Available online at www.sciencedirect.com
ScienceDirect
Journal of Approximation Theory 238 (2019) 3–25
JOURNAL OF Approximation Theory
www.elsevier.com/locate/jat

Editorial

Michael J.D. Powell's work in approximation theory and optimisation

M.D. Buhmann

Mathematics Institute, Justus-Liebig University, 35392 Giessen, Germany

Received 31 May 2017; accepted 24 August 2017
Available online 6 October 2017

Communicated by Paul Nevai

Abstract

We summarise some of the substantial contributions of the late M.J.D. Powell to approximation theory and to optimisation, focusing specifically e.g. on uni- and multivariate splines (for instance Powell–Sabin split), including radial basis functions, but also on optimisation methods for minimising functions without constraints, with and without using first derivative approximations, such as the famous DFP formula.

1. Introduction

In this paper we shall outline the vast contributions of the late M.J.D. Powell (29.7.1936–19.4.2015) to approximation theory and optimisation. The approximation theory conference at Rauischholzhausen (Justus-Liebig University) in Germany – whose proceedings this volume of the journal is – was indeed dedicated to him and his work. Due to the subject area of this journal we shall write to a much smaller proportion about optimisation, although Mike Powell was of course a giant in that field too. For that reason and because Mike began his research work in optimisation, let us begin with a few remarks about optimisation and then add some more in the last section.

E-mail address: martin.buhmann@math.uni-giessen.de.

https://doi.org/10.1016/j.jat.2017.08.004

Mike's research in optimisation, including the development and practical implementation of methods, concerned the search for optima of functions of many variables which has important applications in practice, and science and engineering for instance. He always emphasised the universality of its applications ranging from optimising the routes for the Reading City rubbish collection to the Apollo programme, both of these extremely different applications he has worked on. Mike of course never stopped after proving convergence theorems for the optimisation algorithms for instance, but sought for most efficient ways of *implementing them,* and did just that himself afterwards. This then applied to approximation theory too. These codes were not only written most carefully and supported by profound theory but also made publicly available at no cost. He was always, not only in optimisation, successful in looking for efficient and highly accurate methods, including proofs of the efficiency and accuracy. This was the case whichever application in numerical analysis was asked for. He was also excellent in constructing numerical examples to identify particular deficiencies of inefficient algorithms which he then rectified.

Moreover, on top of his profound and celebrated contributions to optimisation, Mike has worked very influentially on approximation with uni- and multivariate polynomial splines, and ground-breaking on radial basis functions (convergence, fast algorithms for their computation or evaluation, Krylov-space methods, thin-plate splines, etc.), rational functions, ℓ_p-approximation, best Chebyshev approximations and convex or monotone approximations. In many contexts, his interests were again focused on convergence questions, efficient and accurate numerical computations and implementations, best choices of approximants, e.g., with respect to the uniform norm, knots and control points, reduction of number of data and norms of operators, adaptive methods, existence and characterisations of optimality and best approximations.

It should be particularly pointed out that Mike has written a first-class textbook on approximation theory (Approximation Theory and Methods, Cambridge University Press, [81]). His important book of 1983 is a central aspect to Mike's contribution to approximation theory up to that year.

It covers mostly univariate approximation theory, and within that characterisation theorems for best approximation, polynomial interpolation and approximation, rational approximation, splines, interpolation, error estimates and convergence results, ℓ_1-, ℓ_2- and Chebyshev approximations, Jackson's theorems, etc. It is written in Mike's typical most readable style, more explanations and less formulae to make for a particularly fluent and clear text. Speaking about clarity and style, many of us will always remember not only the quality of Mike's writing but

also his remarkable talks using coloured hand-written slides in his typical long-hand writing with carefully colour-coded texts and formulae.

The book includes new results (or new, streamlined proofs of previous results), very well thought-out problems, looking for and paying attention to details, computational considerations in Mike's precise and profound way. The author of this article uses it constantly in bachelor and master student education, both students and he himself finding it most reliable and useful. The work on ℓ_1-approximations – that is described in much detail in the volume – becomes, incidentally, again most interesting right now in connection for example with compression methods which currently receive much attention.

2. Univariate approximations

Some of the first contributions specifically to univariate approximation theory cover best uniform approximation and the so-called exchange algorithm. We begin by explaining some of Mike's results on best uniform approximation by polynomials, using the Remez algorithm. His paper with Curtis [13] studies the Remez algorithm for finite dimensional function spaces and its convergence properties.

Formulae are provided in [13] for the first and second partial derivatives of the error of polynomial approximation to sufficiently smooth approximands f, with respect to the reference points. So the error of approximation to an approximand f is studied, where the reference points are changed one-by-one. Indeed, if the error function's partial derivatives are zero, the stationary points are reached with second order convergence speed when the reference points are moved one at a time. There is a later article [20] with a result which estimates the error of approximation specifically by polynomials. It uses operator norm estimates to show that the Chebyshev norm of the interpolation error is always bounded above by a product of the least Chebyshev error obtainable and a factor that is independent of the approximand.

In fact, that paper is more general in that general finite dimensional approximation spaces are considered, and not only polynomial spaces. After the least maximum error and the maximum error of polynomial interpolation are considered in [20], a similar analysis is carried out for ℓ_2-approximations too: Optimal knot positions for polynomial spline least squares approximations are considered in [23].

A few years later, rational Chebyshev approximations are considered using the ordinary differential correction (ODC) algorithm. Using points $x_i \in X$, where the denominator of the rational approximant is positive, rational functions are computed to minimise the maximum error. Ian Barrodale, Frank Roberts and Mike studied in [41] rational Chebyshev approximations with numerator P of degree at most m and denominator Q of degree up to n for a given real finite point-set X. The maximum of the moduli of Q's coefficients is normalised to one and Q is positive on all of X. This approximation task ends up in a linear programming problem. It is, in fact, the original, not the more often used modified differential correction method, as this one surprisingly turns out to be better. Under the mentioned conditions, Barrodale, Roberts and Mike establish convergence, but while the sequence of approximation errors $\{\Delta_k\}_{k=1}^{\infty}$ is a convergent sequence with limit Δ^*, the numerators P_k and denominators Q_k may not converge, unless the following conditions are met: The cardinality of X must be at least $n + m + 1$, P^* and Q^* have no common zeros, and their leading – full degree – coefficients p_m^* and q_n^* do not vanish.

In that case these P_k and Q_k sequences converge to P^* and Q^*, respectively. The three authors then show at least second order convergence of the coefficient sequences of the polynomials.

The differential correction algorithm is further generalised by Ward Cheney and Mike in a paper about general rational approximations, i.e. fractions whose numerators and denominators

are no longer necessarily polynomials. The authors provide general convergence and order of convergence theorems, where it is shown that convergence may deteriorate from superlinear to linear in numerator and denominator if the best generalised rational approximation does not exist. The context is a set of continuous functions for numerator and denominator defined on a compact Hausdorff space. There must exist a positive function with norm one in that space (as denominator). Also, the distance between approximand and approximation space may in fact not be exactly attained. The authors prove in [108] that nonetheless as soon as the errors of the approximations converge to that distance, and if the spaces for numerator and denominator are finite dimensional with a unique best approximation, then the original differential correction algorithm gives superlinear convergence of the error functions.

Another use of rational approximations is studied by Mike and Arieh Iserles with respect to A-acceptable methods. There, a rational approximation of the same (polynomial) form is sought as the one mentioned earlier, but solely with the exponential function as an approximand. This approximation is called "maximal" if its error function has $m+n+1$ zeros with non-negative real part. If it is also equal (up to a nonzero constant) to $(\cdot)^{m+n+1-k}$ plus higher order terms, then we now define the set $E_{n,m,k}$ by such approximants. The two authors establish in [84] that a maximal approximation is A-acceptable if and only if it is either in $E_{n,n,k}$ for k at most two, or in $E_{n,n-2,0}$, or in $E_{n,n-1,k}$ for k at most one.

A further contribution to approximation theory for ℓ_p concerns the case $p = 1$ and the discrete test choice for ℓ_p-approximations from a space A of continuous approximants. Here, the conditions for the solution of the equivalent linear programming problem (we see again a nice link between approximation theory and optimisation) are expressed in terms of the data $f(x)$, $x \in X$, X a finite set, i.e., the error $\sum_{x\in X} |f(x) - \psi(x)|$ is to be minimised over all $\psi \in A$. Due to the finiteness of the problem we get a finite number of linear inequalities, very well suitable for a linear programming formulation, and expressible in terms of the data [71].

Mike wrote another article related to optimisation and approximation theory with Ya-Xiang Yuan in 1984 [98] that concerns ℓ_p-approximations for p either one or infinity. Here solutions are sought by iterative methods providing (specifically) superlinear convergence of overdetermined nonlinear functional equations

$$f(x) = 0, \quad x \in \mathcal{R}^n, \qquad f : \mathcal{R}^n \to \mathcal{R}^m.$$

Expressed differently, we wish to minimise the continuous, but not continuously differentiable expression

$$F(x) := \|f(x)\|_p, \qquad x \in \mathcal{R}^n.$$

Assuming f itself to be twice continuously differentiable, conditions on the estimates B_k to the Hessian matrix of f denoted by $\nabla^2 f$ are provided to obtain the desired superlinear convergence of descent methods so long as less than n equations are satisfied in a solution for $p = 1$ and at most n maximum residuals occur in a solution for $p = \infty$. The descent method requires at each step k the minimisation with respect to $s \in \mathcal{R}^n$ of the constrained function

$$\|f(x_k) + \nabla f(x_k)^T s\|_p + \frac{1}{2} s^T B_k s \approx F(x_k + s).$$

The demanded superlinear convergence phenomenon, with the mentioned minimum attained at $s = s_k$, means therefore nothing else than

$$\|x_k + s_k - x^*\| = o(\|x_k - x^*\|), \qquad k \to \infty.$$

The B_k may be indefinite (to allow for the so-called trust region methods to be applied, where even if the desired Hessian is positive definite, its approximations need not be at times because of the trust region limitations on the lengths of the search directions, and even then descent directions exist).

The trust region algorithms (see also our last section for optimisation methods) incorporate a step length restriction by radii and decide on increasing or decreasing the radius by comparing the actual reduction of the objective function with the predicted reduction. Using the ratio of these as a criterion for increasing or decreasing the radii – with the goal of achieving global convergence – is a method introduced by Mike. Moreover, a trust region scheme modifies the Hessian in the iteration step by adding a multiple of the identity matrix which gives a bias towards the steepest descent direction. We will mention more of Mike's vast contributions to the trust region algorithm in the final section about optimisation.

It is usually necessary to incorporate second derivative information into the calculation of the B_k for the cases $p = 1, \infty$, unless the solution is of strong uniqueness. It is interesting that the conditions on the second derivative approximations are expressed alternatively both in the Hessian matrices of the components of f at the solution and at the iteration points.

With Ioannis Demetriou in [128, 129], least-squares smoothing of univariate data was carried out to achieve piecewise monotonicity and to minimise the sum of squares of the required changes to those data providing convexity (precisely: to provide non-negative second divided differences). A dynamic programming procedure is given for minimising the global ℓ_2-error in $O(n^2 + kn \log n)$ operations for calculating with n data $k - 1$ (this number being prescribed) joins of individual monotonic sections. The goal is to make the least sum of squares change to the data so that the piecewise linear approximant is composed of those at most k sections. In theory, this combinatorial problem could have up to the order of n^k minima and the number of operations would greatly increase accordingly. Here, they achieve that there are only as few as stated required, and indeed only $O(n)$ suffice for k at most two. The principal advance in this work is that a recursive scheme has been found to reduce efficiently the number of data to be considered when finding the optimal knots (or "breakpoints") which are integer variables.

Further, an algorithm for quadratic programming is given for regaining convexity when the original function where the data stem from, is convex but (by errors, due to noise or rounding for example) the convexity is lost. Again, the changes to the data measured in the ℓ_2-norm are minimised while the second divided differences become non-negative, so that convexity is regained; an $O(n)$ operations starting procedure is required and a convergence proof of the algorithm is given. The method which is employed after generating the starting point is related to the algorithm of Goldfarb and Idnani that Mike researched into in his work on optimisation, see for instance [102].

Much of Mike's important work within approximation theory concerned splines in various ways, uni- and then mostly bivariate, radial function "splines" etc. In [27] for example, he considered norms of univariate polynomial spline operators and the deterioration of the Lagrange functions' localisation (that is, decay) along the real line when their polynomial degrees grow. In [29], again localisations of spline approximation operators were analysed, but now in the context of weighted least-squares approximations. In this case periodised and cubic splines were employed. The weight functions are of course always non-negative and h-periodic. The knots used are $0, h, 2h, \ldots, Mh$, M being a positive integer.

The fundamental functions are computed by a recursion for this approximation problem. The fastest obtainable decay rate of the associated fundamental functions was $(2-\sqrt{3})$ and its powers when the weights are concentrated on the knots. The worst case occurs when the weights are

concentrated in the *middle* between the knots. The spline can of course be expressed by a sum of truncated third powers plus a cubic polynomial. The optimal factor for this is provided in the paper for the case when the worst weights are chosen (as has been said, in the middle between the knots). It behaves like a power 0.3616^{ℓ} if its argument is ℓh, ℓ a positive integer.

The localisation can even be improved when the weighted least-squares norm of the error is augmented by the discrete ℓ_2-norm of the coefficients, again weighted suitably. Uniqueness is also guaranteed in these cases when the aforementioned damping term is added, but it may be lost when the weight function is no longer periodic. Finally, divided differences of the approximand are taken into account. If, for instance, the eighth derivative of the approximand is bounded, these differences are bounded above by a multiple of h^8 and if the approximand is a polynomial of degree at most seven, the error is zero except for dependencies at the ends of the interval of the approximation, when we work on a finite interval.

In [32], a weighted sum of squares of the discrete error function of an approximation plus a smoothing term is considered. It is to be minimised again by a cubic spline, but now with free knots. The augmenting smoothing term is a weighted sum of squares in the form of coefficients of the truncated cubic polynomials that appear in the cubic spline. They reflect the contribution of the third derivative discontinuities of the spline and the latter's smoothness. Amusingly, Mike makes a distinction between knots and "gnots" of the cubic splines, the latter being added whenever a trend turns up in the residual, i.e. the error function's values at the data points. The test for trends is applied locally between current gnots. The knots are added whenever needed for balancing the distribution of spline breakpoints against the distribution of gnots. He gives particular attention to the weights in the smoothing term which depends on knots' and gnots' distances, and on the weighted sum of squares of residuals.

Splines of degree n and using N knots are studied, now, for approximations with respect to standard norms without weights, in [23], where they should minimise the square-integral

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

for a given bounded square-integrable approximand f, defined over the interval $[a, b]$. It is necessary for a jth knot to be optimally placed that the integral

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

vanishes, where $\tilde{c}$ is the $(2n+1)$st degree spline with the given N knots in the open interval (a, b) and having $\hat{x}_j$ as *an extra knot.* That spline must satisfy the Hermite interpolation conditions at the boundaries, the N knots and $\hat{x}_j$. These conditions follow from the orthogonality of $\tilde{c}$'s $(n+1)$st derivative to all splines with knots $x_1 < \cdots < x_N$ and from explicit expressions for the n-fold integral of the error at $\hat{x}_j$. They are derived by integration by parts applied to the expression for orthogonality which result in point evaluations of other splines (those of higher piecewise polynomial degree). The error functional is minimised – here of course we see Mike's expertise in optimisation again – by a minimisation method from the article [8].

So-called optimal interpolation is the subject-matter of a paper [63], where a minimal $c(x)$ for all x is sought such that for an approximand f, a given non-negative integer k, and the approximand not being a polynomial of degree less than k,

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

s being required as a linear combination of the approximand's function values at the m data points of which we have at least k. Its coefficients will naturally depend on the argument. The

derivative of the suitably smooth approximand on the right-hand side is assumed to be non-zero without loss of generality. As a result, both s and c come out as splines of degree less than k with $m - k$ knots.

We now go into the direction of multivariable approximations. Indeed, in [68] a review of bivariate approximation tools is given, e.g. tensor-product methods derived from univariate schemes, considering interpolation and other linear conditions, i.e. least-squares, etc. Both equally-spaced and scattered data are addressed. Piecewise polynomials of degree one and higher are discussed, in connection with methods to generate Delaunay's triangulations. The discussion includes Shepard's method, moving least-squares and natural neighbourhood interpolants.

3. Multivariable approximation and radial basis functions

In a joint work [69] with Malcolm Sabin, Mike sought a globally continuous piecewise quadratic two-dimensional spline approximation, given function values of the approximand and first partial derivatives at each vertex of a provided triangulation of a domain. This ended up in the celebrated Powell–Sabin split, where each such triangle is the split into either six or twelve subtriangles with interior continuous differentiability conditions. The latter are exactly suitable (nine conditions, the right number of conditions for the degrees of freedom) for the mentioned six subtriangles, otherwise normals on edges are required which may be computed by linear interpolation. The edges of the interior triangles go from the midpoints of the "big" edges to the circumcentre of the triangle. The triangles are always acute. As a result the midpoint in the "big" triangle is the intersection of the normals at the midpoints of the big edges. Therefore this midpoint of the big triangle is in the plane spanned by the points between it and the edge's midpoint and that gives the required continuous differentiability. The Powell–Sabin split is used most frequently in finite elements and CAGD.

In an article in Comput. Optim. and Applications [154], again in the union of optimisation and approximation theory, Mike considered an optimal way of moving a sequence of points onto a curve in two-dimensional space. The idea is to identify the smoothest function that maps a two-dimensional picture for example on a square into another one preserving some fixed points. The goal of this is for instance to identify the same faces on several photographs taken at different times or, as Ian Barrodale, one of Mike's frequent co-authors, once put it to keep secrecy for a navy application, "long cigar-shaped objects in the sea, e.g. in a harbour". The next step is then to have fixed *curves* on top of fixed points as before. For the latter, curves in the original picture are mapped into sequences of points on the curves which in turn are made into continuously differentiable curves out of circular arcs and straight lines. The article [147] on "A thin-plate spline method for mapping curves into curves in two dimensions" is also here relevant.

Indeed, if $s : [0, T] \mapsto \mathcal{R}^2$ is the mentioned curve, and $x_1, \ldots, x_m$, are points in $\mathcal{R}^2$, the object is to find a monotonically increasing sequence of t_is, so as to minimise

$$\max_i \|x_i - s(t_i)\|_2,$$

where the smoothness of the sought s will depend on the t_is, the smoothness being desired to be as high as possible. This is an optimisation problem with the linear constraints that come from the monotonicity of the sequence of $t_i \in [0, T]$. Mike's presented method usually requires only $O(n)$ operations.

Our next subject-matter is the multivariable approximations using radial basis functions. Although having been researched into a lot by others, Mike was at that time one of the main persons in that area beginning with the review paper [107] that already contained new proofs

of important results. The paper addresses properties of the radial basis function interpolation matrices

$$\Phi = \{\phi(\|x_i - x_j\|_2)\}_{i,j=1}^m,$$

in particular nonsingularity of the interpolation problem at distinct points $x_i \in \mathcal{R}^n$, where for example the "radial basis function" ϕ could be the identity $\phi(r) = r$ (sometimes called the "linear radial basis function") or the multiquadric function with a real parameter c or the Gauß-kernel $\phi(r) = \exp(-c^2 r^2)$ positive real parameter c. All those matrices are indeed nonsingular if there are at a minimum two points used, therefore admitting unique existence of real coefficients λ_j such that the interpolant

$$s(x) = \sum_{j=1}^m \lambda_j \phi(\|x - x_j\|_2), \qquad x \in \mathcal{R}^n,$$

satisfies interpolation conditions $s(x_i) = b_i$ for any given right-hand sides and all i. Usage of the Euclidean norm here is relevant, as for instance in the Chebyshev- or ℓ_1-norm the interpolation matrices Φ could be singular for some of those simple kernel functions in spite of the points being distinct. It is interesting to note that one of the reasons why Mike was interested in the radial basis function (multivariate) interpolation or more generally approximation from the linear spaces spanned by the kernel functions $\phi(\|\cdot - x_j\|_2)$, $j = 1, \ldots, m$, was that they might be used for local approximation required in optimisation methods. Indeed, a Ph.D. student (Hans Martin Gutmann) much later-on addressed such questions. It is amusing that initially Mike found out that the radial basis function approximations are not so useful for the local approximations required in optimisation after all, but he then developed a great interest for their properties independently of that. Hans Martin's work started much later after Mike's initial interest in radial basis functions which began after a conversation with Carl de Boor in the mid-80s,

Motivated by the remarkable non-singularity results of Micchelli and others for radial basis function interpolation in n variables, Mike summarised and improved further early results in the paper [114]. This concerns the question – among others – whether polynomials are contained in the aforementioned linear spaces spanned by the (now infinitely many) shifts of the radial basis function kernels. In the article [122], the Lagrange functions for the interpolation problem with such gridded centres and the appropriate basis functions were computed for finite grids. Therefore a matrix is to be inverted to allow for the calculation of the coefficients of the cardinal functions. For that, a Gauß–Seidel type iteration was applied to a preconditioned matrix Ψ replacing Φ of the form

$$\Psi = \{\psi(x_i - x_j)\}_{i,j=1}^m,$$

the centres x_i coming from the cardinal grid and the ψs being linear combinations of the form

$$\psi(x) = \sum_{j=1}^m \mu_j \phi(\|x - x_j\|_2), \qquad x \in \mathcal{R}^n.$$

These functions decay for large argument rather than increase such as multiquadrics for instance as soon as their coefficients μ_j are chosen suitably. These linear combinations (of the form of componentwise divided differences in the simplest case, for example of second degree in one-dimensional multiquadrics) can decay in fact quite fast — of integer order; this decay property is linked to their ability to recover polynomials when infinite linear combinations of

the form of *quasi-interpolation* are taken

$$s(x) = \sum_j p(x_j)\psi(x - x_j), \qquad x \in \mathcal{R}^n.$$

There is even a link in this work with an optimisation program Mike wrote [119] because the sums of moduli of the off-diagonal elements of the mentioned matrix were minimised in order to improve convergence of the method (where Gauß–Seidel is applied) subject to a normalisation condition. Further there are coefficient constraints such as on the μ_j which give the needed algebraic decay of the ψs. The first step from gridded data to scattered data was due to Mike again [126] in which article a careful analysis of the multiquadric kernels' translates led to quasi-interpolants in one dimension when centres x_i are scattered. Central to this work is the description using the Peano-kernel theorem of the ψs, for multiquadrics for instance the Peano-kernel being the second derivative of the multiquadric function.

Still in one dimension, but with infinitely many centres diverging to $\pm\infty$, Mike asked in [126] whether it would be possible to have $s = f$, where s is of the form

$$s(x) = \sum_{j=-\infty}^{\infty} f(x_j)\psi_j(x)$$

with ψ_j defined below, when f is a polynomial of degree zero or one. The first point to notice is that the now infinite sum s of the $\lambda_j\phi(|\cdot - x_j|)$ is well-defined if and only if the series $\sum_j |\lambda_j x_j|$ converges.

Moreover, with

$$\psi_j \coloneqq \sum_i \mu_{ij}\phi(|\cdot - x_i|),$$

where we leave open over which indices the finite sum ranges, sums of the form $\sum_j \nu_j\psi_j$ are well-defined if and only if the function

$$\sum_{|i-j|\le m} |\nu_j| B_j\left((\cdot)^2 + c^2\right)^{-1.5}$$

is absolutely integrable, B_j being the linear B-spline centred at x_j. Then, for suitable coefficients μ_{ij}, indeed quasi-interpolation gives for all such approximands f that the quasi-interpolant agrees with the approximand identically.

Focusing further-on on interpolation with multiquadrics, Mike studied in [130] the uniform (in Chebyshev norm) convergence of approximant to approximand and its rate when m equally spaced centres are used solely on the univariate unit interval and the multiquadrics constant c is the spacing between the centres. This c is the reciprocal of $m - 1$. The convergence rate is at least one if the approximand is at least Lipschitz-continuous. As soon as it is also continuously differentiable with Lipschitz-continuous first derivative and satisfies certain boundary conditions on f and its first derivative at the extreme points of the interval, quadratic convergence is shown.

Mike and Rick Beatson extended this *Ansatz* in [134] by studying three different quasi-interpolants and scattered data. They are formed by building ψ_is from different finite-dimensional function spaces spanned not only by translates of the multiquadrics kernel. In the first case, constant functions are added to the other, in the third case linear functions, the second one being the plain vanilla shifts by the data points (centres) $x_1 < \cdots < x_m$, of multiquadrics radial function situation.

The first approximant may incidentally be written in the usual shape as a sum of

$$\lambda_j \phi(|\cdot - x_j|)$$

plus a constant — which is interesting because multiquadrics with constants added occur frequently in the analysis of radial basis function interpolants and their convergence properties. The second one by contrast comes only from a space spanned by multiquadrics (shifted) kernels.

The obtainable error estimates (upper bounds) are $(1 + c/h)\omega$, ω being the notation for the approximand's modulus of smoothness and h the maximal distance between two adjacent data points, or, involving f's boundary values $f(x_1)$, $f(x_m)$, and for a Lipschitz-continuous first derivative of the approximand This is the form that corresponds most 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 mentioned in the previous paragraph). If the added linear polynomial is zero, then one cannot obtain more than $O(h)$ unless the approximand satisfies certain boundary conditions. Concretely, the error is at least a constant multiple of $h\|\lambda\|_\infty$ even if the approximand is a nonzero linear polynomial. Superlinear convergence can be obtained if certain boundary conditions on the approximand and its first derivative are met, and if the approximand is twice differentiable. 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 up to zero), then superlinear convergence to twice continuously differentiable approximand f is obtained if and only if zero boundary conditions on the approximand's first derivative are met. If a linear polynomial is added to the sum of shifts of multiquadrics and the extra degrees of freedom are taken up by requiring that the coefficients of the multiquadric functions sum up to zero and the sum of coefficients times centres sum to zero then superlinear convergence to twice continuously differentiable approximands 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 zero and one (we are still using the unit interval for the approximand's argument), interpolating f at $1/2h$ and $1 - 1/2h$, and minimising the sum of squares of interpolation coefficients. The latter is the choice for which superlinear convergence is proved.

In [132], one of Mike's opera magna, he summarises 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 $\phi(r) = 1/\sqrt{r^2 + c^2}$, 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 summarise his and other authors' work, but offers simplifications, more clarity in the exposition and improvements of proofs and results.

A further nice connection between approximation and optimisation techniques can be found in [147] where approximants s in two dimensions and mapping to two dimensions as well are considered that are componentwise thin-plate splines. The goal is to find a mapping between two regions in the two-dimensional plane, where certain control points and control curves are mapped to prescribed positions. Mapping control points to points with the thin-plate splines method is not hard, but a curve must be discretised and it is not clear whether the discretisation is the same in the image region even though the curve retains its shape. Because thin-plate splines yield the interpolant of minimal second derivatives in the least-squares sense, there is already one optimising feature in that approach. In this article, Mike uses once more the universal algorithm from [119] to determine the optimal positions of the discrete points on the curve in the image. The idea is to minimise again the square of 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 on the discretised curve. Precisely, if f_i, g_i, $i = 1, 2, \ldots, m$, are the required image values, the square of the semi-norm of the thin-plate splines interpolant turns out to be a multiple of

$$(f_i)^T \Phi(f_i) + (g_i)^T \Phi(g_i),$$

where the thin-plate spline kernel is used and x_i are the points in the original domain in the two-dimensional plane. Φ is the interpolation matrix $\{\Phi(x_i - x_j)\}$ as before.

If we only want to map points into points from the plane, i.e. $(x_i)_i \mapsto (f_i, g_i)_i$ for all i, then the previous display is the minimal value of the square of the semi-norm that can be obtained. If, for simplicity, all of the data points originate from the discretisation of the curve in the original domain, that display can again be minimised with respect to the $(f_i, g_i)_i$, $i = 1, 2, \ldots, m$, 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 mentioned data points which gives linear inequality constraints to the optimisation 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 the two-dimensional plane, 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 the earlier results that required additional conditions on the domains where the approximand is defined. 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 spline 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 to be in the convex hull of centres (though it does need to be in their h-neighbourhood, due to the definition of h).

One of Mike's latest work on radial basis functions 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 many times, an application for that being 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 for the coefficients. And, of course, the conjugate gradient 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 that are shifted thin-plate spline kernels for data points far away from the argument x into one expression in order to minimise 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 shifted thin-plate spline kernels are lumped together. When this is done efficiently, however, its cost, 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 with at least $10m$ operations for direct evaluation. As soon as in this fashion thin-plate spline interpolants are tabulated on a two-dimensional grid with spacing h, the total cost is a small multiple of the number of grid-points plus an expression of order $m \log(1/h)/\epsilon^{1/3}$, where m is the number of centres and ϵ is the tolerance.

Another approach for computing thin-plate spline interpolants efficiently by Mike Powell and collaborators uses local Lagrange functions, by which we mean Lagrange functions L_j centred at x_j, say, that satisfy the Lagrange conditions only for several data points 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., they are very well localised. 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 the pointwise errors at the centres times the local Lagrange functions centred at x_j and s_k is here 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 one. In a later paper [148], a slightly different approach is used. It is based on the observation that a good approximation to $s(x)$ near a centre can be computed by only looking at f_j with x_j near that centre. This does not work in such an obvious fashion near the edges of a domain in two dimensions for example, but corrections may be applied by using subsets of centres. The ensuing error function is smooth. Now this can be carried out in iterations, where the coefficient of each shifted thin-plate spline kernel is approximated in each step of the iteration by a multiple of the leading coefficient of a local Lagrange function L_i (the multiplier being the pointwise error $f_i - s(x_i)$). Therefore the correcting term to a prior approximation is now a weighted linear combination of the shift of the kernel times the pointwise errors at the centres used for the local Lagrange function L_i. 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.

Under certain conditions Mike and Anita Faul were able in [159] to prove convergence of an iterative method using shifted thin-plate spline kernels in two dimensions. The number of data being m and s^* being the final, sought shifted thin-plate spline kernel interpolant to m data points, q approximately 30 being the number of local "surrounding" centres with respect to a given centre which are used to formulate a local Lagrange function at that centre, $m + 1 - q$

stages are used to form the approximant. At the kth stage, where k is at most $m - q$, q of the previous $m + 1 - k$ interpolation equations are used to have a new value for the kth coefficient, all earlier ones remaining intact. Let $\mathcal{L}_k \subset \{k, k+1, \ldots, m\}$ be the set of the interpolation conditions employed for a specific local cardinal function L_k. The set contains q elements. Then, if additionally all x_j are not collinear, where j is from $\mathcal{L}_k \setminus \{k\}$, then L_k will have non-zero native-space norm. Subsequently s is updated by adding a weighted sum of the pointwise residuals times this local Lagrange function. At the final stage, s is updated to $s + \sigma$ where σ interpolates the residual $f_i - s(x_i)$ for all $i > m - q$.

A question now other than convergence was then – typical for Mike's union of theoretical analysis and hard analysis about efficient implementation and computation – creating a good (Krylov subspace–) algorithm for computing multiquadrics interpolants efficiently in many dimensions. This again was work done with Anita Faul [159]. Here and in a related paper of Mike's with Anita and George Goodsell [168], they are talking about multiquadric interpolants s^* using $\phi(r) = \sqrt{r^2 + c^2}$ with m data in n dimensions by the usual sums of $\lambda_j \phi(\| \cdot - x_j \|_2)$ plus a constant. A direct calculation would of course require $O(m^3)$ operations for calculating the above coefficients through a linear system of equations. Now the method iterates over k, that is, s_k approximates s^* and the value of q is still about 30. For s_k, the sought coefficients are in the linear combination of shifted multiquadrics. We begin with the coefficients all zero and the added constant the average between the minimum and the maximum of the right-hand sides f_i. The errors are $r_i^k = f_i - s_k(x_i)$, which is $r_i^1 = f_i - \alpha_1$ at the start.

Now the coefficients for the Lagrange functions on $\mathcal{L}_\ell$ are computed.

The *search direction* d_k for the iterative process uses a linear combination of shifts of multiquadrics functions. The *step length* γ_k for the iterative process is a linear combination of the residuals. Then he takes $s_{k+1} = s_k + \gamma_k d_k +$ constant and adds $\gamma_k \delta_{ki}$ to λ_{ki} for forming the next one. In order to state a convergence result we let $\check{z}_\ell$ be the Lagrange functions as before for points with indices in $\mathcal{L}_\ell$, but now *without* the constants added. Let moreover Ξ be the operator defined by the projection of the space spanned by the $\check{z}_\ell$, $\ell = \Omega(1), \ldots, \Omega(m-1)$, with respect to the semi-inner product associated to the multiquadrics function (that possesses a semi-norm as the thin-plate splines above). Now suppose that γ_k is positive and that the first k search directions create the span of functions $\Xi^\ell s$ for ℓ at most k, these latter vectors being linear independent, and that the search directions d_ℓ and d_k are orthogonal and $s - s_k$ is to all $d_1, \ldots, d_k$ with respect to the aforementioned semi-inner product. Then termination of the iterative method is guaranteed.

4. Optimisation

In optimisation, Mike's contributions were – as is well known – gigantic and universal, it all beginning among other things on methods for minimisation without constraints with or without derivative information. For instance, an early contribution was to suggest extra line searches in [9] for optimisation schemes which do not require derivatives and would terminate for quadratic objective functions f, using directions that are conjugate to each other, but otherwise be less efficient for non-quadratic f. The iterative method increases the efficiency also in the non-quadratic case.

A superior method and a real breakthrough at the time was of course the DFP (Davidon–Fletcher–Powell) algorithm (a "variable metric method", and such methods had a huge impact on optimisation) that contains Mike's name and that now uses derivative information [8]. At the time, problems with twenty variables were already considered as difficult, but with the DFP method, a hundred variables and more were suddenly no longer impossible to handle.

The variable metric methods were considered in many of his papers such as [34], and an advantage of these are the excellent convergence properties directly enjoyed by the sequences x_j approximating a stationary point x^* as soon as they have entered into a suitable neighbourhood of x^*, so we have an asymptotic superlinear rate of convergence (that is, locally), with no restrictions on the approximation to the Hessian. Another updating method for the second derivative approximations is the BFGS algorithm; in the articles [38] and [62] we find convergence proofs of DFP and BFGS methods (with convex objective function and supposing exact line searches in the former, giving those assumptions up in the latter most important paper), respectively. It turns out in practice that BFGS is very often a little superior as far as speed of convergence is concerned, the question as to why having always been of great interest to Mike: indeed, highly exact worst case analyses were specialities of Mike's as for instance in [104] about the DFP method versus the BFGS (these two algorithms are carefully compared for quadratic objective functions).

In [93] Mike and Ge show convergence of the second derivative matrix approximations B_k for both DFP and BFGS when unit steplength in the search direction is fixed. If the objective function is strictly convex, then we have the said convergence. If the objective function is twice continuously differentiable with the Hessian being Lipschitz continuous and positive definite at the minimum, then DFP gives the same for starting points and starting Hessian approximations being already sufficiently accurate. If the objective function f is strictly convex and quadratic, then we have the convergence restricted to subsets

$$\lim_{k\to\infty} B_k d = \nabla^2 f d \qquad \forall\, d \in S,$$

where

$$S := \operatorname{span}\{[B_1^{-1}\nabla^2 f]^\ell (B_1)^{-1}\nabla f(x_1)\}_{\ell=0}^{\infty}$$

and with the further condition that $B_k u$ be independent of k for all vectors u that are orthogonal to $\nabla^2 f S$.

The next step is requiring no longer convexity of the function f to be minimise. Mike showed in a Mathematical Programming Paper [161] convergence of the DFP algorithm in two dimensions using DFP updates when the objective function f is twice differentiable and has bounded level sets

$$\{x \in \mathcal{R}^2 \mid f(x) \le f(x_1)\}.$$

It turns out that the line search conditions are essential for either convergence or for finding counter examples. In the former case Mike showed that if the steplength α_k is maximal so that $f(x_k + \alpha d_k)$ decreases monotonically in the search direction d_k for all $\alpha \le \alpha_k$, then, for any positive ϵ the gradient has length at most ϵ after only finitely many steps.

Conjugate gradient schemes which are also enormously important to the field are considered with respect to convergence in [64] and [99], for unconstrained optimisation, and in [110] using specifically BFGS conjugate direction updates. In [99], the Fletcher–Reeves formula for non-quadratic optimisation is used to derive a global convergence result, while in fact in [67], the Polak Ribière formula is shown to be better in practice.

Further, Mike analysed schemes for unconstrained minimisation *without using derivative information* and wrote a review of algorithms that do not require first derivatives in [171]. In this context, it is interesting that these methods were of high interest to him early in his career and belonged to his last interests as well.

Constrained optimisation problems are of course studied almost everywhere, for instance in the context of optimising by augmented Lagrangian penalty functions in the fundamental paper [30]. There, not only large multiples of squares of the equality (equal to zero) constraints are added to the objective function – and then an algorithm for unconstrained minimisation is applied – but constant shifts of constraint functions are squared and multiplied with small coefficients instead. Mike's method adjusts multipliers and shifts, the algorithm therefore being equivalent to correcting Lagrange coefficient estimates in the augmented Lagrange function method for constrained optimisation. One of the most important points of the work in [30] is to show that the multiplier in front of the square of the Euclidean norm of the zero constraint functions need not diverge to infinity but may stay bounded once it is sufficiently large and we still get convergence of the stationary points of the augmented Lagrange functions to the x^* sought. This has the particular advantage of improving the conditioning of previous approaches such as quadratic penalty methods.

SQP methods are studied in [72, 73], where quadratic approximations are made to the objective function and linear approximations to the constraints, giving a quadratic programming problem. No second derivative information are required of the objective function but the BFGS updates or DFP are used (of which a particular form is proposed in [72]).

The Maratos effect and a watchdog method for avoiding it are analysed in [87]; Mike used the latter to make the SQP method more efficient. The problem is here that we have to decide whether a change to the current approximation to the minimum in a constrained minimisation problem is acceptable even if it does not reduce both the objective function and the constraint violations at the same time. The Maratos effect occurs when these are both worse although the update is in fact a good one. The watchdog technique is to allow such changes to the points sometimes while possibly rejecting them later on.

Furthermore, a most successful software package TOLMIN for general linearly constrained minimisation is described and analysed in [121, 145]. Here the original problem is that many constraints may greatly reduce the efficiency of an algorithm by producing many small steps when we are working near the boundaries of the feasible set, so long as all constraints are always met. The new TOLMIN method avoids this bad behaviour by keeping points away from those boundaries of constraints where the current step is, so that line searches are only restricted by constraint boundaries with residuals exceeding certain tolerance parameters. The latter are also automatically reduced as the steps are calculated.

A typical form of Mike's most careful analyses of difficult effects in numerical schemes and immediate, carefully worked out suggestions for improvement is also to be found in [125] where he studied Karmarkar's method and gave examples of its limitations. A remarkable convergence proof of the shifted log-barrier method is offered in [145].

A further breakthrough among many is the trust region method [37] where also the Powell symmetric Broyden updating formula is used; later on this is nicely summarised and improved in the paper [154]. In the same year there is an analysis of using band matrices for the approximations of second derivative matrices in trust region algorithms. Mike's precisely analysed improvements of nonlinear minimisation schemes are found for instance in his dogleg method in [34] where steps along the steepest descent direction and towards Newton points are used alternately. In [155], band matrices of band-width s, say, are used for updating the second derivative approximations. Therefore, only $O(sn^2)$ instead of $O(n^3)$ operations are required when a new search direction is computed. The total work turns out to be $O(n^{7/3})$ operations because the updates are made by matrices of rank less than three.

These schemes developed further because interest became generally less focused on quadratic termination criteria but on global convergence behaviour of the minimisation software. In those

papers, trust region methods were developed. Furthermore, the Powell symmetric Broyden updating formula for the second derivative approximations was created. In [174, 176], papers were published about convergence proofs of trust region algorithms for unconstrained minimisation, trust region algorithms for unconstrained minimisation without using derivatives (this uses linear or quadratic models without derivative information and trust region with recipes for updating radius, interpolation points – that are required to keep nonsingularity and a good model intact –, and possibly using fewer interpolation points that minimise the Frobenius norm of the change in the Hessian of the model: the Broyden update). This also includes a review of the state-of-the-art in 2008 of nonlinear optimisation methods since 1959 [175].

Further articles on unconstrained minimisation using quadratic approximations, but no derivative information were in [169, 171] and modelled the first and second derivatives by interpolation where from time to time during the algorithm extra values were used to avoid degeneracy and again we minimise the Frobenius norm of the change in the Hessian of the model. The general algorithm is again a trust region method. The latter question of least Frobenius norm updates (symmetric Broyden method) to quadratic models was specifically addressed again in [166, 167], where in n dimensions and using m interpolation conditions $(m + n + 1)^2$ linear systems of equations are to be solved. Further, a new program to update the Lagrange function's coefficients using $O(m + n)^2$ operations is provided; from time to time extra values are added for example to keep regularity of the interpolation problem. Specifically in [166] the trust region radius and which interpolation points to employ for a good model and regularity, always with a symmetric Broyden update, are given. In [163], questions about efficient updates when an interpolation point is moved and how to avoid degeneracy are answered.

And as late as in 2014 in [179] fast trust region methods with quadratic models were analysed for linearly constrained optimisation. This is related to [165], a paper he wrote on controlling the change of trust region radii by error bounds on models for the optimisation problems; these are quadratic models defined by Lagrange functions (therefore by interpolation), and Mike asked the question whether a change to an interpolation point preserves nonsingularity of the problem, the next question then to be addressed being how to make an efficient update once a point is moved. This fashion of controlling the radii gives excellent convergence properties to the algorithm, and one can conveniently construct initial points beforehand.

The new program UOBYQA using trust region methods for unconstrained optimisation is described in a paper in Mathematical Programming [164], and the algorithm requires no (first) derivatives anymore. The derivatives are not modelled by divided differences (which is unsuitable for noisy data). It models the curvature of the objective function by a quadratic function and interpolation at $(n + 1)(n + 2)/2$ points in n dimensions. At each step, *either* a trust region step is carried out *or* a procedure to improve the said model by replacing a point (and deciding which one). Each new point replaces one single interpolation point, where it has of course to be decided which point is to be replaced and how to update the coefficients of the Lagrange function. The latter provide incidentally an excellent estimate of the current accuracy of the model and information on how to improve it. The interpolation points cannot be too close due to lower bounds on the trust region radii.

He discussed the NEWUOA program too, which was introduced in 2004–2006 [169, 173]. There, m values are used instead of the otherwise $(n + 1)(n + 2)/2$ needed ones in n dimensions to interpolate the objective functions by a quadratic polynomial in n variables, $m \approx 2n + 1$ being a typical value, in other words much less information is needed for the second derivative matrix' approximation. At the beginning, the second derivative is approximated even by a diagonal matrix. At each iteration a single interpolation point only is updated. The remaining

$(n+1)(n+2)/2-m$ degrees of freedom are used up by minimising the Frobenius norm of the alteration to the Hessian, so the symmetric Broyden method is used where also the approximation to the Hessian is kept a symmetric matrix. An earlier paper introduces a precise approximation of the objective function without using derivatives [151].

The algorithms perform very well even for large dimensions n up to, say 100. For such dimensions, Mike's codes perform excellently and they are state-of-the-art of such programmes.

After NEWUOA Mike developed BOBYQA which extends unconstrained derivative free optimisation to so-called box-constraints and then LINCOA which uses constrained derivative free optimisation with linear constraints. Mike's codes are in fact available for the future under https://ccpforge.cse.rl.ac.uk/gf/project/powell/wiki/.

Acknowledgments

The author thanks Naomi Buhmann (who helped with the list of publications), Carl de Boor, and Coralia Cartis for their help with this article. The comments of two anonymous referees also improved the paper; the portrait of Mike Powell in this journal's issue has been taken by Naomi Buhmann who gave permission for publication.

List of Publications

[1] Ground-state splitting for d^{56} ions in a cubic field, (with J.R. Gabriel and D.F. Johnston), Phys. Rev. Letters 5, 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^{++} in a cubic field, (with J.R. Gabriel and D.F. Johnston), Proc. Roy. Soc. A 264, 503–515 (1961).
[4] The crystalline field parameters for dysprosium ethyl sulphate, (with R. Orbach), Proc. Phys. Soc. 78, 753–758 (1961).
[5] An iterative method for finding stationary values of a function of several variables, Computer Journal 5, 147–151 (1962).
[6] Roothaan's procedure for solving the Hartree–Fock equation, in Numerical Solution of Ordinary and Partial Differential Equations (L. Fox, ed.), Pergamon Press (Oxford), 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 272, 371–386 (1963).
[8] A rapidly convergent descent method for minimization, (with R. Fletcher), Computer Journal 6, 163–168 (1963).
[9] An efficient method for finding the minimum of a function of several variables without calculating derivatives, Computer Journal 7, 155–162 (1964).
[10] The volume internal to three intersecting hard spheres, Molecular Physics 7, 591–592 (1964).
[11] A method for minimizing a sum of squares of nonlinear functions without calculating derivatives, Computer Journal 7, 303–307 (1965).
[12] Necessary conditions for a minimax approximation, (with A.R. Curtis), Computer Journal 8, 358–361 (1966).
[13] On the convergence of exchange algorithms for calculating minimax approximations, (with A.R. Curtis), Computer Journal 9, 78–80 (1966).
[14] The general theory of linear approximation, (with D.C. Handscomb and D.F. Mayers).
[15] The exchange algorithm on a discrete point set.
[16] Theory of general nonlinear minimax approximation, and.
[17] An introduction to epsilon entropy, in Methods of Numerical Approximation (D.C. Handscomb, ed.), Pergamon Press (Oxford), 61–71, 73–81, 155–162 and 183–190 (1966).
[18] Minimization of functions of several variables, in Numerical Analysis: An Introduction (J. Walsh, ed.), Academic Press (London), 143–157 (1966).
[19] Weighted uniform sampling –a Monte Carlo technique for reducing variance, (with J. Swann), J. Inst. Maths. Applics. 2, 228–236 (1966).
[20] On the maximum errors of polynomial approximations defined by interpolation and by least squares criteria, Computer Journal 9, 404–407 (1967).

[21] Error analysis of the equal-interval interpolation by cubic splines, (with A.R. Curtis), AERE Report 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 spline approximations, in Numerische Mathematik, Differentialgleichungen, Approximationstheorie Birkhäuser Verlag (Basel), 317–339 (1968).
[24] On the calculation of orthogonal vectors, Computer Journal 11, 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 12, 288–290 (1969).
[27] A comparison of spline approximations with classical interpolation methods, in Proceedings of the IFIP 1968 Congress, North-Holland (Amsterdam), 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), 122–126 (1969).
[29] The local dependence of least squares cubic splines, SIAM J. Numer. Anal. 6, 398–413 (1969).
[30] A method for nonlinear constraints in minimization problems, in Optimization (R. Fletcher, ed.), Academic Press (London), 283–297 (1969).
[31] A survey of numerical methods for unconstrained optimization, SIAM Review 12, 79–97 (1970).
[32] Curve fitting by splines in one variable, in Numerical Approximation to Functions and Data (J.G. Hayes, ed.), Athlone Press (London), 65–83 (1970).
[33] Rank one methods for unconstrained optimization, in Integer and Nonlinear Programming (J. Abadie, ed.), North-Holland (Amsterdam), 139–156 (1970).
[34] A hybrid method for nonlinear equations, in Numerical Methods for Nonlinear Equations (P. Rabinowitz, ed.), Gordon and Breach (London), 87–114 (1970).
[35] A Fortran subroutine for solving systems of nonlinear algebraic equations, in Numerical methods for Nonlinear Equations (P. Rabinowitz, ed.), Gordon and Breach (London), 115–161 (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 (J.B. Rosen, O.L. Mangasarian and K. Ritter, eds), Academic Press (New York), 31–65 (1970).
[38] On the convergence of the variable metric algorithm, J. Inst. Maths. Applics. 7, 21–36 (1971).
[39] Recent advances in unconstrained optimization, Math. Programming 1, 26–57 (1971).
[40] Some properties of the variable metric algorithm, in Numerical Methods for Nonlinear Optimization (F.A. Lootsma, ed.), Academic Press (London), 1–17 (1972).
[41] The differential correction algorithm for rational approximation, (with I. Barrodale and F.D.K. Roberts), SIAM J. Numer. Anal. 9, 493–504 (1972).
[42] Quadratic termination properties of minimization algorithms: 1. Statement and discussion of results, J. Inst. Maths. Applics. 10, 333–342 (1972).
[43] Quadratic termination properties of minimization algorithms: 2. Proofs of theorems, J. Inst. Maths. Applics. 10, 343–357 (1972).
[44] Problems related to unconstrained optimization, in Numerical Methods for Unconstrained Optimization (W. Murray, ed.), Academic Press (London), 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. 7308, Harwell Laboratory (1972).
[47] On search directions for minimization algorithms, Math. Programming 4, 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 9, 326–327 (1973).
[50] Unconstrained minimization algorithms without computation of derivatives, Boll. della Unione Matematica Italiana 9, 60–69 (1974).
[51] On the estimation of sparse Jacobian matrices, (with A.R. Curtis and J.K. Reid), J. Inst. Maths. Applics. 13, 117–119 (1974).
[52] Unconstrained minimization and extensions for constraints, in Mathematical Programming in Theory and Practice (P.L. Hammer and G. Zoutendijk, eds), North–Holland (Amsterdam), 31–79 (1974).

[53] Piecewise quadratic surface fitting for contour plotting, in Software for Numerical Mathematics (D.J. Evans, ed.), Academic Press (London), 253–271 (1974).
[54] On the modification of factorizations, (with R. Fletcher), Maths. of Comp. 28, 1067–1087 (1974).
[55] Introduction to constrained optimization, in Numerical Methods for Constrained Optimization (P.E. Gill and W. Murray, eds), Academic Press (London), 1–28 (1974).
[56] On the convergence of cyclic Jacobi methods, (with K.W. Brodlie), J. Inst. Maths. Applics. 15, 279–287 (1975).
[57] Convergence properties of a class of minimization algorithms, in Nonlinear Programming 2 (O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds), Academic Press (New York), 1–27 (1975).
[58] A view of minimization algorithms that do not require derivatives, ACM Trans. Math. Software 1, 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), 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 (L.C.W. Dixon, ed.), Academic Press (London), 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 (R.W. Cottle and C.E. Lemke, eds), American Mathematical Society (Providence), 53–72 (1976).
[63] Optimal interpolation, (with P.W. Gaffney), in Numerical Analysis Dundee 1975, Lecture Notes in Mathematics No. 506 (G.A. Watson, ed.), Springer-Verlag (Berlin), 90–99 (1976).
[64] Some convergence properties of the conjugate gradient method, Math. Programming 11, 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 11, 141–147 (1977).
[67] Restart procedures for the conjugate gradient method, Math. Programming 12, 241–254 (1977).
[68] Numerical methods for fitting functions of two variables, in The State of the Art in Numerical Analysis (D.A.H. Jacobs, ed.), Academic Press (London), 563–604 (1977).
[69] Piecewise quadratic approximations in triangles, (with M.A. Sabin), ACM Trans. Math. Software 3, 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 (J.R. Rice, ed.), Academic Press (New York), 15–33 (1977).
[71] Algorithms for nonlinear constraints that use Lagrangian functions, Math. Programming 14, 224–248 (1978).
[72] A fast algorithm for nonlinearly constrained optimization calculations, in Numerical Analysis Dundee 1977, Lecture Notes in Mathematics No. 630 (G.A. Watson, ed.), Springer–Verlag (Berlin), 144–157 (1978).
[73] The convergence of variable metric methods for nonlinearly constrained optimization calculations, in Nonlinear Programming 3 (O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds), Academic Press (New York), 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 (R. Glowinski and J.L. Lions, eds), Springer-Verlag (Berlin), 62–72 (1979).
[75] On the estimation of sparse Hessian matrices, (with Ph.L. Toint), SIAM J. Numer. Anal. 16, 1060–1074 (1979).
[76] Gradient conditions and Lagrange multipliers in nonlinear programming, and.
[77] Variable metric methods for constrained optimization, in Nonlinear Optimization Theory and Algorithms (L.C.W. Dixon, ed.), E. Spedicato and G. Szegö, Birkhäuser (Boston), 201–220 and 279–294 (1980).
[78] Optimization algorithms in 1979, in Optimization Techniques, Lecture Notes in Control and Information Sciences No. 22, (K. Iracki, K. Malanowski and S. Walukiewicz, eds), Springer-Verlag (Berlin), 83–98 (1980).
[79] A discrete characterization theorem for the discrete linear approximation problem, (with F.D.K. Roberts), J. Approx. Theory 30, 173–179 (1980).
[80] A note on quasi-Newton formulae for sparse second derivative matrices, Math. Programming 20, 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 20, 353–357 (1981).
[83] On the decomposition of conditionally positive-semidefinite matrices, (with D.H. Martin and D.H. Jacobson), Linear Algebra Applics 39, 51–59 (1981).
[84] On the A-acceptability of rational approximations to the exponential function, (with A. Iserles), IMA J. Numer. Anal. 1, 241–251 (1981).

[85] An upper triangular matrix method for quadratic programming, in Nonlinear Programming (O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds), Academic Press (New York), 1–24 (1981).
[86] The Shanno-Toint procedure for updating sparse symmetric matrices, (with Ph.L. Toint), IMA J. Numer. Anal. 1, 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 16, 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 (G.A. Watson, ed.), Springer-Verlag (Berlin), 26–37 (1982).
[90] Extensions to subroutine VFO2AD, in System Modeling and Optimization, Lecture Notes in Control and Information Sciences No. 38 (R.F. Drenick and F. Kozin, eds), Springer-Verlag (Berlin), 529–538 (1982).
[91] Algorithms for constrained and unconstrained optimization calculations, in Current Developments in the Interface: Economics, Econometrics, Mathematics (M. Hazewinkel and H.G. Rinnooy Kan, eds), Reidel (Dordrecht), 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 27, 123–143 (1983).
[94] Variable metric methods for constrained optimization, in Mathematical Programming: The State of the Art, Bonn 1982 (A. Bachem, M. Grötschel and B. Korte, eds), Springer-Verlag (Berlin), 288–311 (1983).
[95] General algorithms for discrete nonlinear approximation calculations, in Approximation Theory IV (C.K. Chui, L.L. Schumaker and J.D. Ward, eds), Academic Press (New York), 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 29, 297–303 (1984).
[98] Conditions for superlinear convergence in and solutions of overdetermined nonlinear equations, (with Y. Yuan), IMA J. Numer. Anal. 4, 241–251 (1984).
[99] Nonconvex minimization calculations and the conjugate gradient method, in Numerical Analysis Proceedings, Dundee 1983, Lecture Notes in Mathematics No. 1066 (G.A. Watson, ed.), Springer-Verlag (Berlin), 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 (Z. Ciesielski and C. Olech, eds), North-Holland (Amsterdam), 1525–1539 (1984).
[101] An application of Gaussian elimination to interpolation by generalised rational functions, (with T. Havie), in Rational Approximation and Interpolation, Lecture Notes in Mathematics No. 1105 (P.R. Graves-Morris, E.B. Saff and R.S. Varga, eds), Springer- Verlag (Berlin), 442–452 (1984).
[102] On the quadratic programming algorithm of Goldfarb and Idnani, Math. Programming Studies 25, 46–61 (1985).
[103] The performance of two subroutines for constrained optimization on some difficult test problems, in Numerical Optimization 1984 (P.T. Boggs, R.H. Byrd and R.B. Schnabel, eds), SIAM Publications (Philadelphia), 160–177 (1985).
[104] How bad are the BFGS and DFP methods when the objective function is quadratic? Math. Programming 34, 34–47 (1986).
[105] Convergence properties of algorithms for nonlinear optimization, SIAM Review 28, 487–500 (1986).
[106] A recursive quadratic programming algorithm that uses differentiable exact penalty functions, (with Y. Yuan), Math. Programming 35, 265–278 (1986).
[107] Radial basis functions for multivariable interpolation: a review, in Algorithms for Approximation (J.C. Mason and M.G. Cox, eds), Oxford University Press (Oxford), 143–167 (1987).
[108] The differential correction algorithm for generalized rational functions, (with E.W. Cheney), Constructive Approximation 3, 249–256 (1987).
[109] On error growth in the Bartels-Golub and Fletcher-Matthews algorithms for updating matrix factorizations, Linear Algebra Applics 88–89, 597–621 (1987).
[110] Updating conjugate directions by the BFGS formula, Math. Programming 38, 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 (A. Iserles and M.J.D. Powell, eds), Oxford University Press (Oxford), 325–357 (1987).

[113] A biographical memoir of Evelyn Martin Lansdowne Beale, Biographical Memoirs of Fellows of the Royal Society 33, 23–45 (1987).
[114] Radial basis function approximation to polynomials, in Numerical Analysis 1987 (D.F. Griffiths and G.A. Watson, eds), Longman Scientific and Technical (Burnt Mill), 223–241 (1988).
[115] Mathematical Models and their Solutions: Contributions to the Martin Beale Memorial Symposium (editor), Math. Programming B 42(1) (1988).
[116] An algorithm for maximizing entropy subject to simple bounds, Math. Programming B 42, 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 (J. McKenna and R. Temam, eds), SIAM Publications (Philadelphia), 220–232 (1988).
[118] QR factorization for linear least squares problems on the hypercube, (with R.M. Chamberlain), IMA J. Numer. Anal. 8, 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 (M.J.C. Gover and S. Barnett, eds), Oxford University Press (Oxford), 83–100 (1989).
[121] A tolerant algorithm for linearly constrained optimization calculations, Math. Programming B 45, 547–566 (1989).
[122] Radial basis function interpolation on an infinite regular grid, (with M.D. Buhmann), in Algorithms for Approximation II (J.C. Mason and M.G. Cox, eds), Chapman and Hall (London), 146–169 (1990).
[123] Algorithms for linearly constrained optimization calculations, in Proceedings of the Third IMSL User Group Europe Conference (M. Vaccari,ed.), IMSL (Houston), paper A-1, 1–17 (1990).
[124] The updating of matrices of conjugate directions in optimization algorithm, in Numerical Analysis 1989 (D.F. Griffiths and G.A. Watson, eds), Longman Scientific and Technical (Burnt Mill), 193–205 (1990).
[125] Karmarkar's algorithm: a view from nonlinear programming, IMA Bulletin 26, 165–181 (1990).
[126] Univariate multiquadric approximation: reproduction of linear polynomials, in Multivariate Approximation and Interpolation (W. Haussmann and K. Jetter, eds), Birkhäuser Verlag (Basel), 227–240 (1990).
[127] A trust region algorithm for equality constrained optimization, (with Y. Yuan), Math. Programming A 49, 189–211 (1991).
[128] Least squares smoothing of univariate data to achieve piecewise monotonicity, (with I.C. Demetriou), IMA J. Numer. Anal. 11, 411–432 (1991).
[129] The minimum sum of squares change to univariate data that gives convexity, (with I.C. Demetriou), IMA J. Numer. Anal. 11, 433–448 (1991).
[130] Univariate multiquadric interpolation: some recent results, in Curves and Surfaces (P.J. Laurent, A. Le Méhauté and L.L. Schumaker, eds), Academic Press (New York), 371–382 (1991).
[131] A view of nonlinear optimization, in History of Mathematical Programming: A Collection of Personal Reminiscences (J.K. Lenstra, A.H.G. Rinnooy Kan and A. Schrijver, eds), North-Holland (Amsterdam), 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 (W. Light, ed.), Oxford University Press (Oxford), 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. 12, 107–133 (1992).
[134] Univariate multiquadric approximation: quasi-interpolation to scattered data, (with R.K. Beatson), Constructive Approximation 8, 275–288 (1992).
[135] The complexity of Karmarkar's algorithm for linear programming, in Numerical Analysis, 1991 (D.F. Griffiths and G.A. Watson, eds), Longman Scientific and Technical (Burnt Mill), 142–163 (1992).
[136] Tabulation of thin plate splines on a very fine two–dimensional grid, in Numerical Methods in Approximation Theory 9 (D. Braess and L. Schumaker, eds), Birkhäuser Verlag (Basel), 221–244 (1992).
[137] On the number of iterations of Karmarkar's algorithm for linear programming, Math. Programming 62, 153–197 (1993).
[138] Truncated Laurent expansions for the last evaluation of thin plate splines, Numerical Algorithms 5, 99–120 (1993).
[139] Log barrier methods for semi-infinite programming calculations, in Advances on Computer Mathematics and its Applications (E.A. Lipitakis, ed.), World Scientific (Singapore), 1–21 (1993).
[140] An iterative method for thin plate splines interpolation that employs approximations to Lagrange functions, (with R.K. Beatson), in Numerical Analysis, 1993 (D.F. Griffiths and G.A. Watson, eds), Longman Scientific and Technical (Burnt Mill), 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 (S. Gomez and J.-P. Hennart, eds), Kluwer Academic (Dordrecht), 51–67 (1994).
[142] Some algorithms for thin plate spline interpolation to functions of two variables, in Advanced in Computational Mathematics: New Delhi, India (H.P. Dikshit and C.A. Micchelli, eds), World Scientific (Singapore), 303–319 (1994).
[143] The uniform convergence of thin plate spline interpolation in two dimensions, Numerische Mathematik 68, 107–128 (1994).
[144] A 'taut string algorithm' for straightening a piecewise linearpath in two dimensions, Report No. DAMTP 1994/NA7, University ofCambridge, IMA J. Num. Anal. 18, 1–35 (1998).
[145] Some convergence properties of the modified log barrier method for linear programming, SIAM J. Optimization 5, 695–739 (1995).
[146] An algorithm that straightens and smoothes piecewise linear curves in two dimensions, in Mathematical Methods for Curves and Surfaces (M. Daehlen, T. Lyche and L.L. Schumaker, eds), Vanderbilt University Press (Nashville), 439–453 (1995).
[147] A thin plate spline method for mapping curves into curves in two dimensions, in Computational Techniques and Applications: CTAC95 (R.L. May and A.K. Easton, eds), World Scientific (Singapore), 43–57 (1996).
[148] On multigrid techniques for thin plate spline interpolation in two dimensions, (with R.K. Beatson and G. Goodsell), in The Mathematics of Numerical Analysis (J. Renegar, M. Shub and S. Smale, eds), American Mathematical Society (Providence), 77–97 (1996).
[149] A review of algorithms for thin plate spline interpolation in two dimensions, in Advanced Topics in Multivariate Approximation (F. Fontanella, K. Jetter and P.J. Laurent, eds), World Scientific (Singapore), 303–322 (1996).
[150] A review of methods for multivariable interpolation at scattered data points, Report No. DAMTP 1996/NA11, University of Cambridge.
[151] Least squares fitting to univariate data subject torestrictions on the signs of the second differences, (withI.C. Demetriou), Approximation Theory and Optimization: Tributes to M.J.D. Powell, M.D. Buhmann and A. Iserles, eds, Cambridge UniversityPress (Cambridge), 1997, 109–132.
[152] A new iterative algorithm for thin plate spline interpolation in two dimensions, Annals of Numerical Mathematics 4, 519–527 (1997).
[153] A review of methods for multivariable interpolation at scattered data points, in The state of the art in Numerical Analysis (I.S. Duff and G.A. Watson, eds), Oxford University Press (New York), 283–309 (1997).
[154] Trust region calculations revisited, in Numerical analysis 1997 (Dundee), Longman (Harlow), 193–211 (1998).
[155] The use of band matrices for second derivative approximations in trust region algorithms, in Advances in nonlinear programming, Kluwer Academic Publ. (Dordrecht), 3–28 (1998).
[156] Direct search algorithms for optimization calculations, in Acta Numerica 7, Cambridge University Press (Cambridge), 287–336 (1998).
[157] An optimal way of moving a sequence of points onto a curve in two dimensions, in Computational optimization –a tribute to Olvi Mangasarian, Part II, Computational Optim. Appl. 13, 163–185 (1999).
[158] Recent research at Cambridge on radial basis functions, in New developments in approximation theory, Birkhäuser Verlag (Basel), 215–232 (1999).
[159] Proof of convergence of an iterative technique for thin plate spline interpolation in two dimensions. Radial basis functions and their applications, (with A.C. Faul), in Adv. Comput. Math. 11, 183–192 (1999).
[160] Krylov subspace methods for radial basis function interpolation, (with A.C. Faul), in Numerical Analysis 1999 (Dundee), Chapman & Hall/CRC (Boca Raton), 115–141 (2000).
[161] On the convergence of the DFP algorithm for unconstrained optimization when there are only two variables, in Studies in algorithmic optimization, Math. Program. 87, 281–301 (2000).
[162] System modelling and optimization. Methods, theory and applications. Proceedings of the 19th IFIP TC7 Conference held in Cambridge, July 12–16, 1999, (edited by M.J.D. Powell and S. Scholtes), Kluwer Academic Publishers (Boston), 2000.
[163] On the Lagrange functions of quadratic models that are defined by interpolation. Dedicated to Professor Laurence C.W. Dixon on the occasion of his 65th birthday, in Optim. Methods Softw. 16, 289–309 (2001).
[164] UOBYQA: unconstrained optimization by quadratic approximation, in Math. Program. 92, 555–582 (2002).
[165] On trust region methods for unconstrained minimization without derivatives. New trends in optimization and computational algorithms, in Math. Program. 97, 605–623 (2003).
[166] On the use of quadratic models in unconstrained minimization without derivatives. The First International Conference on Optimization Methods and Software, in Optim. Methods Softw. 19, 399–411 (2004).

[167] Least Frobenius norm updating of quadratic models that satisfy interpolation conditions, in Math. Program. 100, 183–215 (2004).
[168] A Krylov subspace algorithm for multiquadric interpolation in many dimensions, (with A.C. Faul and G. Goodsell), in IMA J. Numer. Anal. 25, 1–24 (2005).
[169] The NEWUOA software for unconstrained optimization without derivatives, in Large-scale nonlinear optimization, Springer (New York), 255–297 (2006).
[170] Fast evaluation of polyharmonic splines in three dimensions, (with R.K. Beatson and A.M. Tan), in IMA J. Numer. Anal. 27, 427–450 (2007).
[171] A view of algorithms for optimization without derivatives, in Math. Today (Southend-on-Sea) 43, 170–174 (2007).
[172] Foreword (Dedicated to Arieh Iserles on the occasion of his sixtieth birthday), (with S. Nørsett), in Found. Comput. Math. 8, 287–290 (2008).
[173] Developments of NEWUOA for minimization without derivatives, in IMA J. Numer. Anal. 28, 649–664 (2008).
[174] On the convergence of a wide range of trust region methods for unconstrained optimization, in IMA J. Numer. Anal. 30, 289–301 (2010).
[175] On nonlinear optimization since 1959, in The birth of numerical analysis, World Sci. Publ. (Hackensack), 141–160 (2010).
[176] On the convergence of trust region algorithms for unconstrained minimization without derivatives, in Comput. Optim. Appl. 53, 527–555 (2012).
[177] Beyond symmetric Broyden for updating quadratic models in minimization without derivatives, in Math. Program. 138, 475–500 (2013).
[178] The Lagrange method and SAO with bounds on the dual variables, in Optim. Methods Softw. 29, 224–238 (2014).
[179] On fast trust region methods for quadratic models with linear constraints, in Math. Program. Comput. 7, 237–267 (2015).