline{x}-\underline{x}_j\|) + p(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d, \tag{1.1}$$

for some function $\phi(r)$, $r \geq 0$, where the vector norm is Euclidean and where p is a polynomial of degree at most m, say. Therefore m is a nonnegative integer, except that we employ the notation $m=-1$ to denote the important case when p has to be identically zero. For $m \geq 0$, we express p as the sum

$$p(\underline{x}) = \sum_{j=1}^{\hat{m}} c_j \, \pi_j(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d, \tag{1.2}$$

where π_j, $j=1,2,\ldots \hat{m}$, is a basis of the linear space of polynomials of degree at most m from $\mathcal{R}^d$ to $\mathcal{R}$. After choosing ϕ and m, the coefficients of s are required to satisfy the $(n+\hat{m}) \times (n+\hat{m})$ system of linear equations

$$\left(\begin{array}{c|c} \Phi & P \\ \hline P^T & 0 \end{array} \right) \left(\begin{array}{c} \underline{\lambda} \\ \hline \underline{c} \end{array} \right) = \left(\begin{array}{c} \underline{f} \\ \hline 0 \end{array} \right), \tag{1.3}$$

where Φ and P are the $n \times n$ and $n \times \hat{m}$ matrices that have the elements $\Phi_{ij} = \phi(\|\underline{x}_i - \underline{x}_j\|)$ and $P_{ij} = \pi_j(\underline{x}_i)$, respectively. Further, $\underline{\lambda} \in \mathcal{R}^n$ and $\underline{c} \in \mathcal{R}^{\hat{m}}$ are the vectors of coefficients of s, and the components of $\underline{f}$ are the data f_i, $i=1,2,\ldots,n$. Thus the first n rows of expression (1.3) are the interpolation conditions $s(\underline{x}_i) = f_i$, $i=1,2,\ldots,n$.

We assume throughout this paper that, if $m \geq 1$, then the data points $\underline{x}_i$, $i=1,2,\ldots,n$, are such that the rank of P is $\hat{m}$, because otherwise the system (1.3) would be singular. In particular, if $m=1$ and $d=2$, the assumption is that the data points are not collinear. Then the crucial property of the radial basis function method is that several choices of $\phi(r)$, $r \geq 0$, are available such that the equations (1.3) have a unique solution, with no other restrictions on the data points (Micchelli, 1986), except that $n \geq 2$ is required occasionally for $m=-1$. Further, for each of these choices of ϕ there is an integer m_0 such that, if $m \geq m_0$ and if $\underline{v}$ is any nonzero vector that satisfies $P^T \underline{v} = 0$, then $\underline{v}^T \Phi \underline{v}$ is nonzero and its sign is opposite to that of $(-1)^{m_0}$, where $\underline{v}$ can be any nonzero vector in the case $m_0 = -1$. The most popular choices of ϕ are the Gaussian $\phi(r) = \exp(-c\,r^2)$, the multiquadric $\phi(r) = (r^2 + c^2)^{1/2}$ and the thin plate spline $\phi(r) = r^2 \log r$, c being a positive constant. The values of m_0 in these cases are -1, 0 and 1, respectively.

The nonsingularity of the system (1.3) when the data points are in general positions provides an excellent reason for studying the radial basis function method.

Further, the approximation $s \approx f$ is highly useful in a wide range of practical applications (Hardy, 1990), because accuracy and smoothness properties are achieved that compare favourably with those of other methods (Franke, 1982). Many interesting theoretical questions occur too. Some stunning answers are derived from Fourier transforms by Buhmann (1990) when the data points form a regular infinite grid, but, for finite n, it is usual for numerical experiments to suggest orders of accuracy that have not been confirmed theoretically. Therefore the subject is still an active field of research at Cambridge. Three of the topics that are under investigation will be considered in the remaining three sections of this paper.

The first of them is an iterative procedure for solving the equations (1.3) when n is very large, in the important special case of thin plate spline interpolation in two dimensions. Therefore $d=2$ and $m=1$ hold, and ϕ is the function

$$\phi(r) = r^2 \log r, \qquad r \geq 0, \tag{1.4}$$

the value of $\phi(0)$ being zero. The work began about six years ago, as reported by Beatson and Powell (1994), and a breakthrough for data points in general positions is described by Powell (1997). Thus, by applying the Laurent expansion technique of Beatson and Newsam (1992) to generate the product $\Phi\underline{\lambda}$ for trial values of $\underline{\lambda}$, it is possible for n to exceed 10^4, although there is no useful sparsity in Φ. Another major advance was made by Goodsell last year, which reduces substantially the number of products that have to be calculated on each iteration. Thus a close relation between the procedure and Newton's interpolation formula was noticed, that assists a description of the new method. These findings are presented in Section 2.

Section 3 addresses an application of the radial basis function method to global optimization, the least value of $f(\underline{x})$, $\underline{x} \in \mathcal{X}$, being required, where $\mathcal{X}$ is a subset of $\mathcal{R}^d$. We take the view that it is expensive to calculate $f(\underline{x})$ for each choice of $\underline{x}$. Therefore we let the data $f_i = f(\underline{x}_i)$, $i = 1, 2, \ldots, n$, be all the values of the objective function that are known already, and we consider the crucial question of deciding on the position of $\underline{x}_{n+1}$, which is the vector of variables for the next calculation of f. A way of choosing this point is recommended. It depends on a measure of smoothness that is derived from the properties of Φ that are stated in the second paragraph of this section. Specifically, we find that the nonnegativity of $(-1)^{m_0+1}\underline{\lambda}^T\Phi\underline{\lambda}$ provides a useful semi-norm of the interpolant (1.1), and that there is an associated scalar product that assists the calculation of $\underline{x}_{n+1}$ by the recommended procedure. This semi-norm and scalar product belong to the large linear space that is studied by Schaback (1993).

The final topic includes some interesting examples of orders of accuracy that are suggested by numerical experiments. Specifically, we consider thin plate spline interpolation in only one dimension when the data points x_i, $i = 0, 1, \ldots, n$, are spaced uniformly throughout the interval $[0, 1]$. The numerical results indicate the accuracy

$$\max\{|f(x)-s(x)| : 0 \le x \le 1\} = \mathcal{O}(h^{3/2}) \tag{1.5}$$

when f is smooth, where h is the spacing between data points, although $\mathcal{O}(h^3)$ accuracy occurs away from the ends of the range (Bejancu, 1997). Therefore, as suggested by Beatson and Powell (1992) for multiquadrics, we investigate the replacement of the $P^T\underline{\lambda} = 0$ part of expression (1.3) by two other conditions that take up the degrees of freedom that remain in the coefficients of s after the interpolation equations are satisfied. Thus it is shown numerically in Section 4 that it is possible to reduce the right hand side of expression (1.5) from $\mathcal{O}(h^{3/2})$ to $\mathcal{O}(h^{5/2})$ when f is a general smooth function. Work has begun on trying to establish these curious orders of convergence analytically.

2. An iterative procedure for interpolation

The new procedure of this section is analogous to the following version of Newton's method for polynomial interpolation in one dimension. Let the interpolation conditions be $s(x_i) = f_i$, $i = 1, 2, \ldots, n$, where x_i, $i = 1, 2, \ldots, n$, are points of $\mathcal{R}$ that are all different, and where s is now required to be a polynomial of degree at most $n{-}1$ from $\mathcal{R}$ to $\mathcal{R}$. We introduce the Lagrange functions $\chi_n(x) = 1$, $x \in \mathcal{R}$, and

$$\chi_k(x) = \prod_{j=k+1}^{n} \frac{x - x_j}{x_k - x_j}, \qquad x \in \mathcal{R}, \qquad k = 1, 2, \ldots, n-1, \tag{2.1}$$

and we write s in the form

$$s(x) = \sum_{k=1}^{n} \mu_k \, \chi_k(x), \qquad x \in \mathcal{R}. \tag{2.2}$$

It follows that the partial sum

$$s_\ell(x) = \sum_{k=\ell}^{n} \mu_k \, \chi_k(x), \qquad x \in \mathcal{R}, \tag{2.3}$$

is the unique polynomial of degree at most $n-\ell$ that interpolates the last $n-\ell+1$ data, where ℓ is any integer in $[1, n]$. Further, μ_ℓ is the coefficient of $x^{n-\ell}$ in s_ℓ divided by the coefficient of $x^{n-\ell}$ in χ_ℓ. Thus it is possible to calculate the parameters μ_k, $k = 1, 2, \ldots, n$, of formula (2.2). We are going to employ a construction of this kind.

Specifically, for thin plate spline interpolation in two dimensions to the data $f_i = f(\underline{x}_i)$, $i = 1, 2, \ldots, n$, the analogue of expression (2.1) is the Lagrange function

$$\chi_k(\underline{x}) = \sum_{j=k}^{n} \lambda_{kj} \, \phi(\|\underline{x} - \underline{x}_j\|) + p_k(\underline{x}), \qquad \underline{x} \in \mathcal{R}^2, \tag{2.4}$$

whose parameters are fixed by the conditions

$$\left.\begin{array}{ll} \chi_k(\underline{x}_j) = \delta_{jk}, & j=k,k+1,\ldots,n, \\ \sum_{j=k}^{n} \lambda_{kj} = 0 \quad \text{and} & \sum_{j=k}^{n} \lambda_{kj}\, \underline{x}_j = 0 \end{array}\right\}, \tag{2.5}$$

where ϕ is the function (1.4) and where p_k is a polynomial from $\mathcal{R}^2$ to $\mathcal{R}$ of degree at most one. Thus χ_k is defined uniquely for all integers k in $[1,\hat{n}]$, where $\hat{n}$ is an integer from $[1,n-3]$ such that the last $q=n-\hat{n}$ of the points $\underline{x}_i$, $i=1,2,\ldots,n$, are not collinear. Then, corresponding to equation (2.2), we write the required interpolant to all the data in the form

$$s(\underline{x}) = \sum_{k=1}^{\hat{n}} \mu_k\, \chi_k(\underline{x}) + \sigma(\underline{x}), \qquad \underline{x}\in\mathcal{R}^2, \tag{2.6}$$

where σ is the thin plate spline interpolant to the last q data. Thus the analogue of expression (2.3), namely the partial sum

$$s_\ell(\underline{x}) = \sum_{k=\ell}^{\hat{n}} \mu_k\, \chi_k(\underline{x}) + \sigma(\underline{x}), \qquad \underline{x}\in\mathcal{R}^2, \tag{2.7}$$

is the unique thin plate spline interpolant to the last $n-\ell+1$ data, the centres of the radial functions being at the interpolation points as usual, where ℓ is any integer from $[1,\hat{n}]$. It follows that, for each of these integers ℓ, the parameter μ_ℓ is the coefficient of $\phi(\|\underline{x}-\underline{x}_\ell\|)$ in s_ℓ divided by $\lambda_{\ell\ell}$. Fortunately, this remark provides a convenient way of calculating μ_ℓ, which will be explained in the case $\ell=1$.

Let A be the leading $n\times n$ submatrix of the inverse of the partitioned matrix of the system (1.3). It follows from this system that, for general right hand sides f_i, $i=1,2,\ldots,n$, the thin plate spline coefficients of expression (1.1) are the components of the product

$$\underline{\lambda} = A\, \underline{f}. \tag{2.8}$$

In particular, by choosing $\underline{f}$ to be the first coordinate vector in $\mathcal{R}^n$, we find that the elements of the first column of A are the coefficients λ_{1j}, $j=1,2,\ldots,n$, of the Lagrange function χ_1. Now A is a symmetric matrix and equation (2.8) gives $\lambda_1=\sum_{j=1}^{n} A_{1j}\, f_j$. Further, the definition (2.7) with $\ell=1$ provides the formula

$$\mu_1 = \lambda_1 \,/\, \lambda_{11} = \sum_{j=1}^{n} \lambda_{1j}\, f_j \,/\, \lambda_{11}. \tag{2.9}$$

Thus it is straightforward to generate μ_1 if the coefficients of the Lagrange function χ_1 are available. Similarly, the conclusion of the previous paragraph gives the values

$$\mu_\ell = \sum_{j=\ell}^{n} \lambda_{\ell j}\, f_j \,/\, \lambda_{\ell\ell}, \qquad \ell=1,2,\ldots,\hat{n}. \tag{2.10}$$

Therefore equation (2.6) with the coefficients (2.10) can be regarded as Newton's method for radial basis function interpolation.

The new procedure is similar to the one that has just been described, except that it makes approximations to the Lagrange functions that have been mentioned, and it is applied recursively, in order to correct the errors that arise from the approximations. Indeed, the value of q, introduced between expressions (2.5) and (2.6), is about 30, and the estimate of the function (2.4) has the form

$$\hat{\chi}_k(\underline{x}) = \sum_{j \in \mathcal{I}(k)} \hat{\lambda}_{kj}\, \phi(\|\underline{x}-\underline{x}_j\|) + \hat{p}_k(\underline{x}), \qquad \underline{x} \in \mathcal{R}^2, \tag{2.11}$$

where $\mathcal{I}(k)$ is a subset of the integers $\{k, k+1, \ldots, n\}$ that has only q elements. Specifically, $\hat{\chi}_k$ is the Lagrange function of thin plate spline interpolation at the points $\underline{x}_j$, $j \in \mathcal{I}(k)$, that has the properties $\hat{\chi}_k(\underline{x}_j) = \delta_{jk}$, $j \in \mathcal{I}(k)$. Further, it is usual to define the elements of $\mathcal{I}(k)$ by the condition that $\underline{x}_j$, $j \in \mathcal{I}(k)$, are the q points in the set $\{\underline{x}_j : j = k, k+1, \ldots, n\}$ that are closest to $\underline{x}_k$, including $\underline{x}_k$ itself. Then we deduce from the previous two paragraphs that the function

$$\hat{s}(\underline{x}) = \sum_{k=1}^{\hat{n}} \hat{\mu}_k\, \hat{\chi}_k(\underline{x}) + \hat{\sigma}(\underline{x}), \qquad \underline{x} \in \mathcal{R}^2, \tag{2.12}$$

may be a useful approximation to s, if we pick the coefficients

$$\hat{\mu}_\ell = \sum_{j \in \mathcal{I}(\ell)} \hat{\lambda}_{\ell j}\, f_j \,/\, \hat{\lambda}_{\ell\ell}, \qquad \ell = 1, 2, \ldots, \hat{n}, \tag{2.13}$$

where $\hat{\sigma}$ is the thin plate spline solution of the interpolation equations

$$\hat{\sigma}(\underline{x}_i) = f_i - \sum_{k=1}^{\hat{n}} \hat{\mu}_k\, \hat{\chi}_k(\underline{x}_i), \qquad i = \hat{n}+1, \hat{n}+2, \ldots, n, \tag{2.14}$$

in order that $\hat{s}$ interpolates the last $q = n - \hat{n}$ data.

In general, however, $\hat{s}$ does not interpolate all the data, because of the errors of the estimates $\hat{\chi}_k \approx \chi_k$, $k = 1, 2, \ldots, \hat{n}$. It happens typically in numerical experiments that $\hat{s}$ has the property

$$\max\{|f_i - \hat{s}(\underline{x}_i)| : i = 1, 2, \ldots, n\} \approx 0.1 \max\{|f_i| : i = 1, 2, \ldots, n\}. \tag{2.15}$$

Therefore the following iterative refinement method is recommended. We now let s be an estimate of the required interpolant that is set to zero initially. Then the data f_i, $i = 1, 2, \ldots, n$, are replaced by $f_i - s(\underline{x}_i)$, $i = 1, 2, \ldots, n$, throughout the description of the previous paragraph. In particular, equation (2.13) becomes the formula

$$\hat{\mu}_\ell = \sum_{j \in \mathcal{I}(\ell)} \hat{\lambda}_{\ell j}\, [f_j - s(\underline{x}_j)] \,/\, \hat{\lambda}_{\ell\ell}, \qquad \ell = 1, 2, \ldots, \hat{n}. \tag{2.16}$$

The resultant thin plate spline function $\hat{s}$ is added to s, which completes the iteration and which is expected to reduce $\max\{|f_i - s(\underline{x}_i)| : i=1,2,\ldots,n\}$ by about a factor of 10. Termination occurs when the maximum residual becomes acceptably small. All the coefficients of the approximate Lagrange functions are generated before the iterations are begun. Therefore most of the work of an iteration is the calculation of the residuals, which requires only $\mathcal{O}(n \log n)$ operations, because the Laurent expansion technique of Beatson and Newsam (1992) is applied. Thus interpolation equations with more than 10^4 data points can be solved, as claimed in Section 1.

The difference between this iterative procedure and the version that was in use a year ago is that the estimate s of the required interpolant was revised $\hat{n}+1$ times, instead of only once, on each iteration. Specifically, for each integer ℓ in $[1,\hat{n}]$, the coefficient (2.16) was calculated, and then $\hat{\mu}_\ell \hat{\chi}_\ell$ was added to s before increasing ℓ for the next application of formula (2.16). After all of these changes to s, the thin plate spline $\hat{\sigma}$ was defined by the interpolation conditions

$$\hat{\sigma}(\underline{x}_i) = f_i - s(\underline{x}_i), \qquad i=\hat{n}+1, \hat{n}+2, \ldots, n, \tag{2.17}$$

which correspond to expression (2.14), and s was overwritten by $s+\hat{\sigma}$, which completed the iteration. Then the main task of each iteration was generating the residuals $f_j - s(\underline{x}_j)$, $j \in \mathcal{I}(\ell)$, for $\ell = 1,2,\ldots,\hat{n}$. Goodsell noticed, however, that the amount of work per iteration would be reduced by about a factor of $q = |\mathcal{I}(\ell)|$ if the residuals at the beginning of the iteration were employed instead. Therefore he investigated this idea by trying some numerical experiments. He found in most cases that the number of iterations to achieve any prescribed accuracy was about the same as before. Then consideration of a theoretical explanation provided the relation between the procedure and Newton's interpolation method that has been described in this section.

Recent work has included attempts to construct positions of the data points $\underline{x}_i$, $i=1,2,\ldots,n$, so that one or both of the versions of the iterative procedure cause the residuals $f_i - s(\underline{x}_i)$, $i=1,2,\ldots,n$, to diverge. The year old version seems to be very robust, but some examples of failure have been found for the newer version that usually requires much less computation. One of the bad cases is when $n=1000$ and the data points are equally spaced on two concentric quarter circles of radii 1 and $1+10^{-5}$. Therefore research has started on trying to develop an intermediate version that combines the reliability of the year old procedure with some of the savings that can be achieved by calculating fewer residuals. In particular, we have tried with some success to calculate the first half of the coefficients (2.16) using the residuals at the beginning of the iteration. Then the first half of the sum of expression (2.12) is added to s, which gives new residuals that are employed in the calculation of the third quarter of the coefficients (2.16). Further, we alternate between updating s and calculating about half of the remainder of the coefficients. Thus the ordering of the data points $\underline{x}_i$, $i=1,2,\ldots,n$, becomes important, so several questions require attention. Another current topic of investigation is the replacement of the thin plate spline radial function by the multiquadric $\phi(r) = (r^2+c^2)^{1/2}$, $r \geq 0$, in the iterative procedure. Several numerical experiments have been tried, most of them being very successful.

3. A semi-norm for global optimization

The author has become very interested in an application of radial basis function interpolation to global optimization due to an inspiring paper of Jones (1996). The global optimization problem is to find the least value of a function $f(\underline{x})$, $\underline{x} \in \mathcal{X}$, where $\mathcal{X}$ is a subset of $\mathcal{R}^d$, when the value $f(\underline{x})$ can be calculated for any $\underline{x}$ in $\mathcal{X}$, but no derivatives are available. Let the values that have been generated already at the beginning of an iteration of a recursive procedure be $f(\underline{x}_i)$, $i=1,2,\ldots,n$. Then the main task of an iteration is to use these data to select $\underline{x}_{n+1}$, which is the next point at which the value of the objective function will be calculated. Jones described the following technique for making the selection automatically. It may demand much routine computation, which is tolerable if function values are very expensive.

A way of defining a unique function s that satisfies the interpolation equations

$$s(\underline{x}_i) = f(\underline{x}_i) = f_i, \qquad i=1,2,\ldots,n, \tag{3.1}$$

is required, for general n, $\underline{x}_i \in \mathcal{X}$ and f_i, $i=1,2,\ldots,n$. Further, s should be the "smoothest" function from a linear space $\mathcal{A}$ that is allowed by the interpolation equations, where the smoothness of $a \in \mathcal{A}$ is a nonnegative real number, $\sigma(a)$ say, that is defined for every a in $\mathcal{A}$. Therefore, if a is any function in $\mathcal{A}$ with $a(\underline{x}_i)=f_i$, $i=1,2,\ldots,n$, then the smoothness property of s is the inequality $\sigma(s) \leq \sigma(a)$. Now $\mathcal{A}$ has to be large enough to include a function $\hat{s}$ that satisfies the conditions

$$\hat{s}(\underline{x}_{n+1}) = f_* \quad \text{and} \quad \hat{s}(\underline{x}_i) = f_i, \quad i=1,2,\ldots,n, \tag{3.2}$$

for any choice of $\underline{x}_{n+1}$ in $\mathcal{X}$ that is different from $\underline{x}_i$, $i=1,2,\ldots,n$, and any $f_* \in \mathcal{R}$. Further, we pick the $\hat{s}$ that minimizes $\sigma(\hat{s})$ subject to the constraints (3.2), and we denote the dependence of $\sigma(\hat{s})$ on $\underline{x}_{n+1} \in \mathcal{X}$ and $f_* \in \mathcal{R}$ by the notation

$$\sigma(\hat{s}) = \hat{\sigma}(\underline{x}_{n+1}, f_*). \tag{3.3}$$

The technique sets $\underline{x}_{n+1}$ to the vector of variables in $\mathcal{X}$ that minimizes $\hat{\sigma}(\underline{x}, f_*)$, $\underline{x} \in \mathcal{X}$, for some chosen value of f_*.

The technique has some nice features due to the dependence of this value of $\underline{x}_{n+1}$ on f_*, assuming the bound

$$f_* < \min\{s(\underline{x}) : \underline{x} \in \mathcal{X}\}, \tag{3.4}$$

where s is defined at the beginning of the previous paragraph. Indeed, because this bound implies $f_* < f_i$, $i=1,2,\ldots,n$, the smoothness of $\hat{s}$ keeps $\underline{x}_{n+1}$ away from the points $\underline{x}_i$, $i=1,2,\ldots,n$. Furthermore, if f_* is much less than the right hand side of expression (3.4), the value $f_* = -\infty$ being used occasionally, then there is a tendency for the position of $\underline{x}_{n+1}$ to provide a relatively large value of the nearest neighbour distance $\min\{\|\underline{x}_{n+1} - \underline{x}_i\| : i=1,2,\ldots,n\}$. Thus $\underline{x}_{n+1}$ can be placed automatically in a large gap between the interpolation points $\underline{x}_i$, $i=1,2,\ldots,n$, which is important to global optimization. On the other hand, if the global minimum of $s(\underline{x})$, $\underline{x} \in \mathcal{X}$, is close to the required global minimum, which is usual for sufficiently large n, then it may be possible to achieve some good convergence properties by letting the difference between the two sides of expression (3.4) be tiny. The idea of alternating between such choices of f_* on successive iterations is a subject of current research at Cambridge. Hence we hope to develop some useful new algorithms for global optimization.

We will find in the remainder of this section that the radial basis function interpolation method provides the smoothness properties that are required in a form that is suitable for practical computation. In particular, the work of Duchon (1977) suggests a choice of $\mathcal{A}$ and the smoothness measure $\sigma(a)$, $a \in \mathcal{A}$, for thin plate spline interpolation when $d=2$. Specifically, letting $\mathcal{A}$ be the linear space of functions from $\mathcal{R}^2$ to $\mathcal{R}$ with square integrable second derivatives, letting $\sigma(a)$ be the integral

$$\sigma(a) = \frac{1}{8\pi} \int_{\mathcal{R}^2} \left(\frac{\partial^2 a(\underline{x})}{\partial x_1^2} \right)^2 + 2 \left(\frac{\partial^2 a(\underline{x})}{\partial x_1 \, \partial x_2} \right)^2 + \left(\frac{\partial^2 a(\underline{x})}{\partial x_2^2} \right)^2 d\underline{x}, \qquad a \in \mathcal{A}, \quad (3.5)$$

and letting s be the thin plate spline solution of the interpolation equations (3.1) that is studied in Section 2, it is proved by Duchon (1977) that s has the property $\sigma(s) \leq \sigma(a)$, for every $a \in \mathcal{A}$ that satisfies the interpolation conditions. Moreover, the dependence of $\sigma(\hat{s}) = \hat{\sigma}(\underline{x}_{n+1}, f_*)$ on $\underline{x}_{n+1}$ is required, where $\hat{s}$ is the thin plate spline solution of the equations (3.2), but it is not convenient to extract this dependence by estimating the integral (3.5). Integration by parts avoids this difficulty. Indeed, because the bi-Laplacian of $\phi(\|\underline{x}\|)$, $\underline{x} \in \mathcal{R}^2$, in the case (1.4) is a delta function, one can deduce the identity

$$\sigma(a) = \sum_{i=1}^{n(a)} \mu_i \, a(\underline{y}_i), \quad (3.6)$$

when a is any element of $\mathcal{A}$ that has the form

$$a(\underline{x}) = \sum_{j=1}^{n(a)} \mu_j \, \phi(\|\underline{x} - \underline{y}_j\|) + q(\underline{x}), \qquad \underline{x} \in \mathcal{R}^2. \quad (3.7)$$

Here the condition $a \in \mathcal{A}$ imposes the constraints $\sum_{j=1}^{n(a)} \mu_j = 0$ and $\sum_{j=1}^{n(a)} \mu_j \, \underline{y}_j = 0$, but there are no other restrictions on the parameters $\mu_j \in \mathcal{R}$ and $\underline{y}_j \in \mathcal{R}^2$,

$j=1,2,\ldots,n(a)$, and q is any polynomial of degree at most one. For example, equations (1.1), (3.1) and (3.6) give the formula

$$\sigma(s) = \sum_{i=1}^{n} \lambda_i \, s(\underline{x}_i) = \sum_{i=1}^{n} \lambda_i \, f_i. \tag{3.8}$$

A useful and obvious property of the smoothness measure (3.5) is that the inequality

$$\sigma(\hat{s}) = \hat{\sigma}(\underline{x}_{n+1}, f_*) \geq \sigma(s), \tag{3.9}$$

holds, because $\hat{s}$ is an element of $\mathcal{A}$ that satisfies the equations (3.2), while $s \in \mathcal{A}$ minimizes $\sigma(s)$ subject to the constraints (3.1). It is not straightforward, however, to extend the smoothness measure (3.5) to other choices of the radial function ϕ and the number of variables d. An extension suggests itself, however, if we let expression (3.6) be the definition of the smoothness instead, which is equivalent to the definition (3.5) in the $d=2$ thin plate spline case, provided that the linear space $\mathcal{A}$ is reduced. Specifically, the new $\mathcal{A}$ is composed of all functions of the form (3.7) whose parameters satisfy $\sum_{j=1}^{n(a)} \mu_j = 0$ and $\sum_{j=1}^{n(a)} \mu_j \, \underline{y}_j = 0$. Thus $\mathcal{A}$ includes all the functions that are required by the application to global optimization that has been described. We are going to gain some understanding of expression (3.6) as a measure of smoothness by using it to deduce the property (3.9).

It is sufficient to prove that, if the generic function (3.7) of the linear space $\mathcal{A}$ satisfies the interpolation conditions $a(\underline{x}_i)=f_i$, $i=1,2,\ldots,n$, then the inequality $\sigma(a) \geq \sigma(s)$ is achieved. There is no loss of generality in assuming that $\{\underline{x}_i : i = 1,2,\ldots,n\}$ is a subset of $\{\underline{y}_j : j=1,2,\ldots,n(a)\}$, because we can enlarge the latter set if necessary, letting the corresponding coefficients μ_j be zero in expression (3.7). Further, we let Ψ be the $n(a) \times n(a)$ matrix with the elements $\Psi_{ij} = \phi(\|\underline{y}_i - \underline{y}_j\|)$, $1 \leq i,j \leq n(a)$. It follows from equation (3.7) that a takes the values

$$a(\underline{y}_i) = \sum_{j=1}^{n(a)} \Psi_{ij} \, \mu_j + q(\underline{y}_i), \qquad i=1,2,\ldots,n(a), \tag{3.10}$$

so the smoothness measure (3.6) can be expressed in the form

$$\sigma(a) = \sum_{i=1}^{n(a)} \sum_{j=1}^{n(a)} \mu_i \, \Psi_{ij} \, \mu_j = \underline{\mu}^T \Psi \underline{\mu}. \tag{3.11}$$

We see that the nonnegativity of $\sigma(a)$ can be attributed to the conditional positive definiteness of Φ (or Ψ) that is stated in the second paragraph of Section 1.

We continue the proof of $\sigma(a) \geq \sigma(s)$ by adding some zero terms to formula (1.1) for the thin plate spline solution of the equations (3.1). Indeed, we assume without loss of generality that $\underline{x}_i = \underline{y}_i$ holds for $i=1,2,\ldots,n$, and we set λ_j to zero for every integer j in $[n+1, n(a)]$, which gives the identity

$$s(\underline{x}) = \sum_{j=1}^{n(a)} \lambda_j \, \phi(\|\underline{x} - \underline{y}_j\|) + p(\underline{x}), \qquad \underline{x} \in \mathcal{R}^2. \tag{3.12}$$

Further, equation (3.11) shows that $\sigma(s)$ is the nonnegative number $\underline{\lambda}^T \Psi \underline{\lambda}$, where $\underline{\lambda} \in \mathcal{R}^{n(a)}$ has the components λ_j, $j = 1, 2, \ldots, n(a)$. Thus the conditional positive definiteness of Ψ provides the inequality

$$\begin{aligned} \sigma(a) - \sigma(s) = \underline{\mu}^T \Psi \underline{\mu} - \underline{\lambda}^T \Psi \underline{\lambda} \;&=\; (\underline{\mu} - \underline{\lambda})^T \Psi (\underline{\mu} - \underline{\lambda}) + 2\, \underline{\lambda}^T \Psi (\underline{\mu} - \underline{\lambda}) \\ &\geq\; 2\, \underline{\lambda}^T \Psi (\underline{\mu} - \underline{\lambda}). \end{aligned} \tag{3.13}$$

Now equations (3.10) and (3.12) imply that the i-th component of the product $\Psi(\underline{\mu} - \underline{\lambda})$ has the value

$$[a(\underline{y}_i) - q(\underline{y}_i)] - [s(\underline{y}_i) - p(\underline{y}_i)], \qquad i = 1, 2, \ldots, n(a). \tag{3.14}$$

Moreover, the second part of the system (1.3) gives $\sum_{i=1}^{n} \lambda_i \, [p(\underline{y}_i) - q(\underline{y}_i)] = 0$. Therefore, because the last $n(a) - n$ components of $\underline{\lambda} \in \mathcal{R}^{n(a)}$ are zero, inequality (3.13) can be written in the form

$$\sigma(a) - \sigma(s) \geq 2 \sum_{i=1}^{n} \lambda_i \, [a(\underline{y}_i) - s(\underline{y}_i)] = 2 \sum_{i=1}^{n} \lambda_i \, [a(\underline{x}_i) - s(\underline{x}_i)]. \tag{3.15}$$

It follows from the interpolation conditions $s(\underline{x}_i) = f_i$ and $a(\underline{x}_i) = f_i$, $i = 1, 2, \ldots, n$, that the right hand side of expression (3.15) is zero, which completes the proof.

It is straightforward to extend the measure of smoothness (3.6) to the radial basis functions that are introduced at the beginning of Section 1. Recalling the purpose of m_0, we pick a constant integer m that satisfies $m \geq m_0$, and, as in Schaback (1993), we let a general element of $\mathcal{A}$ have the form

$$a(\underline{x}) = \sum_{j=1}^{n(a)} \mu_j \, \phi(\|\underline{x} - \underline{y}_j\|) + q(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d, \tag{3.16}$$

where the coefficients μ_j can have any values that satisfy $\sum_{j=1}^{n(a)} \mu_j \, \pi_k(\underline{y}_j) = 0$, $k = 1, 2, \ldots, \hat{m}$, and where q is any polynomial of degree at most m from $\mathcal{R}^d$ to $\mathcal{R}$. As before, $\{\pi_k : k = 1, 2, \ldots, \hat{m}\}$ is a basis of the space of these polynomials. These conditions imply the identity

$$\sum_{i=1}^{n(a)} \mu_i \, a(\underline{y}_i) = \underline{\mu}^T \Psi \underline{\mu}, \tag{3.17}$$

where Ψ is still the matrix with the elements $\Psi_{ij} = \phi(\|\underline{y}_i - \underline{y}_j\|)$, $1 \le i,j \le n(a)$. Now it is claimed in Section 1 that, if $\underline{\mu}$ is nonzero, then $(-1)^{m_0+1}\underline{\mu}^T\Psi\underline{\mu}$ is positive. Therefore we pick the measure of smoothness

$$\sigma(a) = (-1)^{m_0+1} \sum_{i=1}^{n(a)} \mu_i\, a(\underline{y}_i), \qquad a \in \mathcal{A}. \tag{3.18}$$

Further, we consider the arguments of the previous two paragraphs in this more general setting. It follows that, if s is the element of $\mathcal{A}$ of the form (1.1) whose coefficients are defined by the system (1.3), then $\sigma(s)$ is the least value of $\sigma(a)$, $a \in \mathcal{A}$, subject to the interpolation conditions $a(\underline{x}_i) = f_i$, $i = 1, 2, \ldots, n$. This result is Theorem 4 of Schaback (1993). It provides the highly interesting possibility that several versions of the radial basis function interpolation method may be suitable for our application to global optimization.

Furthermore, there is a scalar product (a,b), between any two elements a and b of $\mathcal{A}$, such that the measure of smoothness is just $\sigma(a) = (a,a)$, $a \in \mathcal{A}$, and such that $(a,a)^{1/2}$, $a \in \mathcal{A}$, is a semi-norm of the linear space $\mathcal{A}$. In particular, if $b \in \mathcal{A}$ is the function

$$b(\underline{x}) = \sum_{j=1}^{n(a)} \nu_j\, \phi(\|\underline{x} - \underline{y}_j\|) + r(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d, \tag{3.19}$$

where r is a polynomial as usual, then equation (3.18) and the homogeneity of the quadratic function (3.17) suggest the value

$$(a,b) = (-1)^{m_0+1}\underline{\mu}^T\Psi\underline{\nu} = (-1)^{m_0+1} \sum_{i=1}^{n(a)} \mu_i\, b(\underline{y}_i) = (-1)^{m_0+1} \sum_{i=1}^{n(a)} \nu_i\, a(\underline{y}_i), \tag{3.20}$$

where the last two identities follow easily from equations (3.19) and (3.16). In general, however, $b \in \mathcal{A}$ is the function

$$b(\underline{x}) = \sum_{j=1}^{n(b)} \nu_j\, \phi(\|\underline{x} - \underline{z}_j\|) + r(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d, \tag{3.21}$$

where the sets $\{\underline{z}_j : j = 1, 2, \ldots, n(b)\}$ and $\{\underline{y}_j : j = 1, 2, \ldots, n(a)\}$ are different. In this case we replace both sets by their union, multiplying the new radial terms of expressions (3.16) and (3.21) by zero coefficients, in order that equation (3.20) becomes valid after increasing $n(a)$. Further, all of the new zero coefficient terms can be removed from the sums of this formula. Thus we derive the bilinear form

$$(a,b) = (-1)^{m_0+1} \sum_{i=1}^{n(a)} \mu_i\, b(\underline{y}_i) = (-1)^{m_0+1} \sum_{i=1}^{n(b)} \nu_i\, a(\underline{z}_i), \tag{3.22}$$

even if the sets $\{\underline{y}_i : i=1,2,\ldots,n(a)\}$ and $\{\underline{z}_i : i=1,2,\ldots,n(b)\}$ have no points in common. It follows that (a,b) is a scalar product with the property that $(a,a)^{1/2}$, $a\in\mathcal{A}$, is a semi-norm of the linear space $\mathcal{A}$ (Schaback, 1993). Further, the sums of the definition (3.22) are convenient for the practical computation of the scalar product.

These remarks imply that expression (3.3) has some very useful properties. In order to describe them, we let $\underline{x}_{n+1}$ be any point of $\mathcal{X}$ that is different from $\underline{x}_i$, $i=1,2,\ldots,n$, and we let $\chi(\underline{x})$, $\underline{x}\in\mathcal{X}$, be the Lagrange function of radial basis function interpolation at the points $\underline{x}_i$, $i=1,2,\ldots,n+1$, that satisfies the conditions

$$\chi(\underline{x}_{n+1})=1 \quad \text{and} \quad \chi(\underline{x}_i)=0, \quad i=1,2,\ldots,n. \tag{3.23}$$

Then equations (1.1) and (3.22) give the identity

$$(s,\chi) = (-1)^{m_0+1}\sum_{i=1}^{n} \lambda_i\,\chi(\underline{x}_i) = 0, \tag{3.24}$$

which shows that χ is orthogonal to s. Now the function

$$\hat{s}(\underline{x}) = s(\underline{x}) + [f_* - s(\underline{x}_{n+1})]\,\chi(\underline{x}), \qquad \underline{x}\in\mathcal{X}, \tag{3.25}$$

is the radial basis function solution of the interpolation equations (3.2). It follows from the orthogonality property (3.24) that the measure of smoothness (3.3) has the value

$$\sigma(\hat{s}) = (\hat{s},\hat{s}) = (s,s) + [f_* - s(\underline{x}_{n+1})]^2\,(\chi,\chi). \tag{3.26}$$

Further, defining $\mu(\underline{x}_{n+1})$ to be the coefficient of $\phi(\|\underline{x}-\underline{x}_{n+1}\|)$, $\underline{x}\in\mathcal{X}$, in χ, equations (3.22) and (3.23) provide $(\chi,\chi)=(-1)^{m_0+1}\mu(\underline{x}_{n+1})$. Thus we deduce the formula

$$\hat{\sigma}(\underline{x}_{n+1},f_*) = \sigma(s) + (-1)^{m_0+1}\,[f_* - s(\underline{x}_{n+1})]^2\,\mu(\underline{x}_{n+1}), \tag{3.27}$$

which shows the dependence of $\hat{\sigma}$ on $\underline{x}_{n+1}$ and f_*. This result is very helpful to the current research on global optimization at Cambridge, not only for seeking a point $\underline{x}_{n+1}$ that minimizes $\hat{\sigma}(\underline{x}_{n+1},f_*)$ approximately, but also for providing information that assists the choice of f_*.

4. On thin plate spline interpolation in one dimension

Throughout this section we consider interpolation to a function $f(x)$, $0\le x\le 1$, of only one variable at the equally spaced points $x_i=ih$, $i=0,1,\ldots,n$, where n is a positive integer and $h=1/n$. The interpolant has the form

$$s(x) = \sum_{j=0}^{n} \lambda_j\,\phi(|x-x_j|) + p(x), \qquad 0\le x\le 1, \tag{4.1}$$

n	$f(x)=x^2$	$f(x)=x^3$
80	0.00013546	0.00009982
160	0.00004780	0.00003554
320	0.00001689	0.00001261
640	0.00000597	0.00000447

Table 1: The usual order of accuracy

where ϕ and p are still the thin plate spline radial function (1.4) and a polynomial of degree at most one, respectively. Corresponding to the system (1.3), the coefficients of the usual interpolant are defined by the equations

$$s(x_i) = f(x_i), \qquad i=0,1,\ldots,n, \tag{4.2}$$

and by the constraints

$$\sum_{j=0}^{n} \lambda_j = 0 \quad \text{and} \quad \sum_{j=0}^{n} \lambda_j x_j = 0, \tag{4.3}$$

but later we will study the replacement of the constraints (4.3) by two other conditions. We give particular attention to the error function

$$e(x) = f(x) - s(x), \qquad 0 \le x \le 1, \tag{4.4}$$

when f has bounded fourth derivatives. In this case the number $\max\{|e(x)| : \varepsilon \le x \le 1-\varepsilon\}$, is bounded above by a multiple of h^3, where ε is any positive constant and the multiplying factor depends on f and ε but not on h (Bejancu, 1997). Unfortunately, the factor grows so rapidly as $\varepsilon \to 0$ that the maximum error

$$\|e\|_\infty = \max\{|e(x)| : 0 \le x \le 1\} \tag{4.5}$$

is hardly ever of magnitude h^3.

Numerical results show that typically the greatest value of $|e(x)|$, $0 \le x \le 1$, occurs when x is in the interval $[0,h]$ or $[1-h,1]$. Further, if $|e(\frac{1}{2}h)|$ tends to zero more slowly than a multiple of h^3 as $h \to \infty$, which happens in all the numerical experiments of this section, then $|e(\frac{1}{2}h)|$ indicates the magnitude of $\max\{|e(x)| : 0 \le x \le \frac{1}{2}\}$. Therefore the main parts of the displays of the given tables are just values of $|e(\frac{1}{2}h)|$ for a range of choices of n and f. These results provide some interesting conjectures. Work has started at Cambridge on trying to establish them analytically.

n	$f(x)=x^3-\frac{3}{4}x^2$	$f(x)=(4.8)$
80	0.0000017728	0.00005749
160	0.0000003149	0.00002069
320	0.0000000558	0.00000738
640	0.0000000099	0.00000262

Table 2: Attempts at higher order convergence

The most usual behaviour is shown in Table 1 for two functions f, where s is defined by the equations (4.2) and (4.3). We see that, when h is halved, then $|e(\frac{1}{2}h)|$ is reduced by a factor of about $2\sqrt{2}$, which suggests the order of accuracy

$$\|e\|_\infty = \mathcal{O}(h^{3/2}). \tag{4.6}$$

This conjecture has been corroborated by many other calculations.

The entries in the third column of Table 1 are about 3/4 of the entries in the second column. Furthermore, if f is any function such that $|e(\frac{1}{2}h)| \sim h^{3/2}$, then it is usual for $e(x)$, $0 \le x \le h$, to be independent of f approximately, apart from a scaling factor. Therefore the $\mathcal{O}(h^{3/2})$ term of e in Table 1 may become zero if we pick $f(x) = x^3 - \frac{3}{4}x^2$, $0 \le x \le 1$. This possibility was investigated numerically, the results being shown in the second column of Table 2. We find that the condition $|e(\frac{1}{2}h)| = \mathcal{O}(h^{5/2})$ is achieved, which raises the question of characterising the smooth functions f that enjoy an order of convergence that is higher than usual. It is answered in Theorem 4.2 of Beatson and Powell (1992) for multiquadric interpolation, and there it is sufficient if f satisfies the equations

$$f(0) = f(1) = f'(0) = f'(1) = 0. \tag{4.7}$$

Therefore we tested the possibility of a similar situation in the thin plate spline case by making the choice

$$f(x) = \begin{cases} 0, & 0 \le x \le \frac{1}{4}, \\ 10^5\,(x-\frac{1}{4})^4\,(x-\frac{3}{4})^4, & \frac{1}{4} \le x \le \frac{3}{4}, \\ 0, & \frac{3}{4} \le x \le 1. \end{cases} \tag{4.8}$$

The resultant values of $|e(\frac{1}{2}h)|$ are shown in the last column of Table 2. We see that the $\mathcal{O}(h^{3/2})$ magnitude of Table 1 occurs again. Therefore we are still seeking a characterisation of the smooth functions f that provide $\|f-s\|_\infty = \mathcal{O}(h^{5/2})$, when s is the thin plate spline radial function that is given by the conditions (4.2) and (4.3).

n	$f(x)=x^2$	$f(x)=x^3$
80	0.00000668	0.00001120
160	0.00000121	0.00000202
320	0.00000022	0.00000036
640	0.00000004	0.00000006

Table 3: The first derivative conditions (4.9)

The form of $e(x)$, $0 \le x \le h$, in the case $|e(\frac{1}{2}h)| \sim h^{3/2}$, mentioned at the beginning of the previous paragraph, has to be due mainly to the thin plate spline interpolation method. Indeed, it seems that $e(x)$, $0 \le x \le h$, is dominated by a function of the form (4.1) that is zero at all the interpolation points. Now different multiples of such functions occur in s if we retain the interpolation conditions (4.2) but alter the constraints (4.3). Further, the magnitude of the error near the ends of the interval $[0,1]$ is likely to be smaller if the two constraints are replaced by the equations

$$s'(0) = f'(0) \quad \text{and} \quad s'(1) = f'(1). \tag{4.9}$$

Therefore this replacement was investigated numerically for the functions of Table 1, the new values of $|e(\frac{1}{2}h)|$ being shown in Table 3. It is clear that the accuracy is much better than before, and there is strong evidence that $\|e\|_\infty = \mathcal{O}(h^{5/2})$ is achieved instead of the order (4.6).

These results suggest that usually the constraints (4.3) cause unnecessary loss of accuracy, but the derivatives $f'(0)$ and $f'(1)$ may not be available. Therefore another technique from Beatson and Powell (1992) was also investigated. It takes up the two degrees of freedom that remain in the coefficients λ_j, $j=0,1,\ldots,n$, of the function (4.1), after the interpolation conditions (4.2) are satisfied, by minimizing the sum of squares $\sum_{j=0}^n \lambda_j^2$. The resultant values of $|e(\frac{1}{2}h)|$ were about four times the corresponding entries in Table 3, so again the order of accuracy seems to be $\mathcal{O}(h^{5/2})$, which was corroborated by other choices of f.

There is a fundamental objection to the methods of the last two paragraphs, however, due to the log term of $\phi(r)=r^2 \log r$, $r \ge 0$. We explain it by scaling the variable x by a positive constant α say, to give the new variable $\hat{x}=\alpha x$. Then the function $f(x)$, $0 \le x \le 1$, becomes $f(\alpha^{-1}\hat{x})$, $0 \le \hat{x} \le \alpha$, and we seek an interpolant of the form

$$\hat{s}(\hat{x}) = \sum_{j=0}^{n} \hat{\lambda}_j \, \phi(|\hat{x} - \alpha x_j|) + \hat{p}(\hat{x}), \qquad 0 \le \hat{x} \le \alpha, \tag{4.10}$$

that satisfies the conditions

$$\hat{s}(\alpha x_i) = f(x_i), \qquad i=0,1,\ldots,n, \tag{4.11}$$

where the notation $x_i = ih$, $i = 0, 1, \ldots, n$, has not changed. The fundamental objection is that, although $\hat{s}(\alpha x)$, $0 \le x \le 1$, is a thin plate spline function that interpolates the original data, one may not be able to express it in the form (4.1). Indeed, equation (4.10) implies the identity

$$\begin{aligned} \hat{s}(\alpha x) &= \sum_{j=0}^{n} \hat{\lambda}_j \, (\alpha x - \alpha x_j)^2 \log |\alpha x - \alpha x_j| + \hat{p}(\alpha x) \\ &= \sum_{j=0}^{n} \alpha^2 \hat{\lambda}_j \, \phi(|x - x_j|) + \alpha^2 \log \alpha \sum_{j=0}^{n} \hat{\lambda}_j \, (x - x_j)^2 + \hat{p}(\alpha x), \end{aligned} \qquad (4.12)$$

which differs from the form (1.1) by the quadratic polynomial $(\alpha^2 \log \alpha \sum_{j=0}^{n} \hat{\lambda}_j) \, x^2$, $0 \le x \le 1$. Therefore the abandonment of the first of the constraints (4.3) has caused the interpolation method to lose an important invariance property under simple scaling of the variable $x \in \mathcal{R}$.

Equation (4.12) shows that the objection does not occur if we allow p and $\hat{p}$ in expressions (4.1) and (4.10) to be quadratic instead of linear polynomials. Then three degrees of freedom remain in the coefficients of s after the interpolation equations (4.2) are satisfied, and they can also be fixed by minimizing $\sum_{j=0}^{n} \lambda_j^2$. This technique was tested for many smooth functions f, and again it was found that the order of accuracy is $\|e\|_\infty = \mathcal{O}(h^{5/2})$. Further, it is usual for the values of $|e(x)|$, $0 \le x \le 1$, to be smaller than those that occur in all the other versions of thin plate spline interpolation that have been addressed in this section. We do not recommend, however, that the new method should be used in practice instead of cubic spline interpolation.

On the other hand, thin plate splines provide some strong advantages over piecewise polynomials when $d = 2$. Therefore the main purpose of the work of this section is to assist understanding of the errors that occur due to edge effects. It is hoped that these investigations will lead to some new versions of thin plate spline interpolation in two dimensions that will give better accuracy than usual near the edges of the range of data points.

Acknowledgements: George Goodsell carried out most of the calculations that guided the development of the method of Section 2, and he pioneered the highly useful idea of not updating the residuals during an iteration. The topics that are mentioned in the last paragraph of Section 2 and the global optimization technique of Section 3 are under investigation by Anita Faul and Hans-Martin Gutmann, respectively. The numerical results of Section 4 were found by Aurelian Bejancu, and he proposed the inclusion of a quadratic polynomial term in thin plate spline interpolation. It is a great pleasure to express my thanks to all of these people for many very helpful discussions and for their contributions to the research that is described.

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 (1992), "Univariate interpolation on a regular finite grid by a multiquadric plus a linear polynomial", *IMA J. Numer. Anal.*, Vol. 12, pp. 107–133.

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.

A. Bejancu (1997), "Local accuracy for radial basis function interpolation on finite uniform grids", Report No. DAMTP 1997/NA19, University of Cambridge.

M.D. Buhmann (1990), "Multivariate cardinal interpolation with radial-basis functions", *Constr. Approx.*, Vol. 6, pp. 225–255.

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.

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

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

D. Jones (1996), "Global optimization with response surfaces", presented at the Fifth SIAM Conference on Optimization, Victoria, Canada.

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

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.