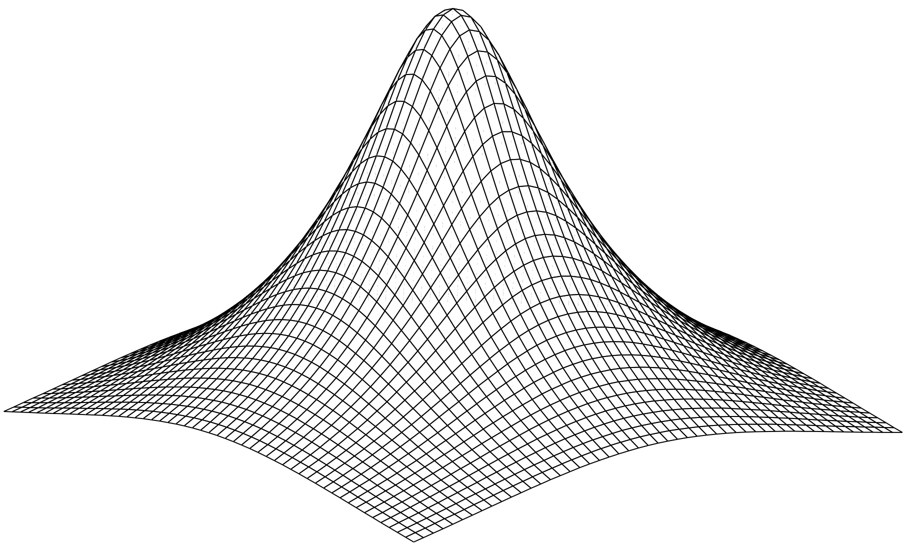
Five Lectures on Radial Basis Functions

M.J.D. Powell

February 10, 2005



Professor Mike J. D. Powell spent three weeks at IMM in November – December 2004. During the visit he gave five lectures on radial basis functions. These notes are a TeXified version of his hand-outs, made by Hans Bruun Nielsen, IMM.

Lecture 1

Interpolation in one dimension

Let values of $f : \mathcal{R} \mapsto \mathcal{R}$ be given.

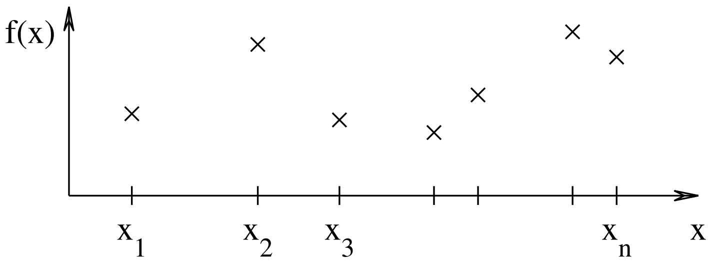


Figure 1.1. *Data points.*

The data are $f(x_i)$, $i = 1, 2, \ldots, n$, where $x_1 < x_2 < \cdots < x_n$.

Pick a function $s(x)$, $x_1 \leq x \leq x_n$, such that $s(x_i) = f(x_i)$, $i = 1, 2, \ldots, n$.

Piecewise linear interpolation. Let s be a linear polynomial on each of the intervals $[x_i, x_{i+1}]$, $i = 1, 2, \ldots, n-1$. Then s can be expressed in the form

$$s(x) = \sum_{j=1}^{n} \lambda_j \, |x - x_j|, \qquad x_1 \leq x \leq x_n \ .$$

To see that, define λ_j, $j = 2, 3, \ldots, n-1$, by the first derivative discontinuities of s, and then pick λ_1 and λ_n to satisfy $s(x_1) = f(x_1)$ and $s(x_n) = f(x_n)$.

Extension to two dimensions

Now $f : \mathcal{R}^2 \mapsto \mathcal{R}$ and the data are $f(\underline{x}_i)$, $i = 1, 2, \ldots, n$, where $\underline{x}_i = \begin{pmatrix} \xi_i \\ \eta_i \end{pmatrix} \in \mathcal{R}^2$.

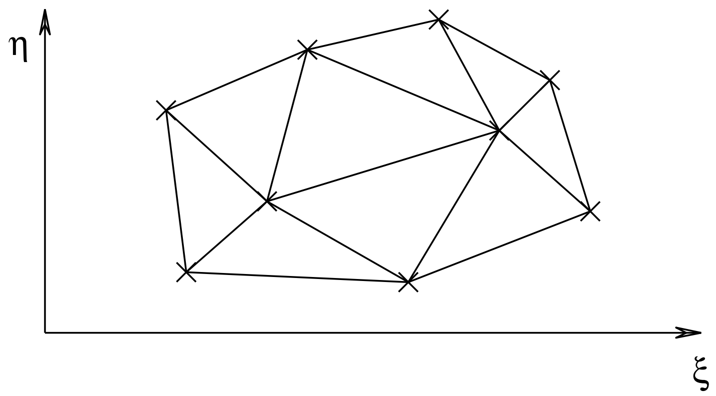


Figure 1.2. *Points in 2D and triangulation.*

Method 1. Form a triangulation and then employ linear polynomial interpolation on each triangle.

Method 2. Set $s(\underline{x}) = \sum_{j=1}^{n} \lambda_j \, \|\underline{x} - \underline{x}_j\|$, $\underline{x} \in \mathcal{R}^2$, where the coefficients λ_j, $j = 1, 2, \ldots, n$, have to satisfy $s(\underline{x}_i) = f(\underline{x}_i)$, $i = 1, 2, \ldots, n$.

Clearly, s is different in the two cases; one way of showing this is to consider where the gradient $\underline{\nabla} s$ is discontinuous.

Higher dimensions

Let $f : \mathcal{R}^d \mapsto \mathcal{R}$ for some positive integer d. Method 2, but not Method 1 allows large values of d.

Radial basis function interpolation

Pick a function $\phi(r),\ r \geq 0$, for example $\phi(r) = r$. Then let s have the form

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

where $\|\cdot\|$ denotes the Euclidean norm (the vector 2-norm). The parameters λ_j should satisfy

$$\Phi \underline{\lambda} = \underline{f} \ , \tag{1.1}$$

where Φ is the $n \times n$ matrix with elements

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

and where $\underline{f} \in \mathcal{R}^n$ has the components $f(\underline{x}_i),\ i = 1, 2, \ldots, n$. Does (1.1) have a unique solution, *ie* is Φ nonsingular ?

In these lectures we consider the following cases of radial basis functions, where α and c are given positive constants.

Gaussian	$\phi(r) = e^{-\alpha r^2},\ r \geq 0,\ \alpha > 0$
inverse multiquadric	$\phi(r) = (r^2 + c^2)^{-1/2},\ r \geq 0,\ c > 0$
linear	$\phi(r) = r,\ r \geq 0$
multiquadric	$\phi(r) = (r^2 + c^2)^{1/2},\ r \geq 0,\ c > 0$
thin plate spline	$\phi(r) = r^2 \log r,\ r \geq 0$
cubic	$\phi(r) = r^3,\ r \geq 0$

Table 1.1. *Cases of radial basis functions.*

The inverse multiquadric case is illustrated on the front page. We do not consider $\phi(r) = r^2,\ r \geq 0$, because then $s(\underline{x}) = \sum_{j=1}^{n} \lambda_j \, \phi(\|\underline{x} - \underline{x}_j\|),\ \underline{x} \in \mathcal{R}^d$, would always be a polynomial of degree at most two, so there would not be enough freedom to interpolate the values $f(\underline{x}_i),\ i = 1, 2, \ldots, n$, for sufficiently large n.

In the **Gaussian case** we have the following theorem.

Theorem 1.1. *If d is any positive integer and if the points $\underline{x}_i \in \mathcal{R}^d,\ i = 1, 2, \ldots, n$, are all different, then the matrix Φ with elements $\Phi_{ij} = e^{-\alpha \|\underline{x}_i - \underline{x}_j\|^2}$ is positive definite.*

Remarks. Each diagonal element of Φ is one. If the points are fixed and α is increased, then Φ becomes diagonally dominant.
The method of proof employs contour integration.

Corollary 1.2. *If ϕ can be written in the form*

$$\phi(r) = \int_0^\infty w(\alpha) e^{-\alpha r^2} d\alpha \ ,$$

where $w(\alpha) \geq 0$ for $\alpha \geq 0$ and $\int_\varepsilon^\infty w(\alpha) d\alpha > 0$ for some $\varepsilon > 0$, then Φ is positive definite.

Proof. If $\underline{v} \in \mathcal{R}^n$, we can write

$$\underline{v}^T \Phi \, \underline{v} \ = \ \sum_{i=1}^n \sum_{j=1}^n v_i \, v_j \, \phi(\|\underline{x}_i - \underline{x}_j\|) \ = \ \int_0^\infty w(\alpha) \sum_{i=1}^n \sum_{j=1}^n v_i \, v_j \, e^{-\alpha \|\underline{x}_i - \underline{x}_j\|^2} d\alpha \ ,$$

which is positive for $\underline{v} \neq \underline{0}$. $\square$

Example 1.1. The matrix Φ corresponding to inverse multiquadric rbf interpolation is positive definite because of the identity

$$\int_0^\infty \alpha^{-1/2} e^{-\alpha(r^2+c^2)} d\alpha = \sqrt{\pi} \left(r^2 + c^2\right)^{-1/2}, \quad r \geq 0 \ .$$

Completely monotonic functions

If $\phi(r) = \int_0^\infty w(\alpha) e^{-\alpha r^2} d\alpha, \ r \geq 0$, where $w \geq 0$, we consider the derivatives of the function

$$\psi(r) = \phi(r^{1/2}) = \int_0^\infty w(\alpha) e^{-\alpha r} d\alpha \ .$$

We find $\psi(r) \geq 0$ and $(-1)^k \psi^{(k)}(r) \geq 0$, $r \geq 0$, for all positive integers k. Such a function is called a *"completely monotonic function"*.

Theorem 1.3. *$\phi(r)$ can be expressed as $\int_0^\infty w(\alpha) e^{-\alpha r^2} d\alpha$, where $w \geq 0$, if and only if $\psi(r) = \phi(r^{1/2})$, $r \geq 0$, is completely monotonic.*

Example 1.2. $\phi(r) = e^{-\beta r^2} \ \Rightarrow \ \psi(r) = e^{-\beta r}$, which is completely monotonic for fixed $\beta > 0$.
$\phi(r) = \left(r^2 + c^2\right)^{-1/2} \ \Rightarrow \ \psi(r) = \left(r + c^2\right)^{-1/2}$, which is also completely monotonic.

Linear rbf. Now $\psi(r) = r^{1/2}$, $r \geq 0$, and we find $(-1)^k \psi^{(k)}(r) < 0$, $r > 0$, for all positive integers k. Hence the function $M - \psi(r)$, $r \geq 0$, "tends" to be completely monotonic for large enough M. Thus the matrix with elements $M - \|\underline{x}_i - \underline{x}_j\|$, $1 \leq i,j \leq n$, is positive definite for large enough M.

Lemma 1.4. *If $\Phi_{ij} = \|\underline{x}_i - \underline{x}_j\|$, $1 \leq i,j \leq n$, and if $\underline{v}$ is a nonzero vector in*

$$\mathcal{U} = \left\{ \underline{v} \in \mathcal{R}^n \,|\, \ \textstyle\sum_{i=1}^n v_i = 0 \right\} \ ,$$

then $\underline{v}^T \Phi \, \underline{v} < 0$.

Proof. For sufficiently large M we have

$$0<\sum_{i=1}^{n}\sum_{j=1}^{n} v_i\,(M-\|\underline{x}_i-\underline{x}_j\|)\,v_j = -\underline{v}^T\Phi\,\underline{v}\ . \qquad \Box$$

Theorem 1.5. *Let d be any positive integer and $n\geq 2$. If the points $\underline{x}_i\in\mathcal{R}^d$, $i{=}1,2,\ldots,n$, are all different, then the $n{\times}n$ matrix Φ with the elements $\Phi_{ij}=\|\underline{x}_i-\underline{x}_j\|$ is nonsingular. It has one positive and $n{-}1$ negative eigenvalues.*

Proof. According to the lemma $\underline{v}^T\Phi\,\underline{v}<0$ for any nonzero $\underline{v}\in\mathcal{U}\subset\mathcal{R}^n$. The dimension of $\mathcal{U}$ is $n{-}1$, so there are $n{-}1$ negative eigenvalues. Moreover, $\Phi_{ii}=0$, $i=1,2,\ldots,n$, implies that the sum of the eigenvalues is zero. Therefore one of the eigenvalues is positive. $\Box$

Multiquadric case

Now $\psi(r)=(r+c^2)^{1/2}$, $r\geq 0$, and $(-1)^k\psi^{(k)}(r)<0$, $r\geq 0$, for all positive integers k. Hence Φ has $n{-}1$ negative and one positive eigenvalues as before.

Multiquadrics with a constant term

Let s have the form

$$s(\underline{x})=\sum_{j=1}^{n}\lambda_j\left(\|\underline{x}-\underline{x}_j\|^2+c^2\right)^{1/2}+\mu, \qquad \underline{x}\in\mathcal{R}^d\ ,$$

with the constraint $\sum_{j=1}^{n}\lambda_j=0$, which we write as $\underline{1}^T\underline{\lambda}=0$, where $\underline{1}$ is the n-vector of all ones. Now the parameters have to satisfy the equations

$$\begin{pmatrix}\Phi & \underline{1}\\ \underline{1}^T & 0\end{pmatrix}\begin{pmatrix}\underline{\lambda}\\ \mu\end{pmatrix} = \begin{pmatrix}\underline{f}\\ 0\end{pmatrix}\ . \tag{1.2}$$

The matrix is nonsingular. Indeed if $\begin{pmatrix}\underline{\lambda}\\ \mu\end{pmatrix}$ is in its nullspace, then premultiplication of $\Phi\underline{\lambda}+\mu\underline{1}=\underline{0}$ by $\underline{\lambda}^T$ gives $\underline{\lambda}^T\Phi\underline{\lambda}+\mu\underline{\lambda}^T\underline{1}=\underline{\lambda}^T\Phi\underline{\lambda}=0$. Since Φ is nonsingular, $\underline{\lambda}=\underline{0}$ holds, which implies $\mu=0$.

This form of multiquadric interpolation will be important in lectures two and three.

Lecture 2

The integer m

Given $\phi(r),\ r \geq 0$, we define $\psi(r) = \phi(\sqrt{r}),\ r \geq 0$. We write $\psi^{(0)} \equiv \psi$, and for each positive integer k we write $\psi^{(k)}(r),\ r \geq 0$ for the kth derivative of ψ. The *"integer m of ϕ"* is the least nonnegative integer such that the sign of $(-1)^k \psi^{(k)}(r),\ r > 0$, $k = m, m{+}1, m{+}2, \ldots$, is independent of r and k. Let σ denote this sign.

Example 2.1. In the cases given in Table 1.1 the integer m takes the values 0, 0, 1, 1, 2, 2, respectively, and the values of σ are $1,\ 1,\ -1,\ -1,\ 1,\ 1$.

The following theorem gives a useful property of m. It holds for points $\underline{x} \in \mathcal{R}^d$ for any positive integer value of d. Π_{m-1} is the space of polynomials of degree at most $m{-}1$ from $\mathcal{R}^d$ to $\mathcal{R}$.

Theorem 2.1. *Let the points $\underline{x}_i \in \mathcal{R}^d$, $i = 1, 2, \ldots, n$, be different, where d is any positive integer, let Φ have the elements $\phi(\|\underline{x}_i - \underline{x}_j\|)$, $1 \leq i, j \leq n$, for some choice of ϕ such that the integer m exists, and let $\underline{v} \in \mathcal{R}^n$ satisfy $\sum_{i=1}^n v_i\, p(\underline{x}_i) = 0$, $p \in \Pi_{m-1}$, except that $\underline{v}$ is unconstrained in the case $m = 0$. Then $\underline{v}^T \Phi\, \underline{v}$ is nonzero if $\underline{v}$ is nonzero, and its sign is equal to σ in the definition of the integer m.*

Proof. Some advanced analysis is required, using the theory of completely monotonic functions, see pp 4–5. □

Rbf interpolation including low order polynomials

Given ϕ, we seek m, and, if $m = 0$, we apply radial basis function interpolation as before. Otherwise, we let s have the form

$$s(\underline{x}) = \sum_{j=1}^{n} \lambda_j\, \phi(\|\underline{x} - \underline{x}_j\|) + p(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d\ , \tag{2.1}$$

with the constraints $\sum_{j=1}^n \lambda_j\, q(\underline{x}_j) = 0$, $q \in \Pi_{m-1}$, and p chosen from Π_{m-1}. Now the interpolation conditions $s(\underline{x}_i) = f(\underline{x}_i)$, $i = 1, 2, \ldots, n$, are satisfied by a unique s for general right hand sides, if the points $\underline{x}_i$, $i = 1, 2, \ldots, n$, are different, and if $q \in \Pi_{m-1}$ with $q(\underline{x}_i) = 0$, $i = 1, 2, \ldots, n$, imply that q is zero.

Proof of the last assertion. We express the polynomial term in the form

$$p(\underline{x}) = \sum_{j=1}^{\widehat{m}} \mu_j\, b_j(\underline{x})\ , \qquad \underline{x} \in \mathcal{R}^d\ ,$$

where the $\{b_j : j = 1, 2, \ldots, \widehat{m}\}$ is a basis of Π_{m-1}, and we let B be the $\widehat{m} \times n$ matrix with elements $B_{ij} = b_i(\underline{x}_j)$, $i = 1, 2, \ldots, \widehat{m}$, $j = 1, 2, \ldots, n$. Then the interpolation conditions and the constraints give the linear system of equations

$$\begin{pmatrix} \Phi & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \underline{\lambda} \\ \underline{\mu} \end{pmatrix} = \begin{pmatrix} \underline{f} \\ \underline{0} \end{pmatrix}\ . \tag{2.2}$$

It is sufficient to prove that the matrix of the system is nonsingular, so we suppose that $\begin{pmatrix} \underline{v} \\ \underline{w} \end{pmatrix}$ is in its nullspace, which is equivalent to

$$\Phi\, \underline{v} + B^T \underline{w} = \underline{0} \quad \text{and} \quad B\underline{v} = \underline{0}\ .$$

Thus we find $\underline{v}^T \Phi\, \underline{v} = 0$, so Theorem 2.1 gives $\underline{v} = \underline{0}$. It now follows that $B^T \underline{w} = \underline{0}$, so the polynomial $\sum_{j=1}^{\widehat{m}} w_j b_j(\underline{x})$, $\underline{x} \in \mathcal{R}^d$, vanishes at $\underline{x}_j$, $j = 1, 2, \ldots, n$. Hence the condition at the end of the assertion gives $\underline{w} = \underline{0}$, which completes the proof. □

Example 2.2. In the multiquadric case $m = 1$, which implies $\widehat{m} = 1$, we choose $b_1(\underline{x}) = 1$ as the basis function of Π_0. This leads to $B = \underline{1}^T$, and we see that in this case (2.2) is identical to (1.2).

The native scalar product and semi-norm

The constant sign of $\underline{v}^T \Phi\, \underline{v}$ that has been mentioned provides the following "native" scalar product and semi-norm, the sign being $(-1)^m$ in all the cases that we consider. We let $\mathcal{S}$ be the linear space of all functions of the form

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

where k is any finite positive integer, where the points $\underline{y}_j$ are all different and can be anywhere in $\mathcal{R}^d$, where q is any element in Π_{m-1}, and where the real numbers μ_j can take any values that satisfy $\sum_{j=1}^{k} \mu_j\, p(\underline{y}_j) = 0$, $p \in \Pi_{m-1}$. The function

$$s(\underline{x}) = \sum_{i=1}^{n} \lambda_i\, \phi(\|\underline{x} - \underline{x}_i\|) + p(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d\ ,$$

is also in $\mathcal{S}$, and we define the scalar product

$$\langle s, t\rangle_\phi = (-1)^m \sum_{i=1}^{n} \sum_{j=1}^{k} \lambda_i\, \phi(\|\underline{x}_i - \underline{y}_j\|)\, \mu_j, \qquad s, t \in \mathcal{S}\ .$$

Further, because $\langle s, s\rangle_\phi$ is nonnegative by Theorem 2.1, we define the semi-norm

$$\|s\|_\phi = \sqrt{\langle s, s\rangle_\phi}, \qquad s \in \mathcal{S}\ .$$

Thus $\|s\|_\phi$ is zero if and only if all the coefficients λ_i, $i = 1, 2, \ldots, n$, of $s \in \mathcal{S}$ are zero.

An identity for the scalar product

$$\langle s, t\rangle_\phi = (-1)^m \sum_{i=1}^{n} \lambda_i\, t(\underline{x}_i) = (-1)^m \sum_{j=1}^{k} \mu_j\, s(\underline{y}_j)\ ,$$

the first part being proved as follows:

$$\begin{aligned} (-1)^m \sum_{i=1}^{n} \lambda_i\, t(\underline{x}_i) &= (-1)^m \sum_{i=1}^{n} \lambda_i \left\{ \sum_{j=1}^{k} \mu_j\, \phi(\|\underline{x}_i - \underline{y}_j\|) + q(\underline{x}_i) \right\} \\ &= (-1)^m \sum_{i=1}^{n} \sum_{j=1}^{k} \lambda_i\, \phi(\|\underline{x}_i - \underline{y}_j\|)\, \mu_j \;=\; \langle s, t\rangle_\phi\ . \end{aligned}$$

Rbf interpolation is "best"

Let $s \in \mathcal{S}$ be the given rbf interpolant to the data $f(\underline{x}_i)$, $i = 1, 2, \ldots, n$, of the form

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

There are infinitely many interpolants to the data from $\mathcal{S}$, but the following argument shows that s is the function in $\mathcal{S}$ that minimizes $\|s\|_\phi$ subject to the interpolation conditions.

Any other interpolant can be written as $s + t$, where t is in $\mathcal{S}$ and satisfies $t(\underline{x}_i) = 0$, $i = 1, 2, \ldots, n$. Thus the scalar product identity gives the inequality

$$\begin{aligned} \|s + t\|_\phi^2 &= \|s\|_\phi^2 + \|t\|_\phi^2 + 2\langle s, t\rangle_\phi \\ &= \|s\|_\phi^2 + \|t\|_\phi^2 + 2 \textstyle\sum_{i=1}^{n} \lambda_i \, t(\underline{x}_i) \\ &= \|s\|_\phi^2 + \|t\|_\phi^2 \;\geq\; \|s\|_\phi^2 \, . \end{aligned}$$

The inequality is strict unless $\|t\|_\phi$ is zero, and then t is in Π_{m-1}, which implies $t = 0$, assuming the condition stated soon after equation (2.1).

Thin plate spline interpolation in two dimensions

We consider this "best" property in the thin plate spline case $\phi(r) = r^2 \log r$, $r \geq 0$, when $d = 2$, the value of m being 2. Let ξ and η denote the components of $\underline{x} \in \mathcal{R}^2$, and, for $s \in \mathcal{S}$, let $I(s)$ be the integral

$$I(s) \;=\; \int\!\!\int_{\mathcal{R}^2} \left[\left(\frac{\partial^2 s}{\partial \xi^2} \right)^2 + 2 \left(\frac{\partial^2 s}{\partial \xi \partial \eta} \right)^2 + \left(\frac{\partial^2 s}{\partial \eta^2} \right)^2 \right] d\xi \, d\eta \, ,$$

which is finite, because the coefficients of s satisfy $\sum_{j=1}^{n} \lambda_j \, q(\underline{x}_j) = 0$, $q \in \Pi_1$. By applying integration by parts to $I(s)$, one can deduce

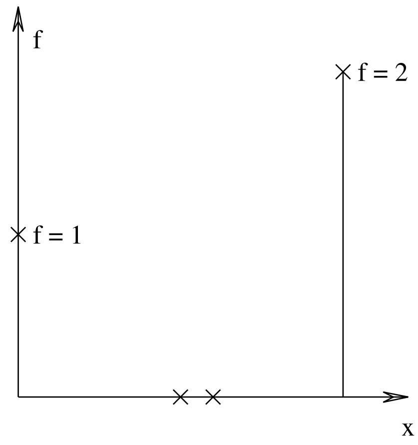
$$I(s) = 8\pi \sum_{i=1}^{n} \lambda_i \, s(\underline{x}_i) = 8\pi \, \|s\|_\phi^2 \, .$$

Thus the rbf interpolant s is the solution to the following problem: Find the function s with square integrable second derivatives that minimizes $I(s)$ subject to $s(\underline{x}_i) = f(\underline{x}_i)$, $i = 1, 2, \ldots, n$.

The rbf interpolation method usually gives good results in practice, and the "best" property may be the main reason for this success. We now turn to a technique for global optimization that uses the "best" property to implement an attractive idea.

The idea for global optimization in one dimension

This idea is due to Don Jones as far as I know. Let a function $f(x)$, $0 \le x \le 1$, take the four values that are shown, the crosses being the points $(0,1)$, $(0.5,0)$, $(0.6,0)$ and $(1,2)$. We seek the minimum of f using only calculated values of f. Therefore, when only $f(x_i)$, $i=1,2,\ldots,n$, are available, we have to pick x_{n+1}, the case $n=4$ being depicted. This task becomes less daunting if we assume that $f^* = \min\{f(x),\ 0 \le x \le 1\}$ is known. For example, we would pick x_5 from $[0.5, 0.6]$ or $[0, 0.5]$ in the cases $f^* = -0.1$ or $f^* = -10$, respectively. Specifically, we are trying to choose x_5 so that the data $f(x_i)$, $i=1,2,3,4$, and the hypothetical value $f(x_5) = f^*$ can be interpolated by as smooth a curve as possible, for the current choice of f^*. The rbf interpolation method is suitable, because we take the view that s is smoothest if $\|s\|_\phi$ is least.



An outline of the procedure for global optimization

We seek the least value of $f(\underline{x})$, $\underline{x} \in \mathcal{D} \subset \mathcal{R}^d$. We choose one of the functions ϕ that have been mentioned, in order to apply rbf interpolation, including the polynomial term $p \in \Pi_{m-1}$. We pick a few well-separated points $\underline{x}_i$, $i=1,2,\ldots,n$, for the first iteration such that $q \in \Pi_{m-1}$ and $q(\underline{x}_i)=0$, $i=1,2,\ldots,n$, imply $q=0$. Then the values $f(\underline{x}_i)$, $i=1,2,\ldots,n$, are calculated. Let $s \in \mathcal{S}$ be the usual interpolant to these values, which minimizes $\|s\|_\phi$. We are now ready to begin the first iteration.

On each iteration, f^* is set to a value that has the property $f^* < \min\{s(\underline{x}) : \underline{x} \in \mathcal{D}\}$. For each $\underline{y} \in \mathcal{D} \backslash \{\underline{x}_1, \underline{x}_2 \ldots, \underline{x}_n\}$, we define $s_{\underline{y}}$ to be the element of $\mathcal{S}$ that minimizes $\|s_{\underline{y}}\|_\phi$ subject to $s_{\underline{y}}(\underline{x}_i) = f(\underline{x}_i)$, $i=1,2,\ldots,n$, and $s_{\underline{y}}(\underline{y}) = f^*$. We seek $\underline{y}^*$, say, which is the $\underline{y}$ that minimizes $\|s_{\underline{y}}\|_\phi$. The value $\underline{x}_{n+1} = \underline{y}^*$ is chosen and $f(\underline{x}_{n+1})$ is calculated. Further, s is changed to the best interpolant to $f(\underline{x}_i)$, $i=1,2,\ldots,n+1$. Then n is increased by one for the next iteration.

On some iterations, f^* is close to $\min\{s(\underline{x}) : \underline{x} \in \mathcal{D}\}$, and on other iterations we pick $f^* = -\infty$. Thus the method combines local convergence properties with exploration of regions where f has not yet been calculated.

Further remarks. For each $\underline{y}$, the dependence of $s_{\underline{y}}$ on f^* is shown explicitly by the formula

$$s_{\underline{y}}(\underline{x}) = s(\underline{x}) + \{f^* - s(\underline{y})\}\, \ell_{\underline{y}}(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d\ ,$$

where $\ell_{\underline{y}}(\underline{x})$, $\underline{x} \in \mathcal{R}^d$, is the function of least norm in $\mathcal{S}$ that satisfies $\ell_{\underline{y}}(\underline{x}_i) = 0$, $i=1,2,\ldots,n$, and $\ell_{\underline{y}}(\underline{y}) = 1$. Further, the scalar product identity provides $\langle s, s - s_{\underline{y}} \rangle_\phi = 0$ and the formula

$$\|s_{\underline{y}}\|_\phi^2 = \|s\|_\phi^2 + \|s - s_{\underline{y}}\|_\phi^2 = \|s\|_\phi^2 + \{f^* - s(\underline{y})\}^2\, \|\ell_{\underline{y}}\|_\phi^2\ .$$

Therefore it may be helpful to calculate $s(\underline{y})$ and $\|\ell_{\underline{y}}\|_\phi$ for several values of $\underline{y}$ before choosing f^*. We see also that, if $f^* = -\infty$, then $\underline{y}^*$ is the value of $\underline{y}$ that minimizes $\|\ell_{\underline{y}}\|_\phi$.

Local adjustments to $\underline{y}$ can be based on the property that, if $\widehat{\underline{y}} \in \mathcal{D}$ satisfies $s_{\underline{y}}(\widehat{\underline{y}}) < s_{\underline{y}}(\underline{y})$, then the replacement of $\underline{y}$ by $\widehat{\underline{y}}$ gives a reduction in $\|s_{\underline{y}}\|_\phi$.

If $\mathcal{D}$ is bounded, and if $f^* = -\infty$ is chosen on every third iteration, say, then the points $\underline{x}_i$, $i = 1, 2, \ldots, n$, tend to become dense in $\mathcal{D}$ as $n \to \infty$. This observation is of hardly any value in practice, but it does provide a theoretical proof of convergence to the least value of $f(\underline{x})$, $\underline{x} \in \mathcal{D}$, when the objective function is continuous.

Further information about the algorithm can be found in *"A radial basis function method for global optimization"* by H-M. Gutmann, Journal of Global Optimization, **19**, pp 201–207 (2001).

Lecture 3

Rbf interpolation when n is large

Imagine that a radial basis function interpolant $s(\underline{x})$, $\underline{x}\in\mathcal{R}^2$, is required to 10,000 values of a function f on the unit square, the distribution of the data points being nearly uniform. Then, when $\underline{x}\in\mathcal{R}^2$ is a general point that is well inside the square, the value of $s(\underline{x})$ should depend mainly on values of f at interpolation points that are close to $\underline{x}$. Therefore it should be possible to construct a good estimate of $s(\underline{x})$ from a small subset of the data. Further, it may be possible to construct an adequate approximation to $s(\underline{x})$, $\underline{x}\in\mathcal{R}^2$, in only $O(n)$ operations, where n is still the number of given values of f. On the other hand, the work of deriving the coefficients of s directly from the interpolation equations would require $O(n^3)$ work, because there is no useful sparsity in the matrix Φ that has the elements $\Phi_{ij}=\phi(\|\underline{x}_i-\underline{x}_j\|)$, $1\le i,j\le n$.

Therefore some iterative procedures have been developed for the calculation of s. This lecture will describe a successful one that has been the subject of much research at Cambridge. It is a Krylov subspace method that employs the semi-norm that we studied in Lecture 2.

Notation. As in the previous lectures, we wish to interpolate the values $f(\underline{x}_i)$, $i=1,2,\ldots,n$, of a function $f:\mathcal{R}^d\mapsto\mathcal{R}$ by a function of the form

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

where $p\in\Pi_{m-1}$ and $m\in\{0,1,2\}$, the term p being dropped in the case $m=0$. Further, the coefficients have to satisfy the constraints

$$\sum_{j=1}^{n}\lambda_j\,q(\underline{x}_j)=0,\qquad q\in\Pi_{m-1}\ ,$$

when $m\ge 1$. We now let $\mathcal{S}$ be the linear space of functions s of this form, the centres $\underline{x}_j$, $j=1,2,\ldots,n$, being fixed, and we let

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

be another element of $\mathcal{S}$. We recall that the relation between m and ϕ provides a scalar product

$$\langle s,t\rangle_\phi\ =\ (-1)^m\sum_{j=1}^{n}\lambda_j\,t(\underline{x}_j)\ =\ (-1)^m\sum_{j=1}^{n}\mu_j\,s(\underline{x}_j)$$

and a semi-norm

$$\|s\|_\phi\ =\ \sqrt{\langle s,s\rangle_\phi}\ =\ \Big\{(-1)^m\textstyle\sum_{j=1}^{n}\lambda_j\,s(\underline{x}_j)\Big\}^{1/2},\qquad s\in\mathcal{S}\ .$$

Outline of the procedure for calculating the interpolant

We let $s_*\in\mathcal{S}$ be the required interpolant that satisfies $s_*(\underline{x}_i)=f(\underline{x}_i)$, $i=1,2,\ldots,n$. The procedure is iterative and constructs $s_{k+1}\in\mathcal{S}$ from $s_k\in\mathcal{S}$, starting with $s_1\equiv 0$. The norms

$\|s_k - s_*\|_\phi$, $k = 1, 2, 3, \ldots$, decrease strictly monotonically, and in theory $\|s_k - s_*\|_\phi = 0$ occurs for an integer k that is at most $n + 1 - \widehat{m}$, where $\widehat{m} = \dim(\Pi_{m-1})$. Far fewer iterations are usually required in practice.

If $\|s_{k+1} - s_*\|_\phi = 0$ happens, then $p \in \Pi_{m-1}$ can be calculated from the interpolation conditions, such that $s_* = s_{k+1} + p$. Therefore the iterations are terminated in this case. We do not consider the possibility $\|s_*\|_\phi = 0$ as no iterations are required.

The iterative procedure employs an operator $A : \mathcal{S} \mapsto \mathcal{S}$ that has the following properties

(1) $\|As\|_\phi = 0 \quad \Leftrightarrow \quad \|s\|_\phi = 0, \qquad s \in \mathcal{S}$ (nonsingularity).

(2) $\langle t, As\rangle_\phi = \langle At, s\rangle_\phi, \qquad s, t \in \mathcal{S}$ (symmetry).

(3) $\langle s, As\rangle_\phi > 0 \quad \Leftrightarrow \quad \|s\|_\phi > 0, \qquad s \in \mathcal{S}$ (strict positivity).

Then, for each iteration number $k > 0$, s_{k+1} is the element of the linear subspace spanned by $A^j s_*$, $j = 1, 2, \ldots, k$, that minimizes $\|s_{k+1} - s_*\|_\phi$. The following argument proves that the dimension of this subspace is k.

Otherwise, there are coefficients θ_j, $j = 1, 2, \ldots, k$, not all zero, such that $\|\sum_{j=1}^k \theta_j\, A^j s_*\|_\phi = 0$. We let ℓ be the least integer that satisfies $\theta_\ell \neq 0$. Then the nonsingularity of A implies the identity

$$\|\textstyle\sum_{j=1}^{k-\ell} (-\theta_{\ell+j}/\theta_\ell) A^j s_* - s_*\|_\phi = 0 \ .$$

Therefore termination would have occurred no later than the $(k-\ell)$-th iteration, which is a contradiction.

The construction of s_{k+1} from s_k

Our procedure is analogous to the conjugate gradient method for minimizing $\|s - s_*\|_\phi^2$, $s \in \mathcal{S}$, using A as a pre-conditioner, and starting at $s = s_1 = 0$. Letting A be the identity operator would provide perfect conditioning, but As_* is required and s_* is not available. Therefore A has a form, given later, that allows As_* to be calculated by the scalar product identity, using the values $s_*(\underline{x}_i) = f(\underline{x}_i)$, $i = 1, 2, \ldots, n$, which are data.

For each $k \geq 1$, s_{k+1} has the form

$$s_{k+1} = s_k + \alpha_k\, d_k \in \mathcal{S} \ ,$$

where $d_k \in \mathcal{S}$ is a "search direction", and where the "step length" α_k is set to the value of α that minimizes $\|s_k + \alpha\, d_k - s_*\|_\phi^2$, $\alpha \in \mathcal{R}$. The Krylov subspace construction is achieved by picking each d_k from $\mathrm{span}\{A^j s_* : j{=}1, 2, \ldots, k\}$ in a way that gives the orthogonality properties $\langle d_k, d_j\rangle_\phi = 0$, $j = 1, 2, \ldots, k{-}1$, for $k \geq 2$, and the descent condition $\langle d_k, s_k{-}s_*\rangle_\phi < 0$. Thus each α_k is positive until termination, and s_{k+1} is the function

$$s_{k+1} = s_k - \frac{\langle d_k, s_k - s_*\rangle_\phi}{\langle d_k, d_k\rangle_\phi}\, d_k \ \in \mathcal{S} \ ,$$

which satisfies $\langle d_k, s_{k+1} - s_*\rangle_\phi = 0$.

After calculating $t_k = A(s_k - s_*)$, the search direction d_k is defined by the formula

$$d_k = \begin{cases} -t_k, & k = 1 \ , \\ -t_k + \dfrac{\langle d_{k-1}, t_k\rangle_\phi}{\langle d_{k-1}, d_{k-1}\rangle_\phi}\, d_{k-1}, & k \geq 2 \ . \end{cases}$$

Thus the identity

$$\langle d_k, s_k - s_* \rangle_\phi \;=\; \langle -t_k, s_k - s_* \rangle_\phi \;=\; -\langle s_k - s_*, A(s_k - s_*) \rangle_\phi$$

holds for every k, the first equation being due to the line search of the previous iteration. It follows from the strict positivity of A that the descent condition is achieved. The orthogonality $\langle d_k, d_{k-1} \rangle_\phi = 0$ is given by the definition of d_k, and the properties $\langle d_k, d_j \rangle_\phi = 0$, $1 \le j \le k{-}2$, can be proved by induction.

On termination

The value of $\|s_{k+1} - s_*\|_\phi$ is not known for any k, because s_* is not available. On the other hand, the residuals

$$s_{k+1}(\underline{x}_i) - s_*(\underline{x}_i) = s_{k+1}(\underline{x}_i) - f(\underline{x}_i), \quad i = 1, 2, \ldots, n \ ,$$

can be calculated. Termination occurs when all the residuals can be less than a prescribed tolerance $\varepsilon > 0$, which is the condition

$$\min_{p \in \Pi_{m-1}} \max\{|s_{k+1}(\underline{x}_i) + p(\underline{x}_i) - f(\underline{x}_i)| : i = 1, 2, \ldots, n\} < \varepsilon \ .$$

The form of the operator A

If $\widehat{m} = \dim(\Pi_{m-1})$ exceeds one, the data points are reordered if necessary so that the zero function is the only element of Π_{m-1} that vanishes at the last $\widehat{m}$ points. Thus, if the coefficients λ_j, $j = 1, 2, \ldots, n{-}\widehat{m}$, of $s \in \mathcal{S}$ are given, the remaining coefficients λ_j, $n{-}\widehat{m}{+}1 \le j \le n$, are defined by the constraints $\sum_{j=1}^n \lambda_j \, q(\underline{x}_j) = 0$, $q \in \Pi_{m-1}$.

We pick functions $\ell_j \in \mathcal{S}$, $j = 1, 2, \ldots, n{-}\widehat{m}$, that have the form

$$\ell_j(\underline{x}) = \sum_{i=j}^{n} \Lambda_{ji} \, \phi(\|\underline{x} - \underline{x}_i\|), \qquad \underline{x} \in \mathcal{R}^d \ ,$$

with the important property that the leading coefficient Λ_{jj} is nonzero for each j. Thus the functions ℓ_j, $j = 1, 2, \ldots, n{-}\widehat{m}$, are linearly independent. Then the operator A is defined by the equation

$$A\, s \;=\; \sum_{j=1}^{n-\widehat{m}} \frac{\langle \ell_j, s \rangle_\phi}{\langle \ell_j, \ell_j \rangle_\phi} \, \ell_j \ , \qquad s \in \mathcal{S} \ .$$

This construction provides the nonsingularity, symmetry and strict positivity conditions that have been mentioned. Further, although s_* is not available, the function $t_k = A(s_k{-}s_*)$ can be calculated by using the formulae

$$\langle \ell_j, s_k - s_* \rangle_\phi \;=\; (-1)^m \sum_{i=j}^{n} \Lambda_{ji} \, \{s_k(\underline{x}_i) - f(\underline{x}_i)\}, \quad j = 1, 2, \ldots, n - \widehat{m} \ .$$

The choice of $\ell_j, \ j = 1, 2, \ldots, n{-}\widehat{m}$

The given procedure picks $s_1 = 0$, $d_1 = -t_1 = A s_*$ and $s_2 = \alpha_1 d_1$, where α_1 minimizes $\|s_2 - s_*\|_\phi = \|\alpha_1 A s_* - s_*\|_\phi$. It follows that the termination condition $\|s_{k+1} - s_*\|_\phi = 0$

is achieved on the first iteration for every s_*, if A has the property $\|As - s\|_\phi = 0$, $s \in \mathcal{S}$, which is equivalent to $\|A\ell_j - \ell_j\|_\phi = 0$, $j = 1, 2, \ldots, n{-}\widehat{m}$. These conditions are satisfied by the following choice of the functions ℓ_j, $j = 1, 2, \ldots, n{-}\widehat{m}$.

For each $j \in [1, n{-}\widehat{m}]$, we apply the radial basis function interpolation method to the data $f(\underline{x}_j) = 1$ and $f(\underline{x}_i) = 0$, $i = j{+}1, j{+}2, \ldots, n$, letting the interpolant be the function

$$\widetilde{g}_j(\underline{x}) = \sum_{i=j}^{n} \Gamma_{ji}\, \phi(\|\underline{x} - \underline{x}_i\|) + p_j(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d \ ,$$

where $p_j \in \Pi_{m-1}$. We define $g_j = \widetilde{g}_j - p_j$, and we consider the choice $\ell_j = g_j$, $j = 1, 2, \ldots, n{-}\widehat{m}$. The interpolation conditions imply $\|\widetilde{g}_j\|_\phi > 0$, and the scalar product identity gives the equation

$$\|\widetilde{g}_j\|_\phi^2 \ = \ (-1)^m \sum_{i=j}^{n} \Gamma_{ji}\, \widetilde{g}_j(\underline{x}_i) \ = \ (-1)^m \Gamma_{jj} \ ,$$

so the coefficients Γ_{jj}, $j = 1, 2, \ldots, n{-}\widehat{m}$, are nonzero as required. Further, if j and k are any integers that satisfy $1 \le j < k \le n{-}\widehat{m}$, we find the orthogonality property

$$\langle g_j, g_k \rangle_\phi \ = \ \langle g_k, \widetilde{g}_j \rangle_\phi \ = \ (-1)^m \sum_{i=k}^{n} \Gamma_{ki}\, \widetilde{g}_j(\underline{x}_i) \ = \ 0 \ .$$

Thus the definition of A on the previous page in the case $\ell_j = g_j$, $j = 1, 2, \ldots, n{-}\widehat{m}$, provides $A\ell_k = \ell_k$, $k = 1, 2, \ldots, n{-}\widehat{m}$. It follows that our iterative procedure takes at most one iteration for general right hand sides $f(\underline{x}_i)$, $i = 1, 2, \ldots, n$.

The bad news, however, is that the calculation of $\widetilde{g}_i$ is about as difficult as the calculation of s_*, because $\widetilde{g}_i$ also has to satisfy n interpolation conditions. In practice, therefore, guided by the remarks on page 12, we impose conditions on $\widetilde{g}_j$ of the form $\widetilde{g}_j(\underline{x}_i) = \delta_{ij}$, $i \in \mathcal{L}_j$, where $\mathcal{L}_j$ is a small subset of the integers $\{j, j{+}1, \ldots, n\}$ that includes j itself. The new $\widetilde{g}_j$ has the form

$$\widetilde{\ell}_j(\underline{x}) = \sum_{i \in \mathcal{L}_j} \Lambda_{ji}\, \phi(\|\underline{x} - \underline{x}_i\|) + \widetilde{p}_j(\underline{x}), \qquad \underline{x} \in \mathcal{R}^d \ ,$$

and we choose $\ell_j = \widetilde{\ell}_j - \widetilde{p}_j$, $j = 1, 2, \ldots, n{-}\widehat{m}$.

From now on we avoid $m \ge 2$ difficulties by restricting attention to linear and multiquadric radial functions, and we change the meaning of q to a prescribed positive integer that bounds the number of elements in each set $\mathcal{L}_j$, the value $q = 30$ being typical. Then $\{\underline{x}_i : i \in \mathcal{L}_j\}$ is chosen to contain the q points from the set $\{\underline{x}_i : i = j, j{+}1, \ldots, n\}$ that are closest to $\underline{x}_j$, except that $\mathcal{L}_j$ is the complete set $\{j, j{+}1, \ldots, n\}$ in the cases $n{-}q{+}1 \le j \le n{-}\widehat{m}$.

The Fortran software

A Fortran implementation of this procedure has been written at Cambridge for the radial basis functions $\phi(r) = \sqrt{r^2 + c^2}$, $r \ge 0$, where c can be zero. Of course the coefficients $\{\Gamma_{ji} : i \in \mathcal{L}_j\}$, $j = 1, 2, \ldots, n{-}1$, are found before beginning the iterations, which takes $O(nq^3) + O(n^2)$ operations, including the searches for nearest points.

The residuals $s_k(\underline{x}_i) - s_*(\underline{x}_i)$, $i = 1, 2, \ldots, n$, and the coefficients of s_k are required at the beginning of the k-th iteration, with the coefficients and values $d_{k-1}(\underline{x}_i)$, $i = 1, 2, \ldots, n$, of the previous search direction if $k \geq 2$. The calculation of the coefficients of $t_k = A(s_k - s_*)$ takes $O(nq)$ operations, and the extra work of finding the coefficients of d_k for $k \geq 2$ is negligible. The values $d_k(\underline{x}_i)$, $i = 1, 2, \ldots, n$, are also required, the work of this task being $O(n^2)$, which is the most expensive part of an iteration, unless fast multipole techniques are applied. This has been done very successfully by Beatson when d, the number of components in $\underline{x}$, is only 2 or 3. Finding the step length α_k is easy. Then the coefficients of s_{k+1} are given by $s_{k+1} = s_k + \alpha_k d_k$, and the new residuals take the values

$$s_{k+1}(\underline{x}_i) - s_*(\underline{x}_i) = s_k(\underline{x}_i) - s_*(\underline{x}_i) + \alpha_k \, d_k(\underline{x}_i), \quad i = 1, 2, \ldots, n \; .$$

The k-th iteration is completed by trying the test for termination. Further information can be found in *"A Krylov subspace algorithm for multiquadric interpolation in many dimensions"* by A.C. Faul, G. Goodsell and M.J.D. Powell, IMA Jnl. of Numer. Anal., **25**, pp 1–24 (2005).

A few numerical results

We sample each $\underline{x}_i$ and $f(\underline{x}_i)$ randomly from the uniform distribution on $\{\underline{x} : \|\underline{x}\| \leq 1\}$ and $[-1, 1]$, respectively, and we set $\varepsilon = 10^{-10}$. Thus 10 problems were generated for each d and n. The ranges of the number of iterations of the Fortran software with $c = 0$ are given in the table below.

	$d=2$		$d=5$	
n	$q=30$	$q=50$	$q=30$	$q=50$
250	$7-8$	$6-6$	$20-21$	$13-14$
500	$8-9$	$6-7$	$26-27$	$17-18$
1000	$9-10$	$7-8$	$34-36$	$22-23$
2000	$10-10$	$8-8$	$44-47$	$29-30$

This performance is very useful, but we will find in Lecture 5 that more iterations occur for $c > 0$.

Lecture 4

Purpose of the lecture

We apply the radial basis function method when the data are the values $f(\underline{j})$, $\underline{j} \in \mathcal{Z}^d$, of a smooth function $f : \mathcal{R}^d \mapsto \mathcal{R}$, where $\mathcal{Z}^d$ is the set of all points in $\mathcal{R}^d$ whose components are integers. We retain the notation $s(\underline{x})$, $\underline{x} \in \mathcal{R}^d$, for the interpolating function. We ask whether $s \equiv f$ occurs when f is a constant or a low order polynomial. If the answer is affirmative for all $f \in \Pi_\ell$, which is the space of polynomials of degree at most ℓ from $\mathcal{R}^d$ to $\mathcal{R}$, and if s_h is the interpolant to the data $f(h\underline{j})$, $\underline{j} \in \mathcal{Z}^d$, then, as $h \to 0$, we expect the errors $f(\underline{x}) - s_h(\underline{x})$, $\underline{x} \in \mathcal{R}^d$, to be of magnitude $h^{\ell+1}$.

Fourier transforms

Some brilliant answers to the questions above have been found by considering Fourier transforms. When $g(\underline{x})$, $\underline{x} \in \mathcal{R}^d$, is absolutely integrable, its Fourier transform is the function

$$\widehat{g}(\underline{t}) = \int_{\mathcal{R}^d} e^{-i\underline{x}\cdot\underline{t}}\, g(\underline{x})\, d\underline{x}, \qquad \underline{t} \in \mathcal{R}^d \ .$$

Further, if $\widehat{g}$ is absolutely integrable and if g is continuous, then the inverse function has the form

$$g(\underline{x}) = \frac{1}{(2\pi)^d} \int_{\mathcal{R}^d} e^{i\underline{x}\cdot\underline{t}}\, \widehat{g}(\underline{t})\, d\underline{t}, \qquad \underline{x} \in \mathcal{R}^d \ .$$

For example, the Fourier transform of the Gaussian $\phi(\|\underline{x}\|) = e^{-\alpha\|\underline{x}\|^2}$, $\underline{x} \in \mathcal{R}^d$, is the expression

$$\widehat{\phi}(\|\underline{t}\|) = (\pi/\alpha)^{d/2}\, e^{-\|\underline{t}\|^2/4\alpha}, \qquad \underline{t} \in \mathcal{R}^d \ .$$

It is elementary that, if g is absolutely integrable, then the function

$$u(\underline{x}) = \textstyle\sum_{j=1}^n \lambda_j\, g(\underline{x} - \underline{x}_j), \qquad \underline{x} \in \mathcal{R}^d \ ,$$

has the Fourier transform

$$\widehat{u}(\underline{t}) = \left\{ \textstyle\sum_{j=1}^n \lambda_j\, e^{i\underline{x}\cdot\underline{t}} \right\} \widehat{g}(\underline{t}), \qquad \underline{t} \in \mathcal{R}^d \ . \tag{4.1}$$

Thus, if $\widehat{u}$, λ_j and $\underline{x}_j$, $j = 1, 2, \ldots, n$, are available, we expect to be able to identify $\widehat{g}(\underline{t})$, $\underline{t} \in \mathcal{R}^d$.

Generalized Fourier transforms

The function $\phi(\|\underline{x}\|)$, $\underline{x} \in \mathcal{R}^d$, is not absolutely integrable for most of the choices of ϕ that have been mentioned, but they do have "generalized" Fourier transforms that can be identified by the relation (4.1) between $\widehat{g}$ and $\widehat{u}$. For example, in the case $\phi(r) = r$ and $d = 1$, the hat function

$$u(x) = \tfrac{1}{2}|x+1| - |x| + \tfrac{1}{2}|x-1|, \qquad x \in \mathcal{R} \ ,$$

satisfies $u(x) = 0$, $|x| \geq 1$, so it is absolutely integrable. Further, it has the Fourier transform

$$\begin{aligned} \widehat{u}(t) \ &= \ \int_0^1 \cos(xt)\,(2-2x)\,dx \ = \ \frac{2}{t^2}\,(1-\cos t) \\ &= \ \{\tfrac{1}{2}\,e^{-it} - 1 + \tfrac{1}{2}\,e^{it}\}(-2/t^2), \quad t \neq 0 \ . \end{aligned}$$

Therefore the relation (4.1) between $\widehat{g}$ and $\widehat{u}$ suggests that $g(x) = |x|$, $x \in \mathcal{R}$, has the transform $\widehat{g}(t) = -2/t^2$, $t \in \mathcal{R} \setminus \{0\}$.

Further, if $g(\underline{x})$, $\underline{x} \in \mathcal{R}^d$, is any function such that $u(\underline{x}) = \sum_{j=1}^n \lambda_j\, g(\underline{x} - \underline{x}_j)$, $\underline{x} \in \mathcal{R}^d$, is absolutely integrable, for some choices of λ_j and $\underline{x}_j$, $j = 1, 2, \ldots, n$, then g has a "generalized transform" $\widehat{g}(\underline{t})$, $\underline{t} \in \mathcal{R}^d$, independent of u, such that u has the Fourier transform $\widehat{u}(\underline{t}) = \left\{ \sum_{j=1}^n \lambda_j\, e^{i\underline{x}\cdot\underline{t}} \right\} \widehat{g}(\underline{t})$, $\underline{t} \in \mathcal{R}^d$.

All the radial basis functions given in Table 1.1 have Fourier transforms or generalized Fourier transforms. We let $\widehat{\phi}(\|\underline{t}\|)$, $\underline{t} \in \mathcal{R}^d$, be the transform of $\phi(\|\underline{x}\|)$, $\underline{x} \in \mathcal{R}^d$, although this notation does not show the dependence of $\widehat{\phi}$ on d. The following cases will be useful later.

ϕ	$\widehat{\phi}(\|\underline{t}\|), \quad \underline{t} \in \mathcal{R}^d$
$\phi(r) = r$	const $\times\ \|\underline{t}\|^{-d-1}$
$\phi(r) = r^2 \log r$	const $\times\ \|\underline{t}\|^{-d-2}$
$\phi(r) = r^3$	const $\times\ \|\underline{t}\|^{-d-3}$
$\phi(r) = e^{-\alpha r^2}$	$(\pi/\alpha)^{d/2}\, e^{-\|\underline{t}\|^2/(4\alpha)}$

Table 4.1. *Fourier transforms of radial basis functions.*

Interpolation on $\mathcal{Z}^d$

We choose $\phi(r)$, $r \geq 0$, and then we seek a function of the form

$$\chi(\underline{x}) = \sum_{\underline{k} \in \mathcal{Z}^d} \mu_{\underline{k}}\, \phi(\|\underline{x} - \underline{k}\|), \qquad \underline{x} \in \mathcal{R}^d ,$$

that satisfies the Lagrange conditions

$$\chi(\underline{j}) = \delta_{\underline{j}0}, \qquad \underline{j} \in \mathcal{Z}^d ,$$

where $\delta_{\underline{j}0}$ is the Kronecker delta. Thus, if all the sums are absolutely convergent, the values $f(\underline{j})$, $\underline{j} \in \mathcal{Z}^d$, are interpolated by

$$s(\underline{x}) = \sum_{\underline{j} \in \mathcal{Z}^d} f(\underline{j})\, \chi(\underline{x} - \underline{j}), \qquad \underline{x} \in \mathcal{R}^d . \tag{4.2}$$

We see that χ is independent of f, and that the convergence of the last sum may impose some restrictions on f. We see also that χ is without a $p \in \Pi_{m-1}$ polynomial term.

The importance of χ

When $\phi(r) = r$ and $d = 1$, χ is the hat function

$$\chi(x) = \tfrac{1}{2}|x+1| - |x| + \tfrac{1}{2}|x-1|, \qquad x \in \mathcal{R} .$$

It follows that the interpolant (4.2) is well defined for any $f(x)$, $x \in \mathcal{R}$, including the case $f(x) = 1$, $x \in \mathcal{R}$. On the other hand, it is not possible to express a nonzero constant function as the sum $\sum_{j=-\infty}^{\infty} \lambda_j |x - j|$, $x \in \mathcal{R}$, as all the coefficients λ_j have to be zero.

More importantly, if $\phi(r)$ is unbounded when $r \to \infty$, then hardly any interpolants on $\mathcal{Z}^d$ can be expressed as $\sum_{\underline{j}\in\mathcal{Z}^d} \lambda_{\underline{j}}\, \phi(\|\underline{x}-\underline{j}\|)$, $\underline{x}\in\mathcal{R}^d$. The value $\chi(x)$, however, usually tends to zero rapidly as $\|\underline{x}\| \to \infty$, which causes the sum in the interpolation formula (4.2) to be absolutely convergent for most functions f that are of interest.

The coefficients and Fourier transform of χ

The coefficients $\mu_{\underline{k}}$, $\underline{k}\in\mathcal{Z}^d$, of χ have an elegant form for all the given choices of ϕ. The function

$$\sigma(\underline{t}) = \left\{ \sum_{\underline{j}\in\mathcal{Z}^d} \widehat{\phi}(\|\underline{t}+2\pi\underline{j}\|) \right\}^{-1}, \qquad \underline{t}\in\mathcal{R}^d ,$$

is well defined, including $\sigma(0)=0$ when $\widehat{\phi}(0)$ is unbounded. We see that σ is periodic. Further, each $\mu_{\underline{k}}$ is defined by the Fourier series

$$\sigma(\underline{t}) = \sum_{\underline{k}\in\mathcal{Z}^d} \mu_{\underline{k}}\, e^{i\underline{k}\cdot\underline{t}}, \qquad \underline{t}\in[-\pi,\pi]^d ,$$

which gives the values

$$\mu_{\underline{j}} = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} e^{-i\underline{j}\cdot\underline{t}}\, \sigma(\underline{t})\, d\underline{t} , \qquad \underline{j}\in\mathcal{Z}^d .$$

This assertion implies that, if χ were absolutely integrable, then it would have the Fourier transform

$$\begin{aligned} \widehat{\chi}(\underline{t}) &= \left\{ \sum_{\underline{j}\in\mathcal{Z}^d} \mu_{\underline{j}}\, e^{i\underline{j}\cdot\underline{t}} \right\} \widehat{\phi}(\|\underline{t}\|) \\ &= \widehat{\phi}(\|\underline{t}\|) \Big/ \sum_{\underline{k}\in\mathcal{Z}^d} \widehat{\phi}(\|\underline{t}+2\pi\underline{k}\|) , \qquad \underline{t}\in\mathcal{R}^d , \end{aligned} \tag{4.3}$$

the first equation being analogous to (4.1). We justify the formula above for $\mu_{\underline{j}}, \underline{j}\in\mathcal{Z}^d$, by showing that the inverse Fourier transform of this $\widehat{\chi}$ does satisfy the Lagrange conditions.

Because $\widehat{\phi}$ does not change sign, we find $0 \le \widehat{\chi}(\underline{t}) \le 1$, $\underline{t}\in\mathcal{R}^d$, and $\widehat{\chi}$ decays fast enough to be absolutely integrable. Further, for each $\underline{j}\in\mathcal{Z}^d$, its inverse transform takes the value

$$\chi(\underline{j}) = \frac{1}{(2\pi)^d} \int_{\mathcal{R}^d} \frac{e^{i\underline{j}\cdot\underline{t}}\, \widehat{\phi}(\|\underline{t}\|)}{\sum_{\underline{k}\in\mathcal{Z}^d} \widehat{\phi}(\|\underline{t}+2\pi\underline{k}\|)}\, d\underline{t} .$$

We divide $\mathcal{R}^d$ into cubes of size $[-\pi,\pi]^d$ centred on $2\pi\underline{\ell}$, $\underline{\ell}\in\mathcal{Z}^d$, and we employ $e^{i\underline{j}\cdot(\underline{t}+2\pi\underline{\ell})} = e^{i\underline{j}\cdot\underline{t}}$. Thus the sum over $\underline{\ell}$ gives the term $\sum_{\underline{\ell}\in\mathcal{Z}^d} \widehat{\phi}(\|\underline{t}+2\pi\underline{\ell}\|)$ in the numerator of the integrand, which is the same as the denominator. These remarks provide

$$\chi(\underline{j}) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} e^{i\underline{j}\cdot\underline{t}}\, d\underline{t} = \delta_{\underline{j}0} , \qquad \underline{j}\in\mathcal{Z}^d ,$$

so the Lagrange conditions hold.

Reproduction of constants

We ask whether the interpolation formula (4.2) achieves $s \equiv f$ in the case $f(\underline{x}) = 1$, $\underline{x}\in\mathcal{R}^d$. Then s is the periodic function

$$s(\underline{x}) = \sum_{\underline{j}\in\mathcal{Z}^d} \chi(\underline{x}-\underline{j}) , \qquad \underline{x}\in\mathcal{R}^d ,$$

and we consider its Fourier series

$$s(\underline{x}) = \textstyle\sum_{\underline{k}\in\mathcal{Z}^d} \theta_{\underline{k}}\, e^{2\pi i\underline{k}\cdot\underline{x}} \;, \qquad \underline{x}\in[0,1]^d \;.$$

For each $\underline{k}$, the coefficient $\theta_{\underline{k}}$ takes the value

$$\theta_{\underline{k}} \;=\; \int_{[0,1]^d} e^{-2\pi i\underline{k}\cdot\underline{x}} \left\{\textstyle\sum_{\underline{j}\in\mathcal{Z}^d} \chi(\underline{x}-\underline{j})\right\} d\underline{x} \;=\; \int_{\mathcal{R}^d} e^{-2\pi i\underline{k}\cdot\underline{x}}\, \chi(\underline{x})\, d\underline{x} \;=\; \widehat{\chi}(2\pi\underline{k}) \;,$$

the two integrals being the same, because $\mathcal{R}^d$ can be split into blocks of size $[0,1]^d$ with their bottom left hand corner at $-\underline{j}$, $\underline{j}\in\mathcal{Z}^d$, while the last equation is just the definition of $\widehat{\chi}(2\pi\underline{k})$. It follows that $s\equiv f$ if and only if the equations $\widehat{\chi}(2\pi\underline{k})=\delta_{\underline{k}0}$, $\underline{k}\in\mathcal{Z}^d$, are satisfied. The formula

$$\widehat{\chi}(\underline{t}) = \widehat{\phi}(\|\underline{t}\|) \Big/ \textstyle\sum_{\underline{j}\in\mathcal{Z}^d} \widehat{\phi}(\|\underline{t}+2\pi\underline{j}\|) \;, \qquad \underline{t}\in\mathcal{R}^d \;,$$

provides these properties when $\widehat{\phi}(0)$ is unbounded and the sum of the terms $\widehat{\phi}(\|2\pi\underline{j}\|)$, $\underline{j}\in\mathcal{Z}^d\setminus\{0\}$, is finite. Therefore $s\equiv f$ holds for constant f for all the choices of ϕ in Table 4.1, except Gaussians. The properties of multiquadrics $\phi(r)=\sqrt{r^2+c^2}$ are similar to those of the linear case $\phi(r)=r$.

Reproduction of linear polynomials

Let interpolation on $\mathcal{Z}^d$ reproduce constants, and let f be a linear polynomial. We consider the case $f(\underline{x})=\xi_1$, $\underline{x}\in\mathcal{R}^d$, where ξ_1 is the first component of $\underline{x}$. Then the interpolant on $\mathcal{Z}^d$ is the function

$$\begin{aligned} s(\underline{x}) &= \textstyle\sum_{\underline{j}\in\mathcal{Z}^d} j_1\, \chi(\underline{x}-\underline{j}) \\ &= \textstyle\sum_{\underline{j}\in\mathcal{Z}^d} (j_1-\xi_1)\, \chi(\underline{x}-\underline{j}) + \xi_1 \sum_{\underline{j}\in\mathcal{Z}^d} \chi(\underline{x}-\underline{j}) \\ &= \xi_1 + \textstyle\sum_{\underline{j}\in\mathcal{Z}^d} (j_1-\xi_1)\, \chi(\underline{x}-\underline{j}), \qquad \underline{x}\in\mathcal{R}^d \;, \end{aligned}$$

the second line being elementary, and the third line being due to the reproduction of constants. Thus, the linear polynomial f is also reproduced if and only if the coefficients $\psi_{\underline{k}}$, $\underline{k}\in\mathcal{Z}^d$, of the periodic function

$$\sum_{\underline{j}\in\mathcal{Z}^d} (j_1-\xi_1)\, \chi(\underline{x}-\underline{j}) = \sum_{\underline{k}\in\mathcal{Z}^d} \psi_{\underline{k}}\, e^{2\pi i\underline{k}\cdot\underline{x}} \;, \qquad \underline{x}\in[0,1]^d \;,$$

are all zero. They have the values

$$\begin{aligned} \psi_{\underline{k}} &= \int_{[0,1]^d} e^{-2\pi i\underline{k}\cdot\underline{x}} \left\{\textstyle\sum_{\underline{j}\in\mathcal{Z}^d} (j_1-\xi_1)\, \chi(\underline{x}-\underline{j})\right\} d\underline{x} \\ &= \int_{\mathcal{R}^d} e^{-2\pi i\underline{k}\cdot\underline{x}}\, (-\xi_1)\, \chi(\underline{x})\, d\underline{x} \;, \qquad \underline{k}\in\mathcal{Z}^d \;. \end{aligned}$$

Fortunately, the definition of $\widehat{\chi}$ gives the equation

$$\frac{d}{dt_1}\,\widehat{\chi}(\underline{t}) \;=\; \int_{\mathcal{R}^d} (-i\xi_1)\, e^{-i\underline{x}\cdot\underline{t}}\, \chi(\underline{x})\, d\underline{x} \;,$$

which provides the coefficients

$$\psi_{\underline{k}} = -i \left[d\widehat{\chi}(\underline{t}) \,/ dt_1 \right]_{\underline{t}=2\pi\underline{k}}, \qquad \underline{k} \in \mathcal{Z}^d \,.$$

We treat other components of $\underline{x}$ in linear polynomials similarly. It follows that linear polynomials, as well as constants, are reproduced by the interpolation method if the gradients $\underline{\nabla}\widehat{\chi}(2\pi\underline{k})$, $\underline{k} \in \mathcal{Z}^d$, are all zero.

These conditions hold when $\widehat{\phi}(\|\underline{t}\|)$, $\underline{t} \in \mathcal{R}^d$, is a constant multiple of $\|\underline{t}\|^{-\mu}$, $\underline{t} \in \mathcal{R}^d$, with $\mu \geq 2$. Indeed, according to (4.3), we can write

$$\widehat{\chi}(\underline{t}) \;=\; \frac{\|\underline{t}\|^{-\mu}}{\sum_{\underline{j}\in\mathcal{Z}^d} \|\underline{t}+2\pi\underline{j}\|^{-\mu}} \;=\; \frac{1}{1+\|\underline{t}\|^{\mu} \sum_{\underline{j}\in\mathcal{Z}^d\setminus\{0\}} \|\underline{t}+2\pi\underline{j}\|^{-\mu}} \;,$$

which shows $\underline{\nabla}\widehat{\chi}(0)=0$. Alternatively, for any $\underline{k}$ in $\mathcal{Z}^d \setminus \{0\}$, we can express $\widehat{\chi}$ in the form

$$\widehat{\chi}(\underline{t}) \;=\; \frac{(\|\underline{t}-2\pi\underline{k}\| \,/\, \|\underline{t}\|\,)^{\mu}}{1+\|\underline{t}-2\pi\underline{k}\|^{\mu} \sum_{\underline{j}\in\mathcal{Z}^d\setminus\{-\underline{k}\}} \|\underline{t}+2\pi\underline{j}\|^{-\mu}} \;,$$

which shows $\underline{\nabla}\widehat{\chi}(2\pi\underline{k})=0$. Therefore linear polynomials are reproduced when $\phi(r)$ is r, $r^2 \log r$ or r^3.

Reproduction of higher order polynomials

The Strang and Fix theorem for finite elements identifies the integer ℓ such that interpolation on $\mathcal{Z}^d$ gives $s \equiv f$ for all $f \in \Pi_\ell$. In addition to $\widehat{\chi}(2\pi\underline{k}) = \delta_{\underline{k}0}$, $\underline{k} \in \mathcal{Z}^d$, all derivatives of $\widehat{\chi}(\underline{t})$, $\underline{t} \in \mathcal{R}^d$, of all orders up to and including ℓ have to vanish at $\underline{t} = 2\pi\underline{k}$, $\underline{k} \in \mathcal{Z}^d$. Therefore, in the case $\widehat{\phi}(\|\underline{t}\|) = \text{const} \times \|\underline{t}\|^{-\mu}$, $\underline{t} \in \mathcal{R}^d$, studied in the previous paragraph, we find $\ell = \mu{-}1$, assuming that μ is a positive integer. It follows from Table 4.1 that ℓ takes the value d, $d{+}1$ and $d{+}2$ for linear, thin plate spline and cubic functions, respectively. The value $\ell = d$ also holds for multiquadrics. These findings were unexpected!!

Order of accuracy

Let χ be the Lagrange function of interpolation on $\mathcal{Z}^d$. The interpolant to $f(h\underline{j})$, $\underline{j} \in \mathcal{Z}^d$, is

$$s_h(\underline{x}) = \textstyle\sum_{\underline{j}\in\mathcal{Z}^d} f(h\underline{j})\, \chi(h^{-1}\underline{x}-\underline{j}), \qquad \underline{x} \in \mathcal{R}^d \,.$$

We seek a bound on $|f(\underline{y})-s_h(\underline{y})|$, assuming $\underline{y} \in [-\frac{1}{2}h, \frac{1}{2}h]^d$. If $s_h \equiv f$ occurs for all $f \in \Pi_\ell$, and if f has bounded $(\ell{+}1)$-th derivatives, we let $p \in \Pi_\ell$ be the Taylor-series polynomial that satisfies $|p(\underline{x}) - f(\underline{x})| = O(\|\underline{x}-\underline{y}\|^{\ell+1})$, $\underline{x} \in \mathcal{R}^d$. Thus

$$\begin{aligned} |f(\underline{y}) - s_h(\underline{y})| \;&=\; \left| \textstyle\sum_{\underline{j}\in\mathcal{Z}^d} \{p(h\underline{j}) - f(h\underline{j})\}\, \chi(h^{-1}\underline{y}-\underline{j}) \right| \\ &\leq\; \textstyle\sum_{\underline{j}\in\mathcal{Z}^d} h^{\ell+1}\, O([\|\underline{j}\|+1]^{\ell+1})\, \max\{|\chi(\underline{j}+\underline{\theta})| : \underline{\theta} \in [-\frac{1}{2},\frac{1}{2}]^d\} \;, \end{aligned}$$

which implies $|f(\underline{y}) - s_h(\underline{y})| = O(h^{\ell+1})$ for fast enough decay of $|\chi(\underline{x})|$ as $\|\underline{x}\| \to \infty$. This decay is known and is sufficient in nearly all the cases that have been mentioned.

Lecture 5

Review of Lagrange functions for $\mathcal{Z}^d$

Let $\phi(r),\ r \ge 0$, be chosen for radial basis function interpolation on $\mathcal{Z}^d$. In the last lecture we considered Lagrange functions of the form

$$\chi(\underline{x}) \;=\; \sum_{\underline{k}\in\mathcal{Z}^d} \mu_{\underline{k}}\, \phi(\|\underline{x}-\underline{k}\|), \qquad \underline{x}\in\mathcal{R}^d \;,$$

that satisfy the conditions

$$\chi(\underline{j}) = \delta_{\underline{j}0}, \qquad \underline{j}\in\mathcal{Z}^d \;.$$

We found that χ can be written as the inverse Fourier transform

$$\chi(\underline{x}) \;=\; \frac{1}{(2\pi)^d}\int_{\mathcal{R}^d} \frac{e^{i\underline{x}\cdot\underline{t}}\,\widehat{\phi}(\|\underline{t}\|)}{\sum_{\underline{k}\in\mathcal{Z}^d}\widehat{\phi}(\|\underline{t}+2\pi\underline{k}\|)}\, d\underline{t} \;, \qquad \underline{x}\in\mathcal{R}^d \;,$$

where $\widehat{\phi}(\|\underline{t}\|),\ \underline{t}\in\mathcal{R}^d$, is the generalized Fourier transform of $\phi(\|\underline{x}\|),\ \underline{x}\in\mathcal{R}^d$.

We noted some cases where $\widehat{\phi}$ is the function $\widehat{\phi}(r) = \text{const}\times r^{-\mu},\ r\ge 0$, the value of μ being $d{+}1$, $d{+}2$ or $d{+}3$ for the linear, thin plate spline and cubic radial functions, respectively. Then the interpolation formula

$$s(\underline{x}) = \textstyle\sum_{\underline{j}\in\mathcal{Z}^d} f(\underline{j})\,\chi(\underline{x}-\underline{j}), \qquad \underline{x}\in\mathcal{R}^d \;,$$

has the amazing property $s\equiv f$ when $f\in\Pi_{\mu-1}$. This property holds also for multiquadric interpolation on $\mathcal{Z}^d$, and in this case $s\equiv f$ is achieved for $f\in\Pi_d$.

The decay of $\chi(\underline{x})$ as $\|\underline{x}\|\to\infty$

Let x_j be the j-th component of $\underline{x}\in\mathcal{R}^d$. If the ℓ-th derivatives of $\widehat{\chi}(\underline{t}),\ \underline{t}\in\mathcal{R}^d$, are absolutely integrable, then integration by parts provides

$$\chi(\underline{x}) \;=\; \frac{1}{(2\pi)^d}\int_{\mathcal{R}^d} \left\{\left(\frac{-1}{ix_j}\right)^{\ell} e^{i\underline{x}\cdot\underline{t}}\right\}\left\{\frac{d^\ell}{dt_j^\ell}\,\widehat{\chi}(\underline{t})\right\} d\underline{t} \;, \qquad \underline{x}\in\mathcal{R}^d \;, \tag{5.1}$$

for each $j\in[1,d]$. Thus $\|\underline{x}\|^\ell|\chi(\underline{x})|$ remains bounded as $\|\underline{x}\|\to\infty$.

In particular, the transform function

$$\widehat{\chi}(\underline{t}) \;=\; \|\underline{t}\|^{-\mu}\Big/\textstyle\sum_{\underline{k}\in\mathcal{Z}^d}\|\underline{t}+2\pi\underline{k}\|^{-\mu} \;, \qquad \underline{t}\in\mathcal{R}^d \;,$$

is infinitely differentiable when μ is even, and its derivatives of order up to $\mu{+}d$ are absolutely integrable when μ is odd. Therefore, in the linear case $\phi(r)=r$, for example, we find

$$|\chi(\underline{x})| = \begin{cases} O(\|\underline{x}\|^{-\ell}) & \text{for any } \ell>0 \text{ if } d \text{ is odd,} \\ O(\|\underline{x}\|^{-2d-1}) & \text{if } d \text{ is even .}\end{cases}$$

Further, the very fast decay of χ (in fact it is exponential) occurs for the thin plate spline or cubic choice of ϕ when d is even or odd, respectively.

The case $\phi(r) = r^2 \log r$ and $d = 1$

In this case the restrictions on the differentiability of

$$\widehat{\chi}(t) \; = \; |t|^{-3} \big/ \textstyle\sum_{k=-\infty}^{\infty} |t+2\pi k|^{-3} \; , \qquad t \in \mathcal{R} \; ,$$

come from the terms

$$\widehat{\chi}(t) \;\approx\; 1 - |t|^3 \textstyle\sum_{k\in\mathcal{Z}\setminus\{0\}} |t+2\pi k|^{-3} \quad \text{near } t=0 \; ,$$

and $\widehat{\chi}(t) \approx |t-2\pi k|^3 \big/ |t|^3$ near $t = 2\pi k$. Thus the part of $\dfrac{d^4}{dt^4}\,\widehat{\chi}(t)$, $t\in\mathcal{R}$, that is not differentiable is $12 \sum_{k\in\mathcal{Z}\setminus\{0\}} |2\pi k|^{-3} \{\delta(t-2\pi k) - \delta(t)\}$, $t\in\mathcal{R}$, where $\delta(t)$, $t\in\mathcal{R}$, is the delta function. Hence, for large $|x|$, (5.1) gives the approximation

$$\chi(x) \;\approx\; \frac{12}{2\pi}\,|x|^{-4} \sum_{k=1}^{\infty} |2\pi k|^{-3} \left\{ e^{2\pi i x k} - 2 + e^{-2\pi i x k} \right\} \;=\; -\frac{3}{\pi^4}\,|x|^{-4} \sum_{k=1}^{\infty} \sin^2(\pi k x)/k^3 \; .$$

We see that the approximation is $|x|^{-4}$ times a nonpositive periodic function of x that is zero at the interpolation points.

On multiquadrics

The function $\phi(\|\underline{x}\|) = \sqrt{\|\underline{x}\|^2 + c^2}$, $\underline{x}\in\mathcal{R}^d$, has the generalized Fourier transform

$$\widehat{\phi}(r) = \text{const} \times (c/r)^{(d+1)/2} K_{(d+1)/2}(cr), \quad r \ge 0 \; ,$$

where $r = \|\underline{t}\|$, $\underline{t}\in\mathcal{R}^d$, and where $K_{(d+1)/2}$ is a modified Bessel function. Thus (5.1) gives the decay rates

$$|\chi(\underline{x})| = \begin{cases} O(\|\underline{x}\|^{-3d-2}) & \text{if } d \text{ is odd} \; , \\ O(\|\underline{x}\|^{-2d-1}) & \text{if } d \text{ is even} \; . \end{cases}$$

When $d = 1$, this rate can be deduced from the approximation

$$\widehat{\phi}(r) \approx -2\left(1/r^2 + \tfrac{1}{2}\,c^2 \log r\right), \quad 0 < r \ll 1 \; , \tag{5.2}$$

which shows the terms that are most important for small r. This approximation also gives the error of multiquadric interpolation to $f(x) = x^2$, $x\in\mathcal{R}$, on $\mathcal{Z}$, in the way that is described briefly below. The property $s \equiv f$, $f\in\Pi_1$, implies that the error is the periodic function

$$f(x) - s(x) \;=\; x^2 - \sum_{j=-\infty}^{\infty} j^2\,\chi(x-j) \;=\; -\sum_{j=-\infty}^{\infty} (x-j)^2\,\chi(x-j) \; , \qquad x\in\mathcal{R} \; .$$

Further, the coefficients of its Fourier series

$$f(x) - s(x) = \tfrac{1}{2}\,\beta_0 + \sum_{j=1}^{\infty} \beta_j \cos(2j\pi x) \; , \qquad x\in\mathcal{R} \; ,$$

take the values

$$\beta_j = -2 \int_{-\infty}^{\infty} x^2\, e^{-2j\pi i x}\, \chi(x) dx = 2\,\widehat{\chi}''(2\pi j), \quad j = 0, 1, \ldots \; .$$

By putting the approximation (5.2) into the formula $\widehat{\chi}(t) = \widehat{\phi}(|t|) / \sum_{j\in\mathcal{Z}} \widehat{\phi}(|t+2\pi j|)$, without changing $\widehat{\phi}(|t+2\pi j|)$, $j \neq 0$, we deduce $\widehat{\chi}''(0) = 2\sum_{j=1}^{\infty} \widehat{\phi}(2\pi j)$, and similarly we find $\widehat{\chi}''(2\pi j) = -\widehat{\phi}(2\pi j)$, $j = 1, 2, \ldots$. Thus the collection of terms gives the identity

$$f(x) - s(x) \;=\; -2\sum_{j=1}^{\infty} \widehat{\phi}(2\pi j)\,\{\cos(2j\pi x) - 1\} \;=\; 4\sum_{j=1}^{\infty} \widehat{\phi}(2\pi j)\sin^2(j\pi x)\;, \quad x \in \mathcal{R}\;.$$

Now the asymptotic form of the Bessel function provides $\widehat{\phi}(r) \approx -2\sqrt{c\pi/(2r^3)}\,e^{-cr}$ for moderate and large r. Therefore, retaining only the first term of the last sum we construct

$$f(x) - s(x) \approx -\frac{2\sqrt{c}}{\pi}\,e^{-2\pi c}\sin^2(\pi x), \quad x \in \mathcal{R}\;,$$

when s is calculated by multiquadric interpolation to $f(x) = x^2$, $x \in \mathcal{R}$, at the points of $\mathcal{Z}$.

The value of $f(\frac{1}{2}) - s(\frac{1}{2})$ was computed to high accuracy for comparison with its estimate. The results for three values of c are as follows

c	$f(\frac{1}{2}) - s(\frac{1}{2})$	$-\frac{2\sqrt{c}}{\pi}\,e^{-2\pi c}$
1	-1.26×10^{-3}	-1.19×10^{-3}
3	-7.32×10^{-9}	-7.18×10^{-9}
5	-3.27×10^{-14}	-3.23×10^{-14}

Table 5.1. *The dependence of accuracy on c.*

We see that huge gains in accuracy occur when c is increased.

Interpolation on manifolds

This subject is studied in the report DAMTP 2001/NA11, that is available under the heading Numerical Analysis at the address www.damtp.cam.ac.uk. Some results will be quoted from it on increases in the multiquadric parameter c. We employ a two dimensional manifold in $\mathcal{R}^d$, where $d \geq 5$, which is without edges. It is constructed in the following way.

Let α_j and β_j, $j = 1, 2, \ldots, d$, be random constants, taken independently from the uniform distribution on $[0, 2\pi]$. For each choice of ξ and η from $[0, 2\pi]$, we let $\underline{x}(\xi, \eta) \in \mathcal{R}^d$ have the components

$$x_j = \widetilde{\sin}(\alpha_j + \xi) + \widetilde{\sin}(\beta_j + \eta), \quad j = 1, 2, \ldots, d\;,$$

where $\widetilde{\sin}(t) = \left(e^{\sin t} - 1\right)/(e-1)$, $t \in \mathcal{R}$. The manifold is the set

$$\{\underline{x}(\xi, \eta)\;:\; 0 \leq \xi, \eta \leq 2\pi\} \subset \mathcal{R}^d\;.$$

We try $d = 5$ and $d = 40$ in the examples that will be given. The interpolation points are $\underline{x}_i = \underline{x}(\xi_i, \eta_i) \in \mathcal{R}^d$, $i = 1, 2, \ldots, n$, where n is a square, and where the (ξ_i, η_i) form a square grid on $[2\pi/\sqrt{n}, 2\pi]^2$. We let f be the Gaussian

$$f(\underline{x}) = \exp(-\theta\,\|\underline{x} - \underline{x}(\pi, \pi)\|^2), \quad \underline{x} \in \mathcal{R}^d\;,$$

$\underline{x}(\pi, \pi)$ being on the manifold as defined already, and the constant θ being defined by the equation

$$\min\{f(\underline{x}_i) : i = 1, 2, \ldots, n\} = 0.01\;.$$

For each interpolant s, we estimate the maximum error

$$\|f-s\|_\infty = \max\{|f(\underline{x}(\xi,\eta)) - s(\underline{x}(\xi,\eta))| : 0 \le \xi,\eta \le 2\pi\} \ .$$

The table below gives results for the case $\phi(r) = r$. Increasing n by the factor 4 corresponds to halving the mesh size of a square grid, so we have support for the $O(h^3)$ convergence result for interpolation on $h\mathcal{Z}^2$ in two dimensions.

n	$d=5$	$d=40$
256	3.1×10^{-2}	1.1×10^{-2}
1024	4.1×10^{-3}	1.5×10^{-3}
4096	5.1×10^{-4}	1.8×10^{-4}

Table 5.2. *Values of* $\|f-s\|_\infty$, $\phi(r) = r$.

In the experiments on increases in the multiquadric parameter c, the data $f(\underline{x}_i)$, $i = 1, 2, \ldots, n$, are as before with n fixed at 4096. We pick $c = (\pi\sigma/32)\sqrt{d}$ for the values of σ in the following table.

σ	$d=5$	$d=40$
0	5.1×10^{-4}	1.8×10^{-4}
0.5	2.1×10^{-5}	2.4×10^{-6}
1.0	3.3×10^{-6}	1.7×10^{-8}
2.0	3.0×10^{-7}	3.8×10^{-12}

Table 5.3. *Values of* $\|f-s\|_\infty$ *when* $n = 4096$, $\phi(r) = \sqrt{r^2+c^2}$.

All the gains in accuracy are due to σ. These experiments are analogous to interpolation on $\mathcal{R}^2$, so one should pick a ϕ that is suitable when $d = 2$, regardless of the actual value of d. The dimension of the manifold may be unknown, however, when the interpolation method is applied.

Edge effects when $d = 1$ and $\phi(r) = r^3$

Let the values $f(x_i)$, $i = 1, 2, \ldots, n$, of $f : \mathcal{R} \mapsto \mathcal{R}$ be interpolated by the function

$$s(x) = \sum_{j=1}^{n} \lambda_j\, |x - x_j|^3 + \alpha + \beta x, \quad x \in \mathcal{R}\ ,$$

where $\sum_{j=1}^n \lambda_j = \sum_{j=1}^n \lambda_j\, x_j = 0$. Thus s is the "natural cubic spline" that minimizes $\int_{-\infty}^{\infty} [s''(x)]^2 dx$ subject to $s(x_i) = f(x_i)$, $i = 1, 2, \ldots, n$. We assume $x_i = (i{-}1)h$, $i = 1, 2, \ldots, n$, where $h = 1/(n{-}1)$, so we are interpolating at equally spaced points on $[0, 1]$.

If f has bounded fourth derivatives, the order of accuracy $\max\{|f(x){-}s(x)| : 0{\le}x{\le}\ 1\} = O(h^4)$ is achieved as $h \to 0$, if the constraints on λ_j, $j = 1, 2, \ldots, n$, are replaced by $s'(0) = f'(0)$ and $s'(1) = f'(1)$. On the other hand, one can verify that the natural cubic spline satisfies $s''(0) = 0$ and $s''(1) = 0$, which introduces errors of magnitude h^2 near the ends of the interval $[0, 1]$ when $f''(0)$ and $f''(1)$ are nonzero. Nevertheless, the natural cubic spline enjoys the property $\max\{|f(x)-s(x)| : \varepsilon \le x \le 1-\varepsilon\} = O(h^4)$ as $h \to 0$, where ε is any positive constant.

Multiquadric interpolation on $[0, 1]$

We retain the data $f(x_i)$, where $x_i = (i-1)h = (i-1)/(n-1)$, $i = 1, 2, \ldots, n$. Beatson and Powell (IMAJNA **12**, pp 107–133, 1992) investigate interpolation by a function of the form

$$s(x) = \sum_{j=1}^{n} \lambda_j \left\{(x-x_j)^2 + h^2\right\}^{1/2} + \alpha + \beta x, \quad x \in \mathcal{R} ,$$

with two constraints on the coefficients. Let f have bounded second derivatives. In the usual case,

$$\textstyle\sum_{j=1}^{n} \lambda_j = \sum_{j=1}^{n} \lambda_j \, x_j = 0 ,$$

the error $|f(x)-s(x)|$ is $O(h^2)$ as $h \to 0$ when x is well inside the interval $[0, 1]$, but in general the accuracy is only $O(h)$ near the ends of the range. They show by numerical experiments, however, that the accuracy seems to become $O(h^2)$ uniformly, if the constraints are replaced by the conditions $s'(0) = f'(0)$ and $s'(1) = f'(1)$. I have a long unpublished proof of this property.

Instead of the constraints, one can take up the two degrees of freedom in the coefficients by minimizing $\sum_{j=1}^{n} \lambda_j^2$. Numerical experiments in this case also suggest $\max\{|f(x)-s(x)| : 0 \le x \le 1\} = O(h^2)$.

Thin plate spline interpolation on $[0, 1]$

We consider interpolation to $f(x_i)$, $i = 1, 2, \ldots, n$, with $x_i = (i-1)h$, $i = 1, 2, \ldots, n$, and $h = 1/(n-1)$, by the thin plate spline function

$$s(x) = \textstyle\sum_{j=1}^{n} \lambda_j (x-x_j)^2 \log|x-x_j| + \alpha + \beta x, \quad x \in \mathcal{R} ,$$

subject to $\sum_{j=1}^{n} \lambda_j = \sum_{j=1}^{n} \lambda_j \, x_j = 0$. Bejancu (Cambridge PhD dissertation) has proved the property

$$\max\{|f(x)-s(x)| : \varepsilon \le x \le 1-\varepsilon\} = O(h^3)$$

as $h \to 0$, when f has bounded third derivatives, ε being any positive constant. This result agrees with the accuracy of interpolation on $h\mathcal{Z}$.

In particular, if $f(x) = x^2$, $x \in \mathcal{R}$, then the $O(h^3)$ accuracy of interpolation on $[0, 1]$ is shown by the following values of $e(x) = f(x) - s(x)$, $x \in \mathcal{R}$.

h	1/32	1/64	1/128
$e(\frac{1}{2}+\frac{1}{2}h)$	-2.08×10^{-6}	-2.60×10^{-7}	-3.24×10^{-8}

Table 5.4. $\phi(r) = r^2 \log r$.

In this case, however, interpolation on $h\mathcal{Z}$ gives $s \equiv f$. Therefore the error function of interpolation on $[0, 1]$ includes contributions of magnitude h^3 that are due to the range $0 \le x \le 1$.

The dependence of $e = f - s$ on h near the ends of the range $[0, 1]$ is shown by the following values of $e(\frac{1}{2}h)$, where we are still employing thin plate spline interpolation to

$f(x) = x^2$, $0 \le x \le 1$, as in the previous paragraph. Other examples and some analysis support the conjecture that $|e(\frac{1}{2}h)|$ is of magnitude $h^{3/2}$ as $h \to 0$.

h	1/80	1/160	1/320	1/640
$e(\frac{1}{2}h)$	1.4×10^{-4}	4.8×10^{-5}	1.7×10^{-5}	6.0×10^{-6}

Table 5.5. $\phi(r) = r^2 \log r$.

In one of the other examples, f has a bounded fourth derivative and is nonzero only on the interval $[\frac{1}{4}, \frac{3}{4}]$. In this case $e(\frac{1}{2}h)$ remains of magnitude $h^{3/2}$ for thin plate spline interpolation at the points $x_i = (i-1)/(n-1)$, $i = 1, 2, \ldots, n$, which was unexpected.

Future research on edge effects

Proving conjectures that arise from numerical experiments is difficult, but I am trying to make some contributions.

On the practical side I intend to study the following idea. Let the data be $f(\underline{x}_i)$, $i = 1, 2, \ldots, n$, as usual, where $\underline{x}_i \in \mathcal{D} \subset \mathcal{R}^d$, and let $\widehat{\mathcal{D}} \subset \mathcal{R}^d$ be such that $\mathcal{D}$ is well inside $\widehat{\mathcal{D}}$. Calculate the interpolant

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

by minimizing $\sum_{j=1}^{\widehat{n}} \lambda_j^2$ subject to $s(\underline{x}_i) = f(\underline{x}_i)$, $i = 1, 2, \ldots, n$, after choosing extra centres $\underline{x}_i \in \widehat{\mathcal{D}} \backslash \mathcal{D}$, $i = n+1, n+2, \ldots, \widehat{n}$. Thus $s(\underline{x})$, $\underline{x} \in \mathcal{D}$, may not be damaged by edge effects.