Eidgenössische Technische Hochschule Zürich
Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

M.J.D. Powell's work in univariate and multivariate approximation theory and his contribution to optimization

M.D. Buhmann and R. Fletcher[1]

Research Report No. 96-16
October 1996

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

[1]Department of Mathematics and Computer Science, University of Dundee, Dundee DD1 4HN, Scotland, UK

M.J.D. Powell's work in univariate and multivariate approximation theory and his contribution to optimization

M.D. Buhmann and R. Fletcher [1]

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Research Report No. 96-16 October 1996

Abstract

Since 1966, exactly 30 years ago, Mike Powell has published more than 40 papers in approximation theory, initially mostly on univariate approximations and then, focussing especially on radial basis functions, also on multivariate methods. A highlight of his work is certainly his book *Approximation theory and methods*, published by CUP in 1981, that summarizes and extends much of his work on ℓ_1, ℓ_2, ℓ_∞ theory and methods, splines, polynomial and rational approximation etc. It is still one of the best available texts on univariate approximation theory. In this short article we attempt to introduce part of Mike's work, with special emphasis on splines in one dimension on the one hand and radial basis functions on the other hand. Only a selection of his papers can be considered, and we are compelled to leave out all of his many software contributions, which for Mike are an integral part of his research work, be it for the purpose of establishing new or better methods for approximation or for making them more accessible to the general public through library systems. We subdivide this chapter into three parts (ℓ_1/ℓ_∞-approximation, rational approximation; splines; multivariate (radial basis function) approximation) although this is in variance with the spirit of many of Mike's articles which often establish beautiful links between different themes (e.g. optimization and ℓ_1-approximation). As will be seen, many of the papers contain optimal results in the sense that constants in error estimates are best (or the best ones known), have also often surprising novelty and always clearly defined goals. One further important contribution that we cannot describe here is Mike's guidance for the seven dissertations in approximation theory that were written under his supervision.

In a second chapter, Mike's contributions to optimization are reviewed with a special emphasis on the historical development of the subject and the impact of Mike's work on it.

[1] Department of Mathematics and Computer Science, University of Dundee, Dundee DD1 4HN, Scotland, UK

1. A brief review of M.J.D. Powell's work in univariate and multivariate approximation theory

We commence with his work on univariate approximation.

1.1 ℓ_1/ℓ_∞-approximation, rational approximation.

Mike's work on these topics begins with contributions to best polynomial approximation step. A particular concern of his is their efficient computation, e.g. with the Remez (or exchange) algorithm. In his paper with Alan Curtis [13][1], the exchange algorithm for best ℓ_∞ approximation from finite dimensional (especially polynomial) spaces is studied with a view to its convergence properties. In particular, formulas for first and second partial derivatives of

$$\max_{\substack{x=x_i\\ i=0,1,\ldots,n}} |f(x)-\phi(x,\mu_1,\ldots,\mu_n)|$$

are given with respect to the reference points x_i, $i=0,1,\ldots,n$, where f is to be approximated from $\{\phi(x,\mu_1,\ldots,\mu_n)|\mu_i\in\mathbb{R},\, i=0,1,\ldots,n\}$. If, for instance the first and second derivatives are zero, the required stationary points are reached with quadratic convergence rate when the reference prints are moved one at a time. In other words, conditions are given in this paper under which the exchange algorithm is particularly efficient.

A year later, in the same journal, there appeared a paper [20] that considers error estimates of polynomial approximation. Lebesgue numbers (norm of the interpolation operator for polynomial interpolation) are used to show that the maximum error of interpolation is always (within a factor independent of f) a multiple of the least maximum error. Chebyshev points on an interval are demonstrated as a good choice of interpolation points. The same analysis is also applied to best least-squares approximation. In the following work [23], properties of optimal knot positions for the latter approximation method are studied.

Rational, rather than polynomial, ℓ_∞ approximation is the subject of a joint paper with Barrodale and Roberts [41] where it is shown that the ordinary differential correction (ODC) algorithm is, surprisingly, in various senses better than the modified one that is more often used. The ODC is employed to find a best ℓ_∞ approximation by P/Q, $P\in\mathbb{P}_m$, $Q\in\mathbb{P}_n$, on the basis of points $X=\{x_1,x_2,\ldots,x_N\}$, where $Q(x_i)>0$ $\forall i=1,2,\ldots,N$. Although the maximum error Δ_k over X may tend to Δ^* in the course of the algorithm, where P_k, Q_k are computed from P_{k-1} and Q_{k-1} by minimizing

$$\max_{i=1,2,\ldots,N} \frac{|f(x_i)\,Q(x_i)-P(x_i)|-\Delta_{k-1}\,Q(x_i)}{Q_{k-1}(x_i)},$$

the problem need not have a solution P^*, Q^* with the required properties for which Δ^* is attained. However, convergence $P_k\to P^*$, $Q_k\to Q^*$ is proved in [41] when $N\ge n+m+1$. The best approximation exists and satisfies the normalization condition $\max_i |q_i| = 1$.

[1] Numbers in square brackets refer to the general list of M.J.D. Powell's publications.

Here, the q_i are the coefficients of Q. Furthermore, the minimization procedure involved is a linear programming problem, and quadratic convergence of the coefficients is established. Incidentally, this is a beautiful link between approximation theory and an application of an optimization (specifically, here, linear programming) method, four more of which we will encounter below.

Often, we find surprising novelties in his papers at points where nothing new was expected. Examples will be in his work on radial basis functions outlined below, but also the problem dealt with in the last paragraph was revisited for new results in [108]. There, general rational approximations are considered, i.e. ones where the numerator and the denominator need no longer be polynomial. However, it is sensible to restrict them to finite dimensional spaces, G and H say, and the denominator should be from the set $H_+ := \{h \in H \,|\, h > 0,\ \|h\| = 1\}$ which is assumed to be non empty, and $\|\cdot\|$ is a prescribed norm.

Again, the unmodified, ordinary ODC algorithm is analyzed and superlinear convergence is shown if the best approximation with respect to $\|\cdot\|$ exists and is unique and if $\Delta_k \to \Delta^*$. If $\inf Q_n \to 0$ and no unique best approximation exists, an example of just linear convergence is given, even if P_k and Q_k tend to a unique limit for $k \to \infty$. Thus the former result is optimal with respect to its hypotheses. It is well-known that the best approximation may not exist uniquely even if $\Delta_k \to \Delta^*$ but $\inf Q_k \to 0$, although P_k and Q_k may have a unique limit. Another approach to solve this problem, that has been suggested elsewhere, is to restrict Q_k away from zero, $Q_k \geq \epsilon > 0$, where ϵ is independent of k, during the algorithm.

Another link between approximation and optimization occurs in ℓ_1 theory. In [79], a linear programming test for ℓ_1 optimality of an approximation is given. It consists of a finite number of linear inequalities and is therefore suitable for the application of an LP method. These inequalities are expressed in terms of the original data, which is highly suitable for practical computations of best ℓ_1 approximations. Those approximations have been, incidentally, much neglected elsewhere in the literature, with the exception of A. Pinkus' excellent book on ℓ_1 approximations. The characterization theorem is best stated explicitly. We let X be a discrete finite set, $C(X)$ be the set of functions that are defined on $X, A \subset C(X)$ an n-dimensional subspace. Furthermore, we let $Z = \{z_j\}_{j=1}^n \subset X$ be the zero set with n elements of the function $s(x) = f(x) - \phi^*(x)$, where $f \in C(X)$ is the function to be approximated, $\phi^* \in A$. Then ϕ^* is best ℓ_1 approximant to f if and only if the following inequalities hold:

$$\Big| \sum_{x \in X \setminus Z} \operatorname{sign} s(x)\ \ell_i(x) \Big| \leq 1 \quad \forall i\ ,$$

where $\ell_i \in A$ satisfy $\ell_i(z_j) = \delta_{ij},\ i = 1, \ldots, n$. There are many advantages to using ℓ_1 approximations in practice and this paper is a concrete aid to their application.

Finally in this subsection we mention the articles [128], [129], where approximants to discrete noisy data are constructed to obtain either piecewise monotonicity of the data or convexity. Precisely, if, for instance, at most k monotone sections of the data are desired (k is prescribed) and n (univariate) data are given, the k optimal breakpoints and the least changes to the data are computed in $0(n^2 + kn \log n)$ operations. "Least changes" is understood in the sense of a global sum of squares of changes. The special cases $k = 1, 2$ give the minimum complexity $O(n)$. The principal advancement in this work is the substantial reduction of the number of data that need be considered when finding the optimal breakpoints. A recursive method is applied with respect to k, and certain subsets of the data have to be used in the computation at each stage.

In [129], the least sum of squares of changes to the data is sought to achieve convexity. The method uses the iterative optimization algorithm of Goldfarb and Idnani, for which a starting value is computed in $O(n)$ operations. Precisely, the statement of the convexity constraints in terms of second divided differences of the data gives rise to a strictly convex quadratic programming problem which is subsequently solved by the above algorithm. Mike's talk at the conference celebrating his 60^{th} birthday this year was also closely related to this topic. In [89], similar questions are considered with respect to the least uniform change to the data. Algorithms are given to compute the least maximum change to the data in order to achieve monotonicity, piecewise monotonicity, or piecewise convex/concave data.

1.2. Splines

In this subsection, a short summary of some of Mike's work on splines is presented. The best source to this work is generally his book [81] where much attention is given to polynomial splines, B-splines, spline interpolation and convergence properties, but we extract his work here from the papers [29], [32] and [63].

Cubic splines are considered in [29] and they are employed for the purpose of providing least squares approximations with weights. The main purpose of the article is an analysis of the locality of the least squares approximation by splines. The (finite number of) knots $\ell h, \ell \in \mathbb{Z} \cap [0, M]$, say, of the spline are equidistant with spacing h and the weight function is h-periodic and nonnegative. Using recurrence relations, the fundamental functions for this spline approximation problem are computed where, for simplicity, the knots are assumed to be all ℓh, $\ell \in \mathbb{Z}$. They decay exponentially unless the weight is concentrated solely at the midpoints between the knots by a δ-function $h \cdot \delta(x - \frac{1}{2}h)$ (precisely, by its h-periodisation). The fastest decay of $(2-\sqrt{3})^\ell$, $\ell \to \pm\infty$, is obtained when the weight is concentrated at the knots, by $h \cdot \delta(x)$. A further consideration is given to the case when the weighted ℓ_2-norm is augmented by the sum of squares of the coefficients c_j where the spline is $s(x) = \sum_{j=0}^{M} c_j (x - jh)_+^3$ plus a cubic polynomial, and this augmentation may again be weighted by a positive factor ϑ. Indeed, by employing this factor, the localisation of the spline's dependence on the data can be strengthened, and also the best ϑ is given for the most unfavourable choice of the weight function in the least squares integral, namely $h \cdot \delta(x - \frac{1}{2}h)$, when the knots are still presumed to be integer multiples of $h > 0$. The fundamental function centred at 0 for the best approximation using that weight function and ϑ is shown to diminish as $(0.3613)^\ell$, when its argument is ℓh, $\ell \in \mathbb{Z}$. This damping term can also give uniqueness of the best approximant, which may otherwise be lost when the weight function is not periodic. Moreover, differences of the approximant are considered, e.g. if the eighth derivative of the approximand is bounded, they are $O(h^8)$ and if $f \in \mathbb{P}_7$, the error is zero except for dependencies at the ends of the range, when the approximand and approximant are defined on a finite interval.

In [27], the norm of the spline interpolation operator is estimated as well as the deterioration of the localisation of the fundamental functions when the degree of the splines becomes larger. In [32], least squares approximations to discrete data, and in particular an adaptive method for computing them when the knots are allowed to move are considered and we describe the approach in some detail now. These papers reflect Mike's interest in smoothing techniques, just like the articles discussed at the end of the previous subsection, which are highly relevant to many applications.

In [32], as in [29], a weighted sum of squares of the discrete error plus a smoothing term is to be minimized by a cubic spline, whose knots are to be determined. The smoothing term is itself a weighted sum of squares, namely of the coefficients of the truncated third powers that appear in the spline; those coefficients reflect the contribution of the third derivative discontinuities of the spline and their size is therefore a measure of its smoothness. Mike makes a distinction between knots and gnots of the spline in this paper, the latter being added whenever there is a trend found in the residual (the error function at the data ordinates). The test for trends is applied locally between the current gnots. The former knots are added so that the total distribution of the splines breakpoints remains balanced in a certain sense even if gnots are accumulating. Particular attention is given to the weights in the smoothing expression which depend on the distance of knots and gnots and on the weighted sum of squares of the residuals. We recall from [29] that the smoothing term can also cause the approximation to depend more locally on the data.

Another view of optimal knot positions, now for plain, unweighted least-squares approximation, is presented in the theoretical work [23]. Splines of degree n with N knots are studied and the goal is to minimize

$$\int_a^b (f(x) - s(x))^2 \, dx ,$$

where the function $f \in L^2[a, b]$ is bounded. A necessary condition for the optimality of a knot x_j is

$$\int_a^b f(x) \, \tilde{c}_j^{(n+1)}(x) \, dx = 0 ,$$

where $\tilde{c}_j$ is the spline of degree $2n + 1$ with simple knots $a < x_1 < x_2 < \ldots < x_N < b$ and an extra knot at x_j, which satisfies the Hermite conditions

$$\begin{aligned} \tilde{c}_j^{(p)}(a) &= \tilde{c}_j^{(p)}(b) , \quad 0 \le p \le n , \\ \tilde{c}_j(x_\ell) &= 0 , \qquad\quad 1 \le \ell \le N , \\ \tilde{c}_j'(x_j) &= 1 . \end{aligned}$$

The above condition is a consequence of the orthogonality of $\tilde{c}_j^{(n+1)}$ to all splines with knots $x_1 < x_2 < \ldots < x_N$ and an explicit expression for the n-fold integral of the approximation's error function evaluated at x_j. A suitable minimization algorithm is available from [8] for minimizing the error functional with respect to the varying knots. An essential tool in deriving the above condition is the application of integration by parts in order to reformulate orthogonality relations with splines to point evaluations of splines of higher degree. The article [23] is one of the first papers where this highly useful technique is used.

In the 1970's, the optimal interpolation problem was widely discussed. In particular, much attention was paid to finding the least pointwise error bound using a multiple (that is to be determined pointwise) of $\|f^{(k)}\|_\infty$, k prescribed. Here f is the function to be approximated, of sufficient smoothness, and $\|\cdot\|_\infty$ is the uniform norm over the interval where f is defined. Hence an approximant s and a function c are sought, where c is smallest for every x while satisfying

$$|f(x) - s(x)| \leq c(x)\, \|f^{(k)}\|_\infty ,$$

and s should be a linear combination of the given values of f at $\{x_i\}_{i=1}^m$, $m \geq k$, in the interval; the coefficients depending on x of course. It turns out that s is a spline of degree $k-1$ with $m-k$ knots, and so is c. This is identified in [63] jointly with P. Gaffney by solving the problem first for $\|f^{(k)}\|_\infty = M$, i.e. a fixed value, by

$$\begin{aligned} s = s_M &= \frac{1}{2}\,(u(x,M) + \ell(x,M)) \\ c = c_M &= \frac{1}{2M}\,|u(x,M) - \ell(x,M)| \,. \end{aligned}$$

Here u and ℓ are defined by

$$\frac{\partial^k}{\partial x^k}\, u = \delta_u(x), \qquad \frac{\partial^k}{\partial x^k}\, \ell = \delta_\ell(x) \,,$$

and δ_ℓ, δ_u are the piecewise constant functions with $|\delta_u| = |\delta_\ell| = \|\delta_u\|_\infty = \|\delta_\ell\|_\infty = M$ and have $m-k$ sign changes consistent with the signs of the data. Their signs alternate, beginning with $+1$ for δ_u and -1 for δ_ℓ, and their discontinuity points have to be computed by solving a certain nonlinear system of equations. The final s and c are thus obtained by letting $M \to \infty$.

In [68] a review of bivariate approximation tools is given, mostly tensor product methods, according to the state-of-the-art at the time, where not just point interpolation, but general linear operators are considered; e.g. least-squares approximation is considered. Also irregular data receive attention; Mike especially discusses weighted local least square approximation by piecewise linears, considers triangulations of domains in two dimensions, etc. A new state-of-the-art conference has taken place this year (1996) in York, and Mike spoke again on multivariate approximation. It is interesting to compare the two papers – twenty years apart – and observe the development of the subject. Indeed, in the new paper there is much more attention given to general multivariate tools, albeit only for interpolation. Most of the work reviewed concerns radial functions which we describe in the following section. Piecewise linears and higher order polynomials are discussed in connection with algorithms for generating Delauney triangulations. Local interpolation schemes, such as Shepard's method and its generalisations are mentioned too, as are moving least squares and natural neighbourhood interpolants.

The article [69] deserves an especially prominent mentioning because the Powell–Sabin-split is very well-known and much used in the CAGD and finite element communities. The theme of that paper is the creation of a globally continuously differentiable piecewise quadratic surface on a triangulation. Function values and derivatives are given at the vertices of each triangle of a partition. In order to have the necessary number of degrees of freedom, each triangle is subdivided into six (or twelve) subtriangles which require additional, interior C^1 conditions. In total, the number of degrees of freedom turns out to be precisely right (nine) in the six subtriangle case; extra degrees of freedom can be taken up by prescribing also normals on edges which may be computed by linear interpolation in an approximation. How are the interior subtriangles selected? In the six subtriangle case, that is the one most often referred to as the Powell–Sabin-element, one takes the edges of the interior triangles from the midpoints of the edges of the big triangle to its circumcentre. The triangles are required to be always acute. In other words, the midpoint inside the big triangle is the intersection of the normals at the midpoints of the edges. This ensures that the midpoint inside the big triangle lies in the plane spanned by the points exactly between it and the edge's midpoint, which is needed for the interior C^1 continuity. As indicated above, extra degrees of freedom are there when the twelve triangle split is used and thus, e.g., the condition that all triangles of the triangulation be acute may be dropped. This paper is just another case where Mike's work (and, of course, that of his co-authors, although most of his papers he has written alone) initiated a large amount of lasting interest in research and applications. Another example displaying much foresight is his work on radial basis functions that we will describe in the next section.

1.3. Radial basis functions

Mike Powell was and still is one of the main driving forces behind the research into radial basis functions. It indeed turned out to be a most successful and fruitful area, leading to many advances in theory and applications. Much of the interest within the mathematical community was stimulated by Mike's first review paper [107] that discusses recent developments in this new field. It addresses the nonsingularity properties of the interpolation matrices $\{\phi(\|x_i - x_j\|_2)\}_{i,j=1}^m$ for interpolation at distinct points $x_j \in \mathbb{R}^n$, where $\phi(r) = r$ or $\phi(r) = e^{-r^2}$ or $\phi(r) = \sqrt{r^2 + c^2}$, c a positive parameter. Indeed, for any n and m, those matrices are always nonsingular, admitting unique solvability of the pointwise interpolation problem from the linear space spanned by $\phi(\| \cdot - x_j\|_2)$. Proofs of these results are provided and examples of singularity are given for 1-norms and ∞-norms replacing the Euclidean norm. It is incidentally indicated in that paper that Mike's interest in radial basis function methods was stimulated by the possibility of using them to provide local approximations to functions required within optimization algorithms.

Motivated by these remarkable nonsingularity results, at first two of Mike's research students worked in this field, their early results being summarized in [114]. Their work concentrated initially on the question whether polynomials are contained in the linear spaces spanned by (multi-) integer translates of the radial basis function $\phi(r) = \sqrt{r^2 + c^2}$, where the underlying space is $\mathbb{R}^n$. For odd $n \geq 1$, all polynomials of degree n were shown to exist in those spaces for $c = 0$. For $n = 1$, $c > 0$, the same was shown for linear polynomials. In both cases, the polynomials p are generated by quasi-interpolation

$$p(x) = \sum_{j \in \mathbb{Z}^n} p(j)\, \psi(x - j)\ , \quad x \in \mathbb{R}^n\ ,$$

where ψ is a finite linear combination of integer translates of those radial basis functions. The first task is to show that ψ exist which decay sufficiently fast to render the above infinite sum absolutely convergent when p is a polynomial. Since this means also that they are **local**, convergence order results on scaled grids $h\mathbb{Z}^n$ follow as well, albeit their proofs made much more complicated by the lack of compact support of ψ (only algebraic decay in comparison to the compact support of multivariate splines). This work, in turn, motivated [122] where such ψ are used to compute fundamental Lagrange functions for interpolation on (finite) cardinal grids. A Gauss–Seidel type algorithm is used to invert the matrix $\{\psi(j - k)\}$, where j and k range over a finite subset of $\mathbb{Z}^n$. The matrix is amenable to this approach because the ψ decay in fact quite fast (linked to the aforementioned polynomial recovery). Faster computation can be obtained by working out ψ as a linear combination of translates of the multiquadric function explicitly. There is even a link in this work with an optimization package [119] that Mike wrote, because the absolute sum of the off-diagonal elements of the Gauss–Seidel matrix was minimized subject to a normalization condition and the coefficient constraints that give algebraic decay. The first step from equally-spaced centres to nonequally-spaced ones was taken again by Mike Powell in [126], where a careful study of the univariate spaces generated by multiquadric translates led to a creation of quasi-interpolants by such radial functions with nonequally spaced x_j in one dimension. The spaces created by the radial function's translates and by the ψs are quite different (both are completely described in [126]) and the latter is shown to contain all linear polynomials. Central to this work is the description of the ψ functions that generate the quasi-interpolants by the Peano kernel theorem, the Peano kernel being the second derivative of the multiquadric function.

Focussing further on multiquadric interpolation, Mike studied in [130] its approximational accuracy on the unit interval when the centres are $h \cdot j$, $j = 0, \ldots, m$, where $c = h = 1/m$. Accuracy for Lipschitz continuous f of $O(h)$ is established when $h \to 0$. A key ingredient to his error estimate, which is valid uniformly on the whole interval, is a careful consideration of the size of the elements of the interpolation matrix' inverse, including the boundary elements. If two conditions on f' at 0 and 1 are satisfied additionally, then $O(h^2)$ accuracy can be shown.

The work of [126] has been carried further by Beatson and Powell [134] by studying three different quasi-interpolants with non-equally spaced x_j. The first quasi-interpolant is in the $(m + 1)$-dimensional space spanned by $\{\phi(\cdot - x_j)\}_{j=1}^m \cup \{\text{constant functions}\}$. The second is in the space spanned just by the m translates of ϕ. Quasi-interpolant number three is in the same space, enlarged by linear polynomials. The orders of accuracy obtained for the three quasi-interpolants are $(1+h^{-1}c)\omega_f(h)$, ω_f being the modulus of continuity of f and $h=\max_{2\le j\le m}(x_j-x_{j-1})$, $c\{|f(x_1)+f(x_m)|\}(x_m-x_1)^{-1}+(1+h^{-1}c)\omega_f(h)$ and

$$\tfrac{1}{4}\,L\Big(c^2(1+2\log(1+(x_m-x_1)/c))+\tfrac{1}{2}h^2\Big)\ ,$$

respectively. For the last estimate, f' is assumed to satisfy a Lipschitz condition with the Lipschitz constant L.

Of course, this quasi-interpolant is the most interesting one which provides the best (i.e. second order except for the $\log c$ term) accuracy. Still, the other quasi-interpolants deserve attention too. The second one because as it is an approximant from the space spanned only by translates of the multiquadric function. This can be considered a very natural space, because one usually interpolates from just that space. Finally, the first one deserves attention because it can be written as

$$s(x)=\text{const}+\sum_{j=1}^{m}\lambda_j\phi(x-x_j),\qquad x\in[x_1,x_m],$$

with an extra condition to take up the additional degree of freedom

$$\sum_{j=1}^{m}\lambda_j=0\ .$$

This is the form that corresponds naturally to the variational formulation of radial basis function approximants such as the thin plate spline approximants.

In [133], the results of [130] are extended by considering interpolation with centres and data points as previously, using translates of the multiquadric function plus a general linear polynomial. The multiquadric parameter c is always a positive multiple of h. In the article, various ways to take up these extra degrees of freedom in such a way that superlinear convergence $o(h)$ is obtained are suggested (the authors conjecture that this is in fact $O(h^2)$, as supported by numerical evidence and the results of the previous paragraph). If the added linear polynomial is zero, then one cannot obtain more than $O(h)$ convergence for general twice-differentiable functions unless the function satisfies boundary conditions. If a constant is added to the multiquadric approximant (and the extra degree of freedom is taken up by requiring that the coefficients of the multiquadric functions sum to zero), then superlinear convergence to twice continuously differentiable f is obtained if and only if $f'(0)=f'(1)=0$. If a linear polynomial is added and additionally to the display above $\sum_{j=1}^{m}\lambda_j x_j=0$ is required, then superlinear convergence to $f\in\mathcal{C}^2([0,1])$ is obtained if and only if $f'(0)=f'(1)=f(1)-f(0)$.

Apart from providing these necessary and sufficient conditions for superlinear convergence, Beatson and Powell suggest several new ways to take up the extra degrees of freedom in such a way that superlinear convergence is always obtained for twice continuously differentiable approximands; there is a proof that this is indeed the case for one of the methods put forward. They include interpolating f' at 0 and 1, interpolating f at $\frac{1}{2}h$ and $1 - \frac{1}{2}h$, and minimizing the sum of squares of interpolation coefficients. The latter is the choice for which superlinear convergence is proved.

In [132], Mike's opus magnum, he summarizes and explains many recent developments, including nonsingularity results for interpolation, polynomial reproduction and approximation order results for quasi-interpolation and Lagrange-interpolation on cardinal grids for classes of radial basis functions, including all of the ones mentioned above and thin plate splines $\phi(r) = r^2 \log r$, inverse (reciprocal) multiquadrics, and several others. The localisation of Lagrange functions for cardinal interpolation is considered in great detail and several improvements of known approximation order results are given. Much like his earlier review papers and his book, this work also does not just summarize his and other authors' work, but offers simplifications, more clarity in the exposition and improvements of results.

A further nice connection between approximation and optimization techniques can be found in [147] where approximants $s : \mathbb{R}^2 \to \mathbb{R}^2$ are considered that are componentwise thin-plate splines. The goal is to find a mapping between two regions in $\mathbb{R}^2$, where certain control points and control *curves* are mapped to prescribed positions. Mapping control points to points with the TPS method is not hard, but a curve must be discretised and it is not clear whether the discretization is the same in the image region even though the curve retains its shape. Because TPS yields the interpolant of minimal second derivatives in the least-squares sense, there is already one optimizing feature in that approach. In this article, Mike uses once more the universal algorithm [119] to determine the optimal positions of the discrete points on the curve in the image. The idea is to minimize again the semi-norm of the interpolant which consists of the sum of the square-integrals of its second partial derivatives but now with respect to the positions of the points of the discretised curve. Precisely, if f_i, g_i, $i = 1, 2, \ldots, m$, are the required image values, the semi-norm of the TPS interpolant turns out to be

$$(*) \qquad 8\pi(f_i)^T \,\Phi(f_i) + 8\pi(g_i)^T \,\Phi(g_i) \,,$$

where $\Phi = \{\phi(\|x_j - x_k\|_2)\}_{j,k=1}^m$, $\phi(r) = r^2 \log r$, and x_i are the points in the original domain in $\mathbb{R}^2$.

If we only want to map points into points, i.e.

$$\{x_i\}_{i=1}^m \to \{(f_i, g_i)\}_{i=1}^m \,,$$

then $(*)$ is the minimal value of the semi-norm that can be obtained. If, for simplicity, all of the $\{x_i\}_{i=1}^m$ originate from the discretization of the curve in the original domain, $(*)$ can again be minimized with respect to the (f_i, g_i), subject to those points lying on the given curve in the image domain. In particular, they should lie in the same order on the curve as the $\{x_i\}_{i=1}^m$ which gives linear inequality constraints to the optimization procedure. It is useful to write the points in a parametric form for this purpose.

In [143], the most general (with respect to the choice of the domain of convergence) results with regard to the convergence of thin-plate splines are obtained. There are several prior articles about convergence of thin-plate spline interpolants to scattered data on domains in $\mathbb{R}^2$, but the domains have always been required to have at least Lipschitz continuous boundaries. Mike succeeds in proving convergence within any bounded domain. The speed of convergence shown is within a factor of $\log h$ (h being the largest minimum distance between interpolation points and any points in the domain), the same as the best of earlier results. On top of this, he gets the best multiplicative constants for the error estimates for interpolation on a line or within a square or a triangle, i.e. when we measure the error of thin-plate interpolation between two, three or four data points, where in the latter case, they form a square. The $\log h$ term is due to the fact that the point x where we measure the error need not be in the convex hull of the centres (though it does need to be in their h-neighbourhood, due to the definition of h).

At the time of writing this article, Mike's latest work considers the efficient solution of the thin-plate spline interpolation problem for a large volume of data. A closely related problem is the efficient evaluation of a given linear combination $s(x)$ of translates of $\|x\|^2 \log \|x\|$ many times, e.g. for the rendering on a computer screen. These two issues are related because the conditional positive definiteness of the interpolation matrix makes the CG (conjugate gradients) algorithm a suitable tool to solve the interpolation equations. And, of course, the CG algorithm needs many function evaluations of $s(x)$. One approach for evaluating $s(x)$ uses truncated Laurent expansions [136], [138] of the thin plate splines and collecting several terms $\|x - x_j\|^2 \log \|x - x_j\|$ for $\|x_j\| >> \|x\|$ into one expression in order to minimize the number of evaluations of the logarithm, a computationally expensive task. A principal part of the work involved with this idea is deciding which of the $\|x - x_j\|^2 \log \|x - x_j\|$ are lumped together. When done efficiently, however, the cost of this, plus the approximation of the lumps by single truncated Laurent expansions, is $O(\log m)$ for m centres plus $O(m \log m)$ set-up cost, small in comparison to at least $10m$ operations for direct evaluation.

Another approach for computing thin plate spline interpolants efficiently by Mike Powell and collaborators uses local Lagrange functions, i.e. Lagrange functions L_j centred at x_j, say, that satisfy the Lagrange conditions $L_j(x_k) = \delta_{jk}$ only for several x_k near to x_j. The approximant is then constructed by a multigrid-type algorithm that exploits the observation that these local Lagrange functions are good approximations to the full Lagrange functions. This is in recognition of the fact that, at least if the data form an infinite regular grid, the full Lagrange functions decay exponentially, i.e. are very well localized. Therefore it is feasible to compute the interpolant by an iterative method which at each stage makes a correction to the residual by subtracting multiples of those local Lagrange functions. The iteration attempts to reduce the residual by subtracting $\sum_{i=1}^{m}(s(x_i) - f_i)\ L_i(x)$ from s, where L_i are the local Lagrange functions and s is the previous approximation to the thin-plate spline interpolant. It turns out that the iteration converges in many test cases, because the spectral radius of the iteration matrix associated with this procedure is less than 1. In a later paper [148], a slightly different approach is used where the *coefficient* of each $\|x - x_j\|^2 \log \|x - x_j\|$ is approximated in each step of the iteration by a multiple of the leading coefficient of a local Lagrange function $L_i(x)$ (the multiplier being the residual $f_i - s(x_i)$). Therefore the correcting term to a prior approximation is now

$$\sum_{i=1}^{m} \|x - x_i\|^2 \log \|x - x_i\| \sum_{j \in \mathcal{L}_i} \mu_{ij} [s(x_j) - f_j] \; ,$$

where $\mathcal{L}_i \subset \{1, 2, \ldots, m\}$ is the set of centres used for the local Lagrange function L_i and μ_{ij} are its coefficients. This correction is performed iteratively until the required accuracy is obtained. The multigrid idea comes into play in this method as an inner iteration within each stage of the updating algorithm already described. Namely the above iterations are expected to remove the very high frequency components from the error. Therefore, there is now an inner iteration like a fine to coarse sweep of multigrid, where the set of centres is thinned out consecutively, and the updates as above are performed on the thinner sets, until just few centres are left which have not yet been considered. For those centres the correction then consists of solving the interpolation problem exactly. A remarkable observation is that the number of such iterations to obtain a prescribed accuracy seems to depend only weakly on the number of data.

2. The Contributions of Mike Powell to Optimization

I first came across Mike when I was a PhD student at the University of Leeds. I had been fortunate enough to come across a report of Bill Davidon on a variable metric method (strange words to me in those days). Having established that it was much better than anything currently available, I had dashed off a short paper which I was about to send to the *Computer Journal*, then the prime outlet in the UK for articles on numerical analysis. At the same time Mike was booked to give us a seminar on what was probably his first method for unconstrained minimization. This by the way was a ingenious way of obtaining quadratic termination in a gradient method, which however was already about to be superseded. A week before the seminar, Mike phoned and asked if he could change his title: he had come across a report of a much better method that he would like to talk about. Typically Mike had extracted the essentials of the method from the mass of detail in Davidon's flow sheets, and had also implemented it on the IBM machine at Harwell which was much faster than our modest Ferranti Pegasus at Leeds. When he heard of my interest in the method, he generously offered to pool his work with mine. We added some more things on conjugacy and such, and so was born the DFP paper [8]. Mike was also instrumental in promulgating the good news: he stood up at a meeting in London where speakers were elaborating on the difficulties of minimizing functions of 10 variables and told them that he had coded a method which had solved problems in 100 variables without difficulty. Of course this revolutionized the discipline and this type of method (but with the BFGS formula) is still the method of choice to this day.

Since that time our paths have crossed many times, most notably when Mike recruited me to work at Harwell from 1969 to 1973. Although this led to only one other joint publication, I very much benefitted from discussing my work with him, and he was especially good at exposing weak arguments or suggesting useful theoretical and practical possibilities. (Also we made a formidable partnership at table football!) I remember an incident when I first arrived at Harwell and was telling him of some code in which I had used double precision accumulation of scalar products. His reply was something along the lines of "...why bother with that, if an algorithm is numerically stable you should not need to use double precision in order to get adequate accuracy...". I think that this is good advice in the sense that you learn much about an algorithm when you develop it in single precision, which hopefully can be used to good effect in improving numerical stability at any level of precision. *Apropos* of this, I hear that the recent Ariane rocket disaster was caused by a software error arising from the use of double precision floating point arithmetic, emphasizing how important it is to pay attention to errors arising from finite precision.

A lot of interest, particularly in the mid Sixties, centred on methods for unconstrained minimization without derivatives, and Mike published an important paper [9] in this area in 1964. I recently listened to a review paper crediting Mike with discovering in this paper how to obtain quadratic termination with line searches without evaluating derivatives. In fact this idea dates back to a report due to Smith in 1962, which Mike used as the basis of his paper. Mike's contribution was to extend the algorithm by including extra line searches which would be unnecessary in the quadratic case, but would enable the method to work more effectively for nonquadratic functions, whilst retaining termination for quadratics. An important feature was the manipulation of a set of "pseudo-conjugate directions" with a criterion to prevent these directions from becoming linearly dependent. In practice however it turned out to be very difficult to beat the DFP algorithm (and later the BFGS algorithm) with finite difference approximations to derivatives, much to Mike's (and other people's) disappointment. Recently derivative-free methods have become fashionable again, with Mike in the forefront of this research [151]. He has again emphasized the importance of adequately modelling the problem functions, as against using heuristic methods like Nelder and Mead which are still well used despite their obvious disadvantages. His ideas also include a criterion to keep interpolation points from becoming coplanar, which harks back to his 1964 paper.

As methods developed, interest switched away from quadratic termination, towards proving global and local termination results. Here Mike's own contribution has been immense, and I shall say more about this in what follows. Moreover results obtained by subsequent researchers often use techniques of proof developed by him. Nonetheless Mike also has his feet on the ground and is aware of the importance of good performance on test problems and in real applications. I was struck by a phrase of his from an early paper that the performance of a method of his "...can best be described as lively..."!. Mike is also very good at constructing examples which illustrate the deficiencies in algorithms and so help rectify them. For example the Gauss–Newton method for nonlinear least squares (and hence Newton's method for equations) with an ℓ_2 line search was a popular method, especially in the early days. A so-called convergence proof existed, and the method was often used with the confidence of a good outcome. The small print in the proof that the computed search direction $\mathbf{s}$ should be bounded was not usually given much thought, although it was known that convergence could be slow if the Jacobian matrix was rank deficient at the solution. Mike showed that the situation was much more serious by devising an example for which, in exact arithmetic, the algorithm converges to a non-stationary point. This led Mike to suggest his *dog-leg algorithm* [34] (here his enthusiasm for golf is seen) in which the search trajectory consists of step along the steepest descent direction, followed by a step to the Newton point. The inclusion of the steepest descent component readily allows convergence to a stationary point to be proved, an idea that was much copied in subsequent research.

This leads me on to talk about *trust region methods* and it was interesting that there was some discussion at the birthday conference as to the historical facts. Of course the idea of modifying the Hessian by a multiple of a unit matrix, so as to induce a bias towards steepest descent, appears in the much referenced papers by Levenberg and by Marquardt. Marquardt gives the equivalence with minimizing the model on a ball, and attributes it to Morrison in 1960. Mike's contribution, which also first appears in [34] and then in [37], is to use the step restriction as a primary heuristic, rather than as a consequence of adding in a multiple of the unit matrix. Also Mike suggests the now well accepted test for increasing and decreasing the radius of the ball, based on the ratio of actual to predicted reduction in the objective function. Thus the framework which Mike proposed is what is now very much thought of as the prototypical trust region method. The term *trust region* was only coined some years later, possibly by John Dennis. In passing, the paper [37] was also notable for the introduction of what came to be known as the PSB (Powell–Symmetric–Broyden) updating formula, derived by an elegant iterative process. The variational properties of this formula were discovered a little later (by John Greenstadt) and subsequently lead to the important work on the sparse PSB algorithm by Philippe Toint, who generously acknowledges Mike's contribution.

Talking about the value of small examples reminds me of another case where Mike removes the last shred of respectability from an algorithm. The idea of minimization by searching along the coordinate directions in turn dates back who knows when, and is known to exhibit slow convergence in practice. Mike's example [47] showed that it could also converge in exact arithmetic to a non-stationary point, and this removed the last reason for anyone being tempted to use the algorithm.

Without doubt, the conjecture that attracted the most interest in the 1960's was that of whether the DFP algorithm could converge to a non-stationary point. Mike accepted a bet of 1 shilling (£0.05) with Philip Wolfe that he (Mike) would solve the problem by some date. Although the method is a descent method, the result is anything but trivial, since the Hessian approximation may become singular or unbounded in the limit. Mike finally produced a convergence proof for strictly convex functions and exact line searches [38], that was a tour-de-force of analysis. I remember checking through it for him, a job which took me several days and left me thoroughly daunted at the complexity of what had been achieved. Mike later went on to prove [62] that the BFGS algorithm would converge for convex functions and an inexact line search, a more elegant result that has influenced most subsequent work on this topic. The conjecture is still open for non-convex functions, so someone can still make a name for themselves here.

Another issue over many years has been to discuss the relative merits of different formulae in the Broyden family, particularly the DFP and BFGS formulae. Great excitement was caused at a Dundee meeting in 1971 when Lawrence Dixon introduced, in a rather peripheral way, his remarkable result for non-quadratic functions that all Broyden family updates generate identical points if exact line searches are used. Mike could hardly believe this and cross-examined Lawrence closely in question time. He then retired to his room to determine whether or not the result was indeed true. Having ascertained that it was, it is characteristic of Mike that he simplified Lawrence's proof and subsequently gave great credit to Lawrence for having made the discovery. One phenomenon of particular interest is the fact that for inexact line searches, the DFP formula can be much inferior to the BFGS formula, and various not very convincing explanations were advanced. One of my favourite Mike papers [104] is the one in which he analyses a very simple case of a two-variable quadratic and does a worst case analysis. The solution of the resulting recurrence relations is very neat, and comes out with a remarkably simple result about how the DFP formula behaves, which I think provides a very convincing explanation of the phenomenon.

Another fundamental algorithm is the conjugate gradient method, particularly in its form (Fletcher and Reeves) for nonquadratic optimization. Incidentally I would like to say here that Colin Reeves was my supervisor and not, as someone once suggested, my Ph.D. student! It was a great loss to numerical analysis when he decided to become a computer scientist, and I had the greatest respect for his abilities. An important issue was the relative merits of the Fletcher–Reeves and Polak–Ribière formulae. Mike showed [67] that the PR formula was usually better in practice, but later [99] that the FR formula allowed a global convergence result, a result later extended by Al-Baali to allow inexact line searches. Mike also showed in [99] that the PR formula can fail to converge and derived a remarkable counterexample with $n = 3$ in which the sequence $\{\mathbf{x}^{(k)}\}$ is bounded and has six accumulation points, none of which is stationary. However I think Mike's most telling result [64] is for quadratics, namely that if the method is started from an arbitrary descent direction, then either termination occurs, or the rate of convergence is linear, the latter possibility occurring with probability 1. This is bad news for the use of conjugate gradient methods in situations where the sequence must be broken, such as active set methods or limited memory methods, since it forces a restart in order to avoid linear convergence.

Turning to constrained optimization, Mike wrote an important paper [30] in 1969 that originated the idea of augmented Lagrangian penalty functions, at about the same time as a related paper by Hestenes on multiplier methods. Mike introduced the idea in a different way by making shifts θ_i to the constraint functions. After each minimization of the penalty function the shifts would be adjusted so as to get closer to the solution of the constrained problem. I seem to remember a talk in which Mike described how he was led to the idea when acting as a gunnery officer, presumably during a period of national service. Adjusting the θ_i parameters is analogous to the adjustment in elevation of a gun based on how close the previous shell lands from the target. One can have no doubts that Mike's gunnery crew would be the most successful in the whole battery! This method was a considerable improvement on the unshifted penalty function method, and led to considerable interest, and indeed is still used as a merit function in some nonlinear programming codes.

Mike also played an important part in the development of sequential quadratic programming (SQP) algorithms. Although the basic idea dates back to Wilson in 1963, and had been reviewed by Beale and others, the method had not attracted a great deal of interest. This was due to various factors, such as a reluctance to use expensive QP subproblems, the issue of how to update an approximation to the potentially indefinite Hessian of the Lagrangian, and what to do about infeasible QP subproblems. Mike [72] published an idea that handled the latter difficulty and got round the indefinite Hessian problem by ignoring it – or more accurately by suggesting a modification of the BFGS formula that could be used to update a positive definite approximation. However I think his greatest contribution was to popularise the method which he did at conferences in 1977 at Dundee, Madison and Paris. His Madison paper [73] contained a justification of the idea of keeping a positive definite Hessian approximation. The term SQP was coined I think at the Cambridge NATO ASI in 1981 and the method has remained popular since that time, and is now frequently used in applications. Mike also contributed to a paper [87] on the *watchdog technique*, a title which I suspect is due to Mike, which popularised one way of avoiding difficulties such as the Maratos effect which might slow down the local convergence of SQP.

Since the early days at Harwell, Mike has diligently provided computer code so that the fullest use can be made of his ideas (although his style of programming is not what one might call structured!). It was natural therefore that he should write a QP code to support his SQP work, and he became a strong advocate [102] of the ideas in the Goldfarb–Idnani algorithm. These ideas also surfaced in his TOLMIN code for general linear constraint programming [121], and his ideas for updating conjugate directions in the BFGS formula [110]. The latter contains a particularly elegant idea for stacking up an ordered set of conjugate directions in the working matrix Z as the algorithm proceeds. These ideas may not have caught on to the same extent as other papers that he has written, perhaps because they are not addressed towards large sparse calculations which are now attracting most attention. Nonetheless they contain a fund of good ideas that merit study. However Mike has made other contributions to sparse matrix algebra, including the CPR algorithm [51] for estimating sparse Jacobian algorithms by finite differences in a very efficient way, and an extension of the same idea [75] to sparse Hessian updates.

More recently Mike has confirmed his ability to make advances in almost all areas of optimization. The Karmarkar interior point method for linear programming has spawned thousands of papers, but Mike nonetheless managed to make a unique contribution by constructing an example [125] showing that the algorithm can perform very badly as the number of constraints increases in a simple semi-infinite programming problem. Mike also continues to provoke the interior point community by his oft-expressed dislike for the need to use logarithms as a means of solving linear programming problems! Despite holding these views, Mike was nonetheless able to make an outstanding theoretical breakthrough in the convergence analysis of the shifted log barrier method for linear programming [145].

Much has been said about Mike as a formidable competitor in all that he he does, but I would like to close with a little reminiscence that sheds a different light. Mike has always been a faithful supporter of our biennial Dundee conferences, and has contributed a great deal to the success of these meetings. At one of these meetings, my Syrian Ph.D. student Mehi Al-Baali was scheduled to give a talk on his work. As is the way of things for a local Ph.D. student, he had been given a very unfavourable slot late on Friday afternoon. Mike however had to travel abroad the next day and so was planning to leave earlier in the afternoon to make the long 8 hour or so journey by road to Cambridge, to snatch a few hours sleep before his flight. However when he saw Mehi's disappointment that he would not be present at the talk, he immediately agreed to postpone his departure and stay on a extra couple of hours or so for the talk. This is typical of the many kindnesses that Mike has shown to others, and I shall always remember him for that gesture. Happy birthday Mike!

M.J.D. Powell's Publications

1. "Ground-state splitting for d^5 ^6S ions in a cubic field" (with J.R. Gabriel and D.F. Johnston), *Phys. Rev. Letters,* Vol. 5, pp. 145–146 (1960).

2. "A scheme for handling functions on a computer", AERE Report No. R.3830, Harwell Laboratory (1961).

3. "A calculation of the ground state splitting for Mn^{++} ions in a cubic field" (with J.R. Gabriel and D.F. Johnston), *Proc. Roy. Soc. A,* Vol. 264, pp. 503–515 (1961).

4. "The crystalline field parameters for dysprosium ethyl sulphate" (with R. Orbach), *Proc. Phys. Soc.,* Vol. 78, pp. 753–758 (1961).

5. "An iterative method for finding stationary values of a function of several variables", *Computer Journal,* Vol. 5, pp. 147–151 (1962).

6. "Roothaan's procedure for solving the Hartree-Fock equation", in *Numerical Solution of Ordinary and Partial Differential Equations,* editor L. Fox, Pergamon Press (Oxford), pp. 197–202 (1962).

7. "Crystalline-field splittings in cerous magnesium nitrate" (with M.J.M. Leask, R. Orbach and W.P. Wolf), *Proc. Roy. Soc. A,* Vol. 272, pp. 371–386 (1963).

8. "A rapidly convergent descent method for minimization" (with R. Fletcher), *Computer Journal,* Vol. 6, pp. 163–168 (1963).

9. "An efficient method for finding the minimum of a function of several variables without calculating derivatives", *Computer Journal,* Vol. 7, pp. 155–162 (1964).

10. "The volume internal to three intersecting hard spheres", *Molecular Physics,* Vol. 7, pp. 591–592 (1964).

11. "A method for minimizing a sum of squares of nonlinear functions without calculating derivatives", *Computer Journal,* Vol. 7, pp. 303–307 (1965).

12. "Necessary conditions for a minimax approximation" (with A.R. Curtis), *Computer Journal,* Vol. 8, pp. 358–361 (1966).

13. "On the convergence of exchange algorithms for calculating minimax approximations" (with A.R. Curtis), *Computer Journal,* Vol. 9, pp. 78–80 (1966).

14. – 17. "The general theory of linear approximation" (with D.C. Handscomb and D.F. Mayers), "The exchange algorithm on a discrete point set", "Theory of general nonlinear minimax approximation" and "An introduction to epsilon entropy", in *Methods of Numerical Approximation,* editor D.C. Handscomb, Pergamon Press (Oxford), pp. 61–71, 73–81, 155–162 and 183–190 (1966).

18. "Minimization of functions of several variables", in *Numerical Analysis: An Introduction,* editor J. Walsh, Academic Press (London), pp. 143–157 (1966).

19. "Weighted uniform sampling—a Monte Carlo technique for reducing variance" (with J. Swann), *J. Inst. Maths. Applics.,* Vol. 2, pp. 228–236 (1966).

20. "On the maximum errors of polynomial approximations defined by interpolation and by least squares criteria", *Computer Journal,* Vol. 9, pp. 404–407 (1967).

21. "Error analysis of equal-interval interpolation by cubic splines" (with A.R. Curtis), AERE Report No. 5600, Harwell Laboratory (1967).

22. "Using cubic splines to approximate functions of one variable to prescribed accuracy" (with A.R. Curtis), AERE Report No. R.5602, Harwell Laboratory (1967).

23. "On best L_2 spline approximations", in *Numerische Mathematik, Differentialgleichungen Approximationstheorie,* Birkhauser-Verlag (Basel), pp. 317–339 (1968).

24. "On the calculation of orthogonal vectors", *Computer Journal,* Vol. 11, pp. 302–304 (1968).

25. "A Fortran subroutine to invert a rectangular matrix of full rank", AERE Report No. R.6072, Harwell Laboratory (1969).

26. "A theorem on rank one modifications to a matrix and its inverse", *Computer Journal,* Vol. 12, pp. 288–290 (1969).

27. "A comparison of spline approximations with classical interpolation methods", in *Proceedings of the IFIP 1968 Congress,* North-Holland (Amsterdam), pp. 95–98 (1969).

28. "On applying Householder transformations to linear least squares problems" (with J.K. Reid), in *Proceedings of the IFIP 1968 Congress,* North-Holland (Amsterdam), pp. 122–126 (1969).

29. "The local dependence of least squares cubic splines", *SIAM J. Numer. Anal.,* Vol. 6, pp. 398–413 (1969).

30. "A method for nonlinear constraints in minimization problems", in *Optimization,* editor R. Fletcher, Academic Press (London), pp. 283–297 (1969).

31. "A survey of numerical methods for unconstrained optimization", *SIAM Review,* Vol. 12, pp. 79–97 (1970).

32. "Curve fitting by splines in one variable", in *Numerical Approximation to Functions and Data,* editor J.G. Hayes, Athlone Press (London), pp. 65–83 (1970).

33. "Rank one methods for unconstrained optimization", in *Integer and Nonlinear Programming,* editor J. Abadie, North-Holland (Amsterdam), pp. 139–156 (1970).

34. "A hybrid method for nonlinear equations", in *Numerical Methods for Nonlinear Equations,* editor P. Rabinowitz, Gordon and Breach (London), pp. 87–114 (1970).

35. "A Fortran subroutine for solving systems of nonlinear algebraic equations", in *Numerical Methods for Nonlinear Equations,* editor P. Rabinowitz, Gordon and Breach (London), pp. 87–114 (1970).

36. "A Fortran subroutine for unconstrained minimization requiring first derivatives of the objective function", AERE Report No. R.6469, Harwell Laboratory (1970).

37. "A new algorithm for unconstrained optimization", in *Nonlinear Programming,* editors J.B. Rosen, O.L. Mangasarian and K. Ritter, Academic Press (New York), pp. 31–65 (1970).

38. "On the convergence of the variable metric algorithm", *J. Inst. Maths. Applics.,* Vol. 7, pp. 21–36 (1971).

39. "Recent advances in unconstrained optimization", *Math. Programming,* Vol. 1, pp. 26–57 (1971).

40. "Some properties of the variable metric algorithm", in *Numerical Methods for Nonlinear Optimization,* editor F.A. Lootsma, Academic Press (London), pp. 1–17 (1972).

41. "The differential correction algorithm for rational L_∞ approximation" (with I. Barrodale and F.D.K. Roberts), *SIAM J. Numer. Anal.,* Vol. 9, pp. 493–504 (1972).

42. "Quadratic termination properties of minimization algorithms: 1. Statement and discussion of results", *J. Inst. Maths. Applics.,* Vol. 10, pp. 333–342 (1972).

43. "Quadratic termination properties of minimization algorithms: 2. Proofs of theorems", *J. Inst. Maths. Applics.,* Vol. 10, pp. 343–357 (1972).

44. "Problems related to unconstrained optimization", in *Numerical Methods for Unconstrained Optimization,* editor W. Murray, Academic Press (London), pp. 29–55 (1972).

45. “A Fortran subroutine for drawing a curve through a sequence of data points” (with S. Marlow), AERE Report No. R.7092, Harwell Laboratory (1972).

46. “A Fortran subroutine for calculating a cubic spline approximation to a given function”, AERE Report No. R.7308, Harwell Laboratory (1972).

47. “On search directions for minimization algorithms”, *Math. Programming,* Vol. 4, pp. 193–201 (1973).

48. “A Fortran subroutine for plotting a cubic spline function” (with S. Marlow), AERE Report No. R.7470, Harwell Laboratory (1973).

49. “Numerical methods for constrained optimization”, *IMA Bulletin,* Vol. 9, pp. 326–327 (1973).

50. “Unconstrained minimization algorithms without computation of derivatives”, *Boll. della Unione Matematica Italiana,* Vol. 9, pp. 60–69 (1974).

51. “On the estimation of sparse Jacobian matrices” (with A.R. Curtis and J.K. Reid), *J. Inst. Maths. Applics.,* Vol. 13, pp. 117–119 (1974).

52. “Unconstrained minimization and extensions for constraints”, in *Mathematical Programming in Theory and Practice,* editors P.L. Hammer and G. Zoutendijk, North-Holland (Amsterdam), pp. 31–79 (1974).

53. “Piecewise quadratic surface fitting for contour plotting”, in *Software for Numerical Mathematics,* editor D.J. Evans, Academic Press (London), pp. 253–271 (1974).

54. “On the modification of LDL^T factorizations” (with R. Fletcher), *Maths. of Comp.,* Vol. 28, pp. 1067–1087 (1974).

55. “Introduction to constrained optimization”, in *Numerical Methods for Constrained Optimization,* editors P.E. Gill and W. Murray, Academic Press (London), pp. 1–28 (1974).

56. “On the convergence of cyclic Jacobi methods” (with K.W. Brodlie), *J. Inst. Maths. Applics.,* Vol. 15, pp. 279–287 (1975).

57. “Convergence properties of a class of minimization algorithms”, in *Nonlinear Programming 2,* editors O.L. Mangasarian, R.R. Meyer and S.M. Robinson, Academic Press (New York), pp. 1–27 (1975).

58. “A view of minimization algorithms that do not require derivatives”, *ACM Trans. Math. Software,* Vol. 1, pp. 97–107 (1975).

59. “The minimax solution of linear equations subject to bounds on the variables”, in *Proc. of the Fourth Manitoba Conference on Numerical Mathematics,* Utilitas Matematica Publishing Inc. (Winnipeg), pp. 53–107 (1975).

60. "A Fortran subroutine that calculates the minimax solution of linear equations subject to bounds on the variables" (with K. Madsen), AERE Report No. R.7954, Harwell Laboratory (1975).

61. "A view of unconstrained optimization", in *Optimization in Action,* editor L.C.W. Dixon, Academic Press (London), pp. 117–152 (1976).

62. "Some global convergence properties of a variable metric algorithm for minimization without exact line searches", in *Nonlinear Programming SIAM-AMS Proceedings, Vol. IX,* editors R.W. Cottle and C.E. Lemke, American Mathematical Society (Providence), pp. 53–72 (1976).

63. "Optimal interpolation" (with P.W. Gaffney), in *Numerical Analysis Dundee 1975, Lecture Notes in Mathematics No. 506,* editor G.A. Watson, Springer–Verlag (Berlin), pp. 90–99 (1976).

64. "Some convergence properties of the conjugate gradient method", *Math. Programming,* Vol. 11, pp. 42–49 (1976).

65. "A Fortran subroutine for plotting the part of a conic that is inside a given triangle" (with S. Marlow), AERE Report No. R.8336, Harwell Laboratory (1976).

66. "Quadratic termination properties of Davidon's new variable metric algorithm", *Math. Programming,* Vol. 11, pp. 141–147 (1977).

67. "Restart procedures for the conjugate gradient method", *Math. Programming,* Vol. 12, pp. 241–254 (1977).

68. "Numerical methods for fitting functions of two variables", in *The State of the Art in Numerical Analysis,* editor D.A.H. Jacobs, Academic Press (London), pp. 563–604 (1977).

69. "Piecewise quadratic approximations on triangles" (with M.A. Sabin), *ACM Trans. Math. Software,* Vol. 3, pp. 316–325 (1977).

70. "A technique that gains speed and accuracy in the minimax solution of overdetermined linear equations" (with M.J. Hopper), in *Mathematical Software 3,* editor J.R. Rice, Academic Press (New York), pp. 15–33 (1977).

71. "Algorithms for nonlinear constraints that use Lagrangian functions", *Math. Programming,* Vol. 14, pp. 224–248 (1978).

72. "A fast algorithm for nonlinearly constrained optimization calculations", in *Numerical Analysis Dundee 1977, Lecture Notes in Mathematics No. 630,* editor G.A. Watson, Springer–Verlag (Berlin), pp. 144–157 (1978).

73. "The convergence of variable metric methods for nonlinearly constrained optimization calculations", in *Nonlinear Programming 3,* editors O.L. Mangasarian, R.R. Meyer and S.M. Robinson, Academic Press (New York), pp. 27–63 (1978).

74. “Variable metric methods for constrained optimization”, in *Computing Methods in Applied Sciences and Engineering 1977 (1), Lecture Notes in Mathematics No. 704,* editors R. Glowinski and J.L. Lions, Springer–Verlag (Berlin), pp. 62–72 (1979).

75. “On the estimation of sparse Hessian matrices” (with Ph.L. Toint), *SIAM J. Numer. Anal.,* Vol. 16, pp. 1060–1074 (1979).

76. – 77. “Gradient conditions and Lagrange multipliers in nonlinear programming” and “Variable metric methods for constrained optimization”, in *Nonlinear Optimization Theory and Algorithms,* editors L.C.W. Dixon, E. Spedicato and G. Szegö, Birkhauser (Boston), pp. 201–220 and 279–294 (1980).

78. “Optimization algorithms in 1979”, in *Optimization Techniques, Lecture Notes in Control and Information Sciences No. 22,* editors K. Iracki, K. Malanowski and S. Walukiewicz, Springer–Verlag (Berlin), pp. 83–98 (1980).

79. “A discrete characterization theorem for the discrete L_1 linear approximation problem” (with F.D.K. Roberts), *J. Approx. Theory,* Vol. 30, pp. 173–179 (1980).

80. “A note on quasi-Newton formulae for sparse second derivative matrices”, *Math. Programming,* Vol. 20, pp. 144–151 (1981).

81. *Approximation Theory and Methods*, Cambridge University Press (Cambridge) (1981).

82. “An example of cycling in a feasible point algorithm”, *Math. Programming,* Vol. 20, pp. 353–357 (1981).

83. “On the decomposition of conditionally positive-semidefinite matrices” (with D.H. Martin and D.H. Jacobson), *Linear Algebra Applics.,* Vol. 39, pp. 51–59 (1981).

84. “On the A-acceptability of rational approximations to the exponential function” (with A. Iserles), *IMA J. Numer. Anal.,* Vol. 1, pp. 241–251 (1981).

85. “An upper triangular matrix method for quadratic programming”, in *Nonlinear Programming 4,* editors O.L. Mangasarian, R.R. Meyer and S.M. Robinson, Academic Press (New York), pp. 1–24 (1981).

86. “The Shanno-Toint procedure for updating sparse symmetric matrices” (with Ph.L. Toint), *IMA J. Numer. Anal.,* Vol. 1, pp. 403–413 (1981).

87. “The watchdog technique for forcing convergence in algorithms for constrained optimization” (with R.M. Chamberlain, C. Lemaréchal and H.C. Pedersen), *Math. Programming Studies,* No. 16, pp. 1–17 (1982).

88. *Nonlinear Optimization 1981* (editor), Academic Press (London) (1982).

89. “Data smoothing by divided differences” (with M.P. Cullinan), in *Numerical Analysis Proceedings, Dundee 1981, Lecture Notes in Mathematics No. 912,* editor G.A. Watson, Springer–Verlag (Berlin), pp. 26–37 (1982).

90. "Extensions to subroutine VFO2AD", in *System Modeling and Optimization, Lecture Notes in Control and Information Sciences No. 38,* editors R.F. Drenick and F. Kozin, Springer–Verlag (Berlin), pp. 529–538 (1982).

91. "Algorithms for constrained and unconstrained optimization calculations", in *Current Developments in the Interface: Economics, Econometrics, Mathematics,* editors M. Hazewinkel and H.G. Rinnooy Kan, Reidel (Dordrecht), pp. 293–310 (1982).

92. "VMCWD: A Fortran subroutine for constrained optimization", Report No. DAMTP 1982/NA4, University of Cambridge (1982).

93. "The convergence of variable metric matrices in unconstrained optimization" (with R-P. Ge), *Math. Programming,* Vol. 27, pp. 123–143 (1983).

94. "Variable metric methods for constrained optimization", in *Mathematical Programming: The State of the Art, Bonn 1982,* editors A. Bachem, M. Grötschel and B. Korte, Springer–Verlag (Berlin), pp. 288–311 (1983).

95. "General algorithms for discrete nonlinear approximation calculations", in *Approximation Theory IV,* editors C.K. Chui, L.L. Schumaker and J.D. Ward, Academic Press (New York), pp. 187–218 (1983).

96. "ZQPCVX: A Fortran subroutine for convex quadratic programming", Report No. DAMTP 1983/NA17, University of Cambridge (1983).

97. "On the global convergence of trust region algorithms for unconstrained minimization", *Math. Programming,* Vol. 29, pp. 297–303 (1984).

98. "Conditions for superlinear convergence in ℓ_1 and ℓ_∞ solutions of overdetermined nonlinear equations" (with Y. Yuan), *IMA J. Numer. Anal.,* Vol. 4, pp. 241–251 (1984).

99. "Nonconvex minimization calculations and the conjugate gradient method", in *Numerical Analysis Proceedings, Dundee 1983, Lecture Notes in Mathematics No. 1066,* editor G.A. Watson, Springer–Verlag (Berlin), pp. 122–141 (1984).

100. "On the rate of convergence of variable metric algorithms for unconstrained optimization", in *Proceedings of the International Congress of Mathematicians Warsaw 1983,* editors Z. Ciesielski and C. Olech, North-Holland (Amsterdam), pp. 1525–1539 (1984).

101. "An application of Gaussian elimination to interpolation by generalised rational functions" (with T. Håvie), in *Rational Approximation and Interpolation, Lecture Notes in Mathematics No. 1105,* editors P.R. Graves-Morris, E.B. Saff and R.S. Varga, Springer-Verlag (Berlin), pp. 442–452 (1984).

102. "On the quadratic programming algorithm of Goldfarb and Idnani", *Math. Programming Studies,* No. 25, pp. 46–61 (1985).

103. "The performance of two subroutines for constrained optimization on some difficult test problems," in *Numerical Optimization 1984,* editors P.T. Boggs, R.H. Byrd and R.B. Schnabel, SIAM Publications (Philadelphia), pp. 160–177 (1985).

104. "How bad are the BFGS and DFP methods when the objective function is quadratic?", *Math. Programming,* Vol. 34, pp. 34–47 (1986).

105. "Convergence properties of algorithms for nonlinear optimization", *SIAM Review,* Vol. 28, pp. 487-500 (1986).

106. "A recursive quadratic programming algorithm that uses differentiable exact penalty functions" (with Y. Yuan), *Math. Programming,* Vol. 35, pp. 265–278 (1986).

107. "Radial basis functions for multivariable interpolation: a review", in *Algorithms for Approximation,* editors J.C. Mason and M.G. Cox, Oxford University Press (Oxford), pp. 143-167 (1987).

108. "The differential correction algorithm for generalized rational functions" (with E.W. Cheney), *Constructive Approximation,* Vol. 3, pp. 249-256 (1987).

109. "On error growth in the Bartels–Golub and Fletcher–Matthews algorithms for updating matrix factorizations," *Linear Algebra Applics.,* Vols 88–89, pp. 597–621 (1987).

110. "Updating conjugate directions by the BFGS formula", *Math. Programming,* Vol. 38, pp. 29–46 (1987).

111. *The State of the Art in Numerical Analysis* (editor with A. Iserles), Oxford University Press (Oxford) (1987).

112. "Methods for nonlinear constraints in optimization calculations", in *The State of the Art in Numerical Analysis,* editors A. Iserles and M.J.D. Powell, Oxford University Press (Oxford), pp. 325–357 (1987).

113. "A biographical memoir of Evelyn Martin Lansdowne Beale", *Biographical Memoirs of Fellows of the Royal Society,* Vol. 33, pp. 23–45 (1987).

114. "Radial basis function approximation to polynomials", in *Numerical Analysis 1987,* editors D.F. Griffiths and G.A. Watson, Longman Scientific and Technical (Burnt Mill), pp. 223–241 (1988).

115. *Mathematical Models and their Solutions: Contributions to the Martin Beale Memorial Symposium* (editor), Math. Programming B, Vol. 42, No. 1 (1988).

116. "An algorithm for maximizing entropy subject to simple bounds", *Math. Programming B,* Vol. 42, pp. 171–180 (1988).

117. "A review of algorithms for nonlinear equations and unconstrained optimization", in *ICIAM '87: Proceedings of the First International Conference on Industrial and Applied Mathematics,* editors J. McKenna and R. Temam, SIAM Publications (Philadelphia), pp. 220–232 (1988).

118. “QR factorization for linear least squares problems on the hypercube” (with R.M. Chamberlain), *IMA J. Numer. Anal.*, Vol. 8, pp. 401–413 (1988).

119. “TOLMIN: A Fortran package for linearly constrained optimization calculations”, Report No. DAMTP 1989/NA2, University of Cambridge (1989).

120. “On a matrix factorization for linearly constrained optimization problems”, in *Applications of Matrix Theory*, editors M.J.C. Gover and S. Barnett, Oxford University Press (Oxford), pp. 83–100 (1989).

121. “A tolerant algorithm for linearly constrained optimization calculations”, *Math. Programming B*, Vol. 45, pp. 547–566 (1989).

122. “Radial basis function interpolation on an infinite regular grid” (with M.D. Buhmann), in *Algorithms for Approximation II*, editors J.C. Mason and M.G. Cox, Chapman and Hall (London), pp. 146–169 (1990).

123. “Algorithms for linearly constrained optimization calculations”, in *Proceedings of the Third IMSL User Group Europe Conference*, editor M. Vaccari, IMSL (Houston), paper A-1, pp. 1–17 (1990).

124. “The updating of matrices of conjugate directions in optimization algorithms”, in *Numerical Analysis*, editors D.F. Griffiths and G.A. Watson, Longman Scientific and Technical (Burnt Mill), pp. 193–205 (1990).

125. “Karmarkar’s algorithm: a view from nonlinear programming”, *IMA Bulletin*, Vol. 26, pp. 165–181 (1990).

126. “Univariate multiquadric approximation: reproduction of linear polynomials”, in *Multivariate Approximation and Interpolation*, editors W. Haussmann and K. Jetter, Birkhäuser Verlag (Basel), pp. 227–240 (1990).

127. “A trust region algorithm for equality constrained optimization ” (with Y. Yuan), *Math. Programming A*, Vol. 49, pp. 189–211 (1991).

128. “Least squares smoothing of univariate data to achieve piecewise monotonicity” (with I.C. Demetriou), *IMA J. Numer. Anal.*, Vol. 11, pp. 411–432 (1991).

129. “The minimum sum of squares change to univariate data that gives convexity” (with I.C. Demetriou), *IMA J. Numer. Anal.*, Vol. 11, pp. 433–448 (1991).

130. “Univariate multiquadric interpolation: some recent results”, in *Curves and Surfaces*, editors P.J. Laurent, A. Le Méhauté and L.L. Schumaker, Academic Press (New York), pp. 371–382 (1991).

131. “A view of nonlinear optimization”, in *History of Mathematical Programming: A Collection of Personal Reminiscences*, editors J.K. Lenstra, A.H.G. Rinnooy Kan and A. Schrijver, North-Holland (Amsterdam), pp. 119–125 (1991).

132. "The theory of radial basis function approximation in 1990", in *Advances in Numerical Analysis II: Wavelets, Subdivision Algorithms and Radial Basis Functions*, editor W. Light, Oxford University Press (Oxford), pp. 105–210 (1992).

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

134. "Univariate multiquadric approximation: quasi-interpolation to scattered data" (with R.K. Beatson), *Constructive Approximation*, Vol. 8, pp. 275–288 (1992).

135. "The complexity of Karmarkar's algorithm for linear programming", in *Numerical Analysis, 1991*, editors D.F. Griffiths and G.A. Watson, Longman Scientific and Technical (Burnt Mill), pp. 142–163 (1992).

136. "Tabulation of thin plate splines on a very fine two-dimensional grid", in *Numerical Methods in Approximation Theory, Vol. 9*, editors D. Braess and L. Schumaker, Birkhäuser Verlag (Basel), pp. 221–244 (1992).

137. "On the number of iterations of Karmarkar's algorithm for linear programming", *Math. Programming*, Vol. 62, pp. 153–197 (1993).

138. "Truncated Laurent expansions for the fast evaluation of thin plate splines", *Numerical Algorithms*, Vol. 5, pp. 99–120 (1993).

139. "Log barrier methods for semi-infinite programming calculations", in *Advances on Computer Mathematics and its Applications*, editor E.A. Lipitakis, World Scientific (Singapore), pp. 1–21 (1993).

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

141. "A direct search optimization method that models the objective and constraint functions by linear interpolation", in *Advances in Optimization and Numerical Analysis*, editors S. Gomez and J-P. Hennart, Kluwer Academic (Dordrecht), pp. 51–67 (1994).

142. "Some algorithms for thin plate spline interpolation to functions of two variables", in *Advances in Computational Mathematics: New Delhi, India*, editors H.P. Dikshit and C.A. Micchelli, World Scientific (Singapore), pp. 303–319 (1994).

143. "The uniform convergence of thin plate spline interpolation in two dimensions", *Numerische Mathematik*, Vol. 68, pp. 107–128 (1994).

144. "A 'taut string algorithm' for straightening a piecewise linear path in two dimensions", Report No. DAMTP 1994/NA7, University of Cambridge.

145. "Some convergence properties of the modified log barrier method for linear programming", *SIAM J. Optimization*, Vol. 5, pp. 695–739 (1995).

146. “An algorithm that straightens and smooths piecewise linear curves in two dimensions”, in *Mathematical Methods for Curves and Surfaces*, editors M. Daehlen, T. Lyche and L.L Schumaker, Vanderbilt University Press (Nashville), pp. 439–453 (1995).

147. “A thin plate spline method for mapping curves into curves in two dimensions”, Report No. DAMTP 1995/NA2, University of Cambridge.

148. “On multigrid techniques for thin plate spline interpolation in two dimensions” (with R.K. Beatson and G. Goodsell), Report No. DAMTP 1995/NA8, University of Cambridge.

149. “A new iterative algorithm for thin plate spline interpolation in two dimensions”, Report No. DAMTP 1995/NA10, University of Cambridge.

150. “A review of algorithms for thin plate spline interpolation in two dimensions”, Report No. DAMTP 1996/NA4, University of Cambridge.

151. “Trust region methods that employ quadratic interpolation to the objective function”, 5th SIAM Conference on Optimization, Victoria, Canada.

Research Reports

No.	Authors	Title
96-16	M.D. Buhmann, R. Fletcher	M.J.D. Powell's work in univariate and multivariate approximation theory and his contribution to optimization
96-15	W. Gautschi, J. Waldvogel	Contour Plots of Analytic Functions
96-14	R. Resch, F. Stenger, J. Waldvogel	Functional Equations Related to the Iteration of Functions
96-13	H. Forrer	Second Order Accurate Boundary Treatment for Cartesian Grid Methods
96-12	K. Gerdes, C. Schwab	Hierarchic models of Helmholtz problems on thin domains
96-11	K. Gerdes	The conjugated vs. the unconjugated infinite element method for the Helmholtz equation in exterior domains
96-10	J. Waldvogel	Symplectic Integrators for Hill's Lunar Problem
96-09	A.-T. Morel, M. Fey, J. Maurer	Multidimensional High Order Method of Transport for the Shallow Water Equations
96-08	A.-T. Morel	Multidimensional Scheme for the Shallow Water Equations
96-07	M. Feistauer, C. Schwab	On coupled problems for viscous flow in exterior domains
96-06	J.M. Melenk	A note on robust exponential convergence of finite element methods for problems with boundary layers
96-05	R. Bodenmann, H.J. Schroll	Higher order discretisation of initial-boundary value problems for mixed systems
96-04	H. Forrer	Boundary Treatment for a Cartesian Grid Method
96-03	S. Hyvönen	Convergence of the Arnoldi Process when applied to the Picard-Lindelöf Iteration Operator
96-02	S.A. Sauter, C. Schwab	Quadrature for *hp*-Galerkin BEM in $\mathbb{R}^3$
96-01	J.M. Melenk, I. Babuška	The Partition of Unity Finite Element Method: Basic Theory and Applications
95-16	M.D. Buhmann, A. Pinkus	On a Recovery Problem
95-15	M. Fey	The Method of Transport for solving the Euler-equations
95-14	M. Fey	Decomposition of the multidimensional Euler equations into advection equations
95-13	M.D. Buhmann	Radial Functions on Compact Support