(i) Trial and error

This is the simplest method. The parameter x is varied in a more or less systematic manner until a minimum value of the function is found. Such a mapping of the function is useful when it has many minima. A striking example is given by Bernal [6], in a problem arising from crystal field calculations.

The trial and error method has been widely used to determine the formation constant K of a simple complex AB. K can be calculated independently at each datum point and the spread of values so obtained is minimised (i.e. the standard deviation on K). The data may be spectrophotometric [7—9] or calorimetric [10]. The heat of formation of an AB complex has been calculated in this way also [11].

(ii) Golden section

To find the minimum precisely the trials must be conducted efficiently. For example, the minimum is first bracketed, and the bracket is then successively narrowed by division in the ratio of the golden section, $\tau = [(5)^{1/2} - 1]/2$ [1]. This method has been applied to the spectrophotometric determination of K_{AB}, where it converges quite rapidly [12], and to solvent extraction data (S.I. Cheng in ref. 13). However, bracketing the minimum is easy only for functions with one minimum [12], so the method is not guaranteed to converge for general functions.

(iii) Quadratic interpolation

The same defect is possessed by the fastest methods for linear minimisation which use a simple polynomial fit to two or three points in order to predict the position of a minimum within a bracket. Assuming a parabolic approximation to the function, the minimum is estimated by quadratic interpolation. Wentworth et al. [14] used the mid-point within the bracket as the third value, as did Sillén et al. [15]. Very effective strategies have been devised by Davies, Swann and Campey (see ref. 4) and Powell [16]. Quadratic interpolation greatly speeds the linear minimisation as long as the underlying assumption is reasonably valid, but this may only be true when the bracket is already narrow. Powell's method is included in MINIQUAD (FORTRAN, subroutine LIMIN) [17].

(iv) Cubic interpolation

Another very effective method of linear minimisation requires the function value and its derivative with respect to the parameter, but only needs these at the two ends of the bracket. The function is approximated by a cubic polynomial. This method seems to have originated with Davidon [18] and is described succinctly by Fletcher and Powell [19].

C. DIRECT SEARCH METHODS

Direct search methods take their name from the fact that the parameters are varied directly, and the function is minimised by noting the effects of these variations; after one cycle it may be that linear combinations of parameters are varied, rather than the parameters themselves. Direct search methods are generally not as efficient as gradient methods (see Sect. D) but are simple and often are useful for obtaining initial parameter estimates for those methods.

(i) Grid

In its simplest form, an n-dimensional grid is formed and the minimum is located in one unit n-space of the grid. The grid mesh may then be reduced (or increased) in order further to localize the minimum. Such an operation is very expensive in computer time [20] since r^n points are needed for a grid with r points along each side. Nevertheless a grid search is preferable to a random search [4]. The grid search has the advantage of clearly revealing multiple minima [21], but obviously must converge extremely slowly. Grid searches have been used to obtain parameter estimates for fitting nuclear decay [22] and ORD curves [23] and for conformational calculations [24] and to map the function in calorimetric titrations [21]. They have little to recommend them except simplicity.

(ii) Axial iteration

Varying each parameter in turn (axial iteration) reduces the number of function evaluations required, but a smaller region of the n-space is explored, and alternative minima may be missed. If the function is linearly optimised during every axial search, convergence is guaranteed, though perhaps very slow [1]. Axial iteration will fail to converge at all if the function being minimised is virtually insensitive to the change in one or more parameters' value.

Axial iteration has been used to obtain initial parameter estimates in the determination of formation constants spectrophotometrically [25] and potentiometrically [26] and as a prelude to the solution of simultaneous non-linear equations by a regula falsi method [27]. Vibrational force constants have been determined by a kind of axial iteration, in which the function was evaluated many thousand times [28].

(iii) Hooke and Jeeves

More rapid convergence can be expected if information about the objective function obtained during the exploration along the parameter axes is used in subsequent iterations. This is the basis of the following methods, all of which have good convergence properties.

In the method of Hooke and Jeeves [29] a cycle of searches of fixed length along each parameter axis (exploratory move) from x^i to give x^{i+1} as the best local parameter vector is followed by a search along the direction of the vector $x^{i+1} - x^i$ (pattern move). Exploratory and pattern moves then alternate, the length of the exploratory search being reduced as the minimum is approached (direct search, ALGOL) [30]. The method has been used to obtain self-association constants for monomer—dimer equilibria [31] and to obtain the crystal field splitting parameters from Nd^{3+} and Eu^{3+} spectra [32]. It can also be used to obtain the optimal orbital exponents when a Slater type orbital (STO) is expanded as a limited set of Gaussian orbitals [33, 34], and indeed to minimise the self-consistent field (SCF) energy in quantum-mechanical calculations [35—37].

(iv) Rosenbrock

Rosenbrock's method makes more use of the directions generated by the axial searches. In fact the parameter axes are replaced by a progression of orthogonal vector sets designed so that each line search (of fixed length) produces the greatest decrease in the objective function [38]. A particular advantage of this method is that it is easily adapted to deal with constrained minimisation, as in a study to determine the pathway for insertion of a singlet methylene group into a C—H bond. The $:CH_2$ group is constrained to approach the C—H bond symmetrically [39]. The method has been little used in chemistry [40] but has been used extensively in other fields. Its advantages are simplicity, directness and good convergence.

A variant of this method [41] in which each line search is subjected to one quadratic interpolation is claimed to be somewhat more effective, but this method is not referred to in the open literature.

(v) Powell 64

This method [16] is so designated to differentiate it from other methods by the same author who has made so many valuable contributions to the field. It is undoubtedly the most effective of the direct search methods in terms of guaranteed and rapid convergence, but it is also the most sophisticated. It generates a progression of "conjugate directions" for which powerful convergence proofs can be derived [1].

In practice the promise implied by these proofs is often fulfilled. For example, rates of reaction at various pH values were fitted to a nine-parameter model (five rate constants, four equilibrium constants) using Powell 64 where three other methods had failed to converge [42]. The method has been used successfully in many other data-fitting problems such as optical dispersion studies [43], formation constant calculations [12,44,45], fitting NMR contact shifts [46], X-ray powder patterns [47] and deriving second virial coefficients [48]. It is widely used to optimise the Gaussian orbital exponents in STO expansions [49] and a general purpose curve-fitting program in which Powell 64 is optionally available has been described [50]. Powell 64 has been used to minimise Schottky defect energy as a function of atomic displacements [51], but a gradient method (see Sect. D) is probably superior [52]. A similar conclusion was reached in connection with geometry optimisation by a semi-empirical MO method [53]. Powell 64 has also been used to minimise SCF energy directly [54,55]. A FORTRAN subroutine is contained in the A.E.R.E. Harwell Subroutine library (VAO4A).

(vi) Simplex

Here the minimum is localised in a volume of n-space. In the "regular simplex" method [56] the volume is defined by a set of $n + 1$ mutually equidistant vertices, i.e. an equilateral triangle in 2-space, a tetrahedron in 3-space, etc. The local minimum is found by examining those simplexes obtained by reflecting a vertex through the (hyper) plane formed by the other vertices. The vertex with the largest function value is rejected and the search is systematically continued until it becomes necessary to reduce the size of the simplex. Nelder and Mead (NELMIN, FORTRAN) [57] extended the concept by allowing the simplex to vary in size and shape, which allows it to adapt to the local behaviour of the objective function. It is clearly a powerful method for obtaining approximate minima, but the convergence properties are poor.

The Nelder—Mead method has been used to fit polynomials in adsorption studies [58] and to test a new theoretical sorption isotherm [59]. It has also been used to obtain hyperfine splitting parameters in well-resolved ESR spectra [60] and to determine reaction rate constants [61]. It has been used to determine optimal molecular geometries in ground and excited states in the MINDO/2 approximation [62,63]. A novel application is in the determination of the coefficients needed when chromatographic peaks are described by Fourier analysis [64].

(vii) Conclusions

Direct search techniques principally perform two services — to obtain approximate solutions and to minimise functions which cannot be differentiated analytically. For the former purpose the simple methods are adequate, with Hooke and Jeeves [29,30] or Nelder—Mead [57] methods giving a reasonable balance between simplicity and efficiency. However, the relatively slow convergence means that it is difficult to define convergence criteria satisfactorily. If too slack a convergence criterion is imposed, the refinement may terminate before the approximate minimum has been reached. If the convergence criterion is too strict the model will be refined to the true minimum, which is inefficient because of the slow convergence. Another advantage accruing from the simplicity of direct search methods is that they require very modest amounts of computer store, which makes them very suitable for implementation on minicomputers. For complex problems Powell 64 [16] is indicated.

D. GRADIENT TECHNIQUES

Gradient techniques are based on the proposition that the gradient g of a function S is equal to zero at the minimum.

$$g_j = \frac{\partial S}{\partial x_j} = 0, g = \{g_1 \dots g_n\} \tag{3}$$

The problem of finding an unspecified minimum is therefore replaced by the problem of finding the parameter vector x^{min} for which the gradient has the specific value of zero. This is computationally a much simpler problem and was solved by Isaac Newton. An estimate of the minimum, x is improved by adding to it a shift s calculated so that, given certain conditions, $x^{\mathrm{min}} = x + s$. The first condition is that the function is quadratic in the parameters x, as expressed in eqn. (4).

$$S = a + x^{\mathrm{T}}b + \tfrac{1}{2}x^{\mathrm{T}}Gx \tag{4}$$

The gradient is obtained by differentiation,

$$g = \frac{\partial S}{\partial x} = b + Gx, \qquad g^{\mathrm{min}} = b + Gx^{\mathrm{min}} = 0 \tag{5}$$

from which $b = -Gx^{\mathrm{min}}$. The second condition is that G, the matrix of second derivatives, or Hessian, eqn. (6), is positive definite.

$$G_{ij} = \frac{\partial^2 S}{\partial x_i \partial x_j} \tag{6}$$

(In the case of a function of one variable a minimum is specified by $g = 0$ and G positive). Upon rearranging the terms of eqn. (5) we obtain

$$g = -Gx^{\mathrm{min}} + Gx = -G(x^{\mathrm{min}} - x) = -Gs \tag{7}$$

The shift vector s is therefore obtained by solving the n linear simultaneous equations $Gs = -g$ for the n parameters $\{x_1 \dots x_n\}$.

If the function is not everywhere quadratic, x^{min} must be sought by successive approximations, and it will be found if S is quadratic at the minimum. In fact all functions are approximately quadratic in a region "near" to the minimum. Raphson expressed the refinement with the recursion formula

$$x^{k+1} = x^k + s^k \tag{8}$$

where k is an iteration number [65]; the iterative method is therefore known as the Newton(—Raphson) method, and it can be used to minimise any function. However, two specific function forms merit special discussion. If the function is a sum of squares, as in

$$S = \sum_{i=1}^{i=m} v_i^2, \qquad v_i = y_i^{\mathrm{obs}} - y_i^{\mathrm{calc}} \tag{9}$$

it is easy to derive eqns. (10) and (11).

$$g_j = -2 \sum_{i=1}^{i=m} J_{ij} v_i, \qquad J_{ij} = \frac{\partial v_i}{\partial x_j} = \frac{\partial y_i}{\partial x_j} \tag{10}$$

$$G_{jk} = 2 \sum_{i=1}^{i=m} - v_i \frac{\partial^2 y_i}{\partial x_j \partial x_k} + \frac{\partial y_i}{\partial x_j} \frac{\partial y_i}{\partial x_k} \tag{11}$$

By ignoring the first term in eqn. (11) involving the second derivatives $\partial^2 y_i / \partial x_j \partial x_k$ one obtains an approximation to the Hessian that involves first derivates only, $G \approx 2J^{\mathrm{T}}J$, which can be substituted with eqn. (10) into eqn. (7) to give

$$-2J^{\mathrm{T}}Js = -2J^{\mathrm{T}}v \tag{12}$$

If the observations have an associated weight matrix W, it is easy to derive

$$(J^{\mathrm{T}}WJ)s = J^{\mathrm{T}}Wv \tag{13}$$

known as the normal equations for least squares. Equation (13) represents the Gauss—Newton(—Raphson) method. Although its derivation by Gauss, that is, the principle of least squares, was based on quite different considerations [66—68], we have derived it in this way in order to show its computational affinity to the Newton(—Raphson) method.

If $m = n$, we have a set of non-linear simultaneous equations $f_i = 0$ and eqn. (13) simplifies to

$$Js = v \tag{14}$$

This is known as the Newton(—Raphson) method for non-linear equations.

It is regrettable that the names of Newton, Raphson and Gauss are used somewhat indiscriminately in the literature, a fact that can cause the uninitiated reader much confusion!

E. LINEAR LEAST SQUARES

When the functions y_i are linear in the parameters x (as, for example, when y_i is a polynomial in the independent variable z_i) equation (15) results.

$$y_i = \sum_{j=1}^{j=n} x_j J_{ij}, \qquad J_{ij} = f_j(z_i) \tag{15}$$

The sum of squares S is verifiably a quadratic function in x and G is constant. It follows that the Gauss—Newton method should converge to $x^{\min}$ in one step from any trial value x^0. One may therefore choose $x^0 = 0, y^0 = 0$ without loss of generality and eqn. (13) becomes $(J^{\mathrm{T}}WJ)x^{\min} = J^{\mathrm{T}}Wy^{\mathrm{obs}}$. Thus, a linear least squares problem is distinguished by the fact that initial parameter estimates are not required and by the fact that the solution is not iterative. Also, because of the form of eqn. (15), no explicit differentiation is performed.

(i) Polynomials

Polynomials are widely used in order to define empirical relationships between data and an independent variable; an example is the e.m.f. of an Ag/AgCl electrode expressed as a cubic polynomial of media effects [63]. A sixth-degree polynomial was used to fit reaction rate data in order to integrate the rate equation [70]. Cubic polynomials are often used to define the base-line of a spectrum, e.g. γ-ray spectra [71,72]. Formation constants of ML_n complexes may be obtained from polynomial fits, but poor computational procedures may lead to misleading answers [73].

For high-order polynomials, numerical difficulties are almost certain to arise when using conventional techniques to solve the normal equations (13), because the equations are intrinsically likely to become ill-conditioned for large n [74]. The problems are best solved, not by using sophisticated techniques to solve the normal equations (a procedure limited by the precision of the computer) but by reformulating the problem so that y_i is a sum of "orthogonal polynomials". This automatically removes the ill-conditioning from the normal equations and has a further significant advantage: having solved for polynomials of order up to n, polynomials of higher order may be added without having to recalculate the coefficients of the lower-order ones! Thus the task of finding the lowest-order polynomial that gives an adequate fit to the data is greatly simplified. An application of this kind to polarographic data has been given by Karmalkar [75]. In the design of magnets for β-ray spectrometers the standard least squares method failed for $n > 7$, but orthogonal polynomials were satisfactory [76].

Because of the advantages described above, the use of orthogonal polynomials is recommended even when n is small.

(ii) Linear models

Examples of linear models are provided in those situations where the observed effect is a simple sum of contributing factors. For example, a heat change is the sum of heat changes for each competing reaction [77], or a mass spectrum peak is the sum of peaks belonging to ions of similar mass [78,79]. These problems are solved by conventional techniques. A set of absorption spectra might be expressible as combinations, in various proportions of two or more simple spectra as, for example, when two chemical components are present in varying proportions. A sophisticated linear least squares technique then provides both the spectra of the components and the proportions present in the observed spectra even when the component spectra cannot be observed independently [80].

(iii) Library spectra

In neutron activation analysis an observed spectrum consists of a sum of the spectra of the constituent elements [81]. If the spectra of the elements are stored in standard form in a "library", linear least squares techniques may be used to determine both the elements present and their proportions [81,78]. Eckhoff [82] applied this method to γ-ray spectra and applied statistical tests to determine whether the proportion of any component is significant, given that even the library spectra are subject to uncertainty. The method is useful with the overlapped spectra obtained when using a scintillation detector, but has perhaps been superseded by the use of semiconductor detectors which give more clearly resolved spectra. A similar technique has been applied to the automatic identification of compounds from their six most intense mass spectrum peaks, using the ASTMS catalogue of reference spectra [83].

(iv) Rigorous linear least squares

The term rigorous is applied to those problems where the independent variable, z in eqn. (1), is subject to experimental error. (The usual least squares treatment assumes z is error free.) Cumming et al. [84] describe the computation of two formation constants by "rigorous" straight-line fitting. In this situation proper weighting becomes extremely important, as shown by the fact that pK values may be more than one pK unit in error when the calculation is not properly carried out.

(v) Smoothing and differentiation

A curve comprising a set of data at equal intervals of the independent variable may be smoothed to remove excessive experimental "noise" by piecewise fitting of a polynomial to small sections of the curve and by replacing each experimental point by one calculated from the polynomial's coefficients. The smoothed differentials of the original curve may also be obtained from the same coefficients. In practice, because of the equal abscissa intervals, the whole process is arithmetically trivial; one calculates the convolute, or weighted mean of the experimental points multiplied by appropriate convoluting integers, which may be obtained from tables [85]. Data smoothing is usually performed prior to some other data reduction process. Differentiation finds a major application in the automatic location of peak maxima in spectra (see Sect. K(iv)).

Spline functions may also be used for piece-wise fitting of data curves. A spline function of order n is a set of nth order polynomials with continuous derivatives up to but not including the nth derivative. Thus, a first-order spline function is simply a set of straight-line segments. Cubic spline functions have continuous first and second derivatives, i.e. the slope and curvature are continuous, and each segment of the curve is a cubic polynomial [86]. Spline functions allow any experimental curve to be described wholly by one empirical function, including skewed spectrophotometric absorption curves [87].

F. NON-LINEAR LEAST SQUARES

An iteration of non-linear least squares refinement consists of setting up and solution of the symmetric linear equations (13). Experience has shown that while there are many cases where the iterative refinement proceeds smoothly to the required minimum there are many other cases where this is not so. We concentrate on the latter since methods which converge with difficult problems are a priori methods which will converge most efficiently on general problems. The extra time and effort required to implement the more elaborate methods should be amply repaid if non-convergence is avoided.

From eqns. (10) and (11), the whole Jacobian need not be stored; only the n elements corresponding to the ith data point are needed at any one time. The principal cause of convergence difficulties lies in the approximation that $G \approx 2J^TWJ$. There is therefore no point in using double-length arithmetic in calculating the approximate G. However, care must be exercised in solving the normal equations. Although the solution can be expressed as $s = (J^TWJ)^{-1}J^TWv$ it should not be obtained by first inverting the matrix (J^TWJ). The method usually preferred is based on Choleski factorisation [74] using single-length arithmetic, though iterative improvement of the solution [74] costs little in extra store and time. The Gauss—Jordan method with partial pivoting (MTRXIN, FORTRAN) [88] is often satisfactory, though it does not have the stability of the Choleski method. The matrix $(J^TWJ)^{-1}$ which is needed for the standard deviations of the parameters should be calculated separately; a routine that performs both operations simultaneously is included in ref. 17 (LINEQ, FORTRAN).

An alternative method of solving the normal equations, due to Golub, involves orthogonal transformations [89]. This is faster than elimination methods on well-conditioned matrices and not inferior on badly conditioned ones [1]. It has been used in force constant calculations [90]. Its main drawback is that the whole Jacobian must be stored.

If the shift vector calculated is nevertheless such that the sum of squares does not decrease, this indicates that s is too long and perhaps also points in the wrong direction. The former defect is corrected by a technique we call step cutting, the latter by damping.

(i) Step cutting

The basic iteration equation (8) is modified so that a fraction of the shift vector is applied

$$x^{k+1} = x^k + \alpha s^k \tag{16}$$

Aldous and Mills chose $\alpha = \frac{1}{2}$ for the first few cycles [91]. Law and Bailey choose the largest useful value of α in the series $1, \frac{1}{2}, \frac{1}{4}, \ldots$ [92], and this approach has worked well for Mössbauer spectra [93,94] and chromatography peaks [95,96]. $\alpha = \frac{1}{10}$ has been suggested [97] as has the series $1, \frac{1}{10}, \frac{1}{100}, \ldots$ [98]. Wentworth et al. found an optimal value by quadratic interpolation between $\alpha = 0, \frac{1}{2}, 1$ [14]. A quadratic interpolation, followed by a second one if deemed necessary, was used by Sabatini and Vacca [99] and in a generalisation of their program [17] α was optimised to a tolerance of 20% using Powell's method [16].

If α is linearly optimised, then convergence of the Gauss—Newton method is guaranteed [100]. Therefore a good step-cutting strategy gives a refinement that fails only when the normal equations matrix appears to be not positive definite.

(ii) Damping

Step cutting may appear to be inefficient if the search vector s is pointing in the wrong direction; a rotation towards the direction $-g$, along which the function has the greatest local rate of decrease, might be helpful. This result is achieved if the function is minimised subject to the condition that the length of s is also minimised. As a result eqn. (13) becomes

$$(J^TWJ + \lambda D)s = J^TWv \qquad (17)$$

D is a diagonal "weight" matrix usually taken either as a unit matrix or as equal to the diagonal elements of (J^TWJ). Equation (17) has been derived independently on at least five occasions [101—105], and Marquart's coding has become widely used through the IBM SHARE library (FORTRAN), and λ is now known as the Marquart parameter.

The chief difficulty with the damped least squares method is to choose a good strategy for the scalar λ. Frazer and Suzuki give a graphic illustration of this for absorption curve fitting in ref. 78. Fletcher has modified Marquart's strategy in such a way that on many iterations a second function evaluation is not needed [106]. Such an efficiently coded algorithm (VAO7A, FORTRAN) [106] may be much faster than Marquart's, with the result that comparisons of efficiency (see Sect. J) may be invalidated.

Step cutting may also be applied to the damped shift [107]: Meyer and Roth claimed that their method converged where Marquart's failed on double exponential decay data. The step fraction was obtained by cubic interpolation for force constant calculations [108].

An early application of Levenberg's method to force constant calculation [109] clearly has stimulated others in this field [110—114], and in fitting NMR spectra [115] and chromatograms [116].

Marquart's method has been used to fit microwave spectra [117,118], resonance spectra [119], GLC peaks [120] and osmotic coefficient data [121], and to calculate second virial coefficients [122,123], optical constants [124] and formation constants [125,126]. It has also been used to determine interatomic potentials numerically using poor parameter estimates [127] and with solvent extraction [128,129] and kinetic studies [130,131]. A general purpose curve-fitting program has been described [50] and another published by Frazer and Suzuki in ref. 78 (BANDFIT, FORTRAN).

From the wealth of successful applications it will be clear that the Marquart modification of the Gauss—Newton (Raphson) method is the most useful general purpose method for non-linear least squares problems. Its advantage over step cutting is that the latter will fail if the normal equations matrix (J^TWJ) is not positive definite, whereas λD can always be chosen so that ($J^TWJ + \lambda D$) is positive definite.

(iii) Numerical derivatives

When the Jacobian J cannot be evaluated analytically, it is obvious that it can be obtained by a finite difference method, $J_{ij} = \delta y_i / \delta x_j$, requiring $m \times (n + 1)$ function evaluations. This approach is used in the program SCOGS (FORTRAN) [132] (see also ref. 133) to determine formation constants, a program that occasionally fails to converge [134,135]. It has also been used to fit chromatographic peaks with empirical shape functions [95]. However, this method suffers from three serious defects: (1) the increments for the finite difference must be chosen very carefully to be neither too large nor too small; (2) any derivative approximated by a straight line segment is subject to truncation error [67]; (3) the errors in J are multiplied together in forming (J^TWJ). These defects are avoided in a method we call Powell 65 [136]. Here, a set of directions and an approximation to (J^TWJ) are updated. This excellent algorithm is not in wide circulation (A.E.R.E. Harwell library program VAO2A, FORTRAN). Anderson et al. used it for chromatography peaks and found that it used slightly more store than that of Law and Bailey [95]. Hyperfine and quadrupole tensors were obtained from ESR spectra using Powell 65 [137]. It has also been applied to Mössbauer spectra [94] and to refine extinction and anharmonic parameters in the crystal structure of calcium fluoride [138].

A modification of Marquart's method to accept numerical estimates of the derivatives was applied to the least squares fit of atom—atom scattering angles [139]. This must be very expensive in computer time. Another algorithm that uses numerical derivatives, due to Peckam [140], uses fewer function evaluations than does Powell 65, but is a little less efficient in terms of the number of routine computing operations. Yet another algorithm exists (A.E.R.E. Harwell library program VAO5A, FORTRAN) which uses a hybrid Marquart—steepest descent approach. Details of this have not been published, though it is an extension of Powell's method for non-linear equations (FORTRAN) [141]. More details about these methods may be found in a comprehensive review by Powell in ref. 142.

(iv) Rigorous non-linear least squares

Although the basic theory has been described by Deming [68], there are few reported applications. Wentworth [143] describes a program used in conjunction with teaching experiments, and Lingane and Hugus [144] a program for the spectrophotometric determination of formation constants.

It seems likely that, as the precision of measurements improves, errors in the independent variables will become significant and that rigorous least squares will be more widely used. A startlingly unconventional differentiation technique is claimed to give a rapidly convergent algorithm [145].

(v) Miscellaneous

In SPIRAL [146] the end of the trial search vector *s* traces a trajectory between the Gauss—Newton direction and the steepest descent direction. Jones applied this program to a variety of chemical data and found that convergence was more rapid than with Marquart's coding of the damped least squares method. It is doubtful if SPIRAL has much advantage over Fletcher's coding [106].

The generalised or pseudo-inverse has recently attracted attention, e.g. in force constant calculations [147]. However, the damped least squares method turns out to be a particular example of the pseudo-inverse [148,149].

Crystallographers often adopt a block-diagonal approximation to the normal equations matrix. In this case compensation for the errors introduced by ignoring the off block-diagonal elements is obtained by adjustment of the length of *s* [150].

G. SEPARABLE VARIABLES

If often happens that some parameters appear linearly in the defining equations; examples occur with calorimetric titration [151] and the Arrhenius equation. Linear and non-linear parameters are then said to be separable. Given the "non-linear" parameters, the "linear" parameters can be obtained by linear least squares [152]. The advantages of this device as a general procedure are a reduced dimension and a simpler surface to the function leading to faster and more stable refinements. Application to, for example, double exponential decay data [152] brings this out clearly. However, the differential coefficients of the "non-linear" parameters must be evaluated with care. In SCF calculations orbital coefficients have long been treated separately from orbital exponents [153].

A truly startling development in technique for separable variables has been presented by Golub and Pereyra [153(a)] from the original variable projection algorithm of Scolnik. In this, the "linear parameters" apparently vanish from consideration; no initial estimates are required for them, and the optimal values appear in the solution! The method is based on the differentiation of the pseudo-inverse and uses Golub's orthogonal factorisation technique [89]. It is particularly useful for fitting sums of exponentials, i.e. Gaussian functions, which arise in decay and curve-fitting problems.

H. GENERAL FUNCTIONS

General functions can be minimised by direct search methods, or, when derivatives are available by Newton's method, by eqn. (7). When the Hessian is obtained by successive approximations using first derivative (gradient) information only we have a class of methods known as quasi-Newton methods.

(i) Least squares functions by Newton's method

Equation (11) applies to the minimisation of a sum of squares by Newton's method. Its advantage is more rapid convergence "near" the minimum, but against this must be set the time required to evaluate the second derivatives $\partial^2 y_i/\partial x_j \partial x_k$ and the fact that away from the minimum the second derivatives are irrelevant.

Orioli and Vacca compared the Newton—Raphson and Gauss—Newton methods for crystal structure refinements [154] and found that, while fewer iterations were needed, the overall time required was similar. Parameters in atomic Thomas—Fermi—Dirac potential functions were obtained by the Newton—Raphson method where the Gauss—Newton method failed [155]. However it was found necessary to optimise the shifts by quadratic interpolation [155]. (N.B. these are sums of exponentials with linear coefficients.) Newton—Raphson and Gauss—Newton methods with shift optimisation in each case were compared for formation constant calculation [99], but inclusion of the second derivatives increased the time needed for convergence.

We conclude that the Newton—Raphson method is not particularly suitable for least squares problems, except perhaps when exponentials are involved.

(ii) Quasi-Newton methods

The basis of these methods is that an approximation to the Hessian, or more usually its inverse, is updated at each iteration, in which a shift vector is calculated from eqn. (7) $G^i s^i = -g^i$ or its inverse $s^i = -H^i g^i$. The approximate Hessian G^i or its inverse H^i is required to be positive definite and will usually be "exact" after n iterations if the function is quadratic.

Davidon published the first report of a quasi-Newton method [18], and his method has been applied to least squares fitting of Mössbauer [156], ESR [157] and γ-ray spectra [158], and to calculate stability constants [159]. Davidon's method was improved by Fletcher and Powell [19] (FLEPOMIN, ALGOL) [160] and convergence proofs were given. Although the method has theoretically a numerical instability [161], proofs that it will converge for general functions [162] are supported by the widespread successful use of the method [52,163,164]. This is particularly true of SCF calculations [165–168].

An alternative updating formula which is better than the original one has been developed by Fletcher [169] and others, and a program using this has been published (VAO9A, FORTRAN) [170]. Another updating scheme [171] was used to minimise SCF energy and was found to be five times faster than Powell 64 [52]. Biggs has developed a scheme in which H^i maintains positive definiteness when the function is not quadratic [172]. This was successful on some exponential sum problems where the other updating formula failed to converge.

All these methods update the approximate inverse Hessian. Gill et al. update a factorised form of the Hessian itself in an attempt to improve numerical stability [173] and have published their program [174].

(iii) Damped Newton's method

As with the Gauss–Newton method, the Hessian may be modified by adding to it a diagonal matrix, particularly when the unmodified Hessian is not positive definite. This underlies Greenstadt's method [175] which has been applied to the evaluation of two-centre integrals [176]. A similar technique [177] has been modified and used for spin-optimised SCF wave functions where the undamped Newton–Raphson method failed [178]. A comparison on test functions shows the damped Newton method to perform well when compared with Davidon, Powell 64 and Hooke and Jeeves [5].

(iv) Numerical derivatives for quasi-Newton methods

Stewart has described how numerical differentiation may be incorporated into the Fletcher–Powell method, using the approximate inverse Hessian in order to choose those increments which minimise the truncation error [179], and these suggestions are embodied in DAPODMIN(ALGOL) [180]. This has been applied to potentiometric titration data [181] and to the general computation of formation constants (in a FORTRAN translation) [182]. A program that incorporates alternative updating formulae is given by Fletcher (VAIOA, FORTRAN) [170].

(v) Pit mapping

This name was coined by Dryssen et al. [183] to describe the process whereby the function is assumed to be quadratic, and the coefficients a, b and G of the quadratic expression (4) are obtained numerically. It is therefore equivalent to Newton's method with numerical derivatives. It requires $\frac{1}{2}(n+1)(n+2)$ evaluations of the function per iteration, compared with $n+1$ for Stewart's method. Greatly improved results are obtained if the axes, along which the increments are taken, are "twisted" to give maximum function decrease for each parameter increment (LETAGROPVRID, ALGOL) [184]. The method works well if the underlying quadratic assumption is a good one, though often a precalculation is needed to obtain this condition [185]. Apart from potentiometric titration work, for which it was originally developed [186], pit mapping has been applied to NMR coupling constants [187] (initial estimates by simulation), solvent extraction data [188] and formation constant determination from spectrophotometric [25,189,190] and calorimetric data [191].

Pit mapping was "rediscovered" for fitting radioactive decay curves [22] (initial grid search) and for minimising SCF energy [37]. In the latter case, however, the minimum obtained by the method of Hooke and Jeeves was not improved by pit mapping.

(vi) Miscellaneous

Use of $s = -g$, known as the method of steepest descent, has guaranteed convergence, but in practice extremely slow convergence, on general functions, for reasons discussed in ref. 1. Early use of this method for SCF energy minimisation has now been discontinued.

$-g$ may be used as the initial vector for generating a set of "conjugate directions"; after n of these have been searched a quadratic function will be minimised. Fletcher and Reeves implemented this idea (FUMICOG, ALGOL) [192] in an algorithm which has minimal storage requirements. For this reason it may be well suited for problems with many parameters, and it is effective for the linear parameters in SCF calculations [153,166]. It has also been used to minimise SCF energy [193] but, while it is better than steepest descent, Roothaan's method is better still [168].

I. NON-LINEAR EQUATIONS

Least squares problems which are not greatly overdetermined ($m \approx n$) can lead to meaningless fits [194]. When $m = n$, eqn. (14) is appropriate, $Js = v$. A fundamental difficulty arises because J may be unsymmetric and/or not positive definite (unlike J^TWJ which should always be so) and convergence is not guaranteed. This has led Sillén's group to prefer a regula falsi method to Newton's method [27,195]. Convergence problems were noted by Bard and King [98] and avoided in COMICS(FORTRAN) [196] by a most unusual method of successive approximations. These problems all arise in the field of solution equilibria, in the calculation of free reactant concentrations.

The state of the art has been surveyed recently [197].

Powell's hybrid method which does not require derivatives has been mentioned [141]. Powell also states [198] that an algorithm due to Broyden [199] is faster and more reliable than generalised secant methods, see ref. 1.

The computation of vibrational force constants by fitting vibration frequencies is notoriously liable to non-convergence, mainly because the parameters are often physically underdetermined ($m < n$). Many efforts have been made to overcome this problem [147], usually using step cutting or damped least squares. For the refinement of diagonal symmetry force constants ($m = n$) J may be singular at the minimum, but in spite of this the Fletcher—Powell method always minimised the sum of squares smoothly and efficiently [200]. In certain circumstances the Hessian may also be singular at the minimum. Refinements using the Fletcher—Powell method were also successful here, though the refinements became relatively slow [201].

J. COMPARISONS OF METHODS

The efficiency of a minimisation algorithm can be measured by its reliability, i.e. whether it always converges satisfactorily, and its speed. For the chemist speed, though desirable, is less important and fortunately the most reliable algorithms are often the fastest. Unfortunately the reliability of an algorithm is affected by external factors such as the formulation of the problem and the quality of the initial parameter estimates.

Bard compared the efficiency of 13 algorithms on 10 test problems of least squares, maximum likelihood and Bayesian estimation types [202]. The Gauss—Newton method with "crude" shift optimisation came out best, closely followed by Marquart's method. Quasi-Newton methods were consistently slower and sometimes unreliable.

Duke and Gibb compared six methods for fitting Mössbauer spectra [94]. Gauss—Newton with crude shift optimisation, following Law and Bailey [92], sometimes failed to converge but was, with Powell 65, the fastest method. Storage presents difficulties with large problems when using Powell 65. Fletcher—Powell and Powell 64 both failed even with good parameter estimates. Fletcher—Reeves always converged, but slowly. Nelder—Mead was slow but did converge with good parameter estimates. Marshall et al. found Davidon's method unsuitable for small computers, and settled for Law—Bailey [93].

Pitha and Jones compared methods for fitting IR spectra [203], but did not consider a shift-optimised Gauss—Newton method. Levenberg's method with optimisation of λ, eqn. (17), sometimes failed, but a modified version proved to be better than Marquart's method using $D_{ii} = (J^{T}WJ)_{ii}$ and λ optimisation. Fletcher—Powell was speeded up by using $H^0 = (J^{T}WJ)^{-1}$ in place of $H^0 = E$ * but was still relatively slow.

It is clear that the Gauss—Newton method with shift optimisation or with damping is the best for least squares problems. Shift optimisation involves no arbitrary or ad hoc devices, whereas the damping factor λ is not theoretically determined. Fletcher's strategy [106] probably offers the best compromise between speed and reliability for most least squares problems. An exception comprises models with sums of exponentials, for which a Newton or quasi-

---

* A similar observation is in ref. 201.

Newton method may succeed where a Gauss method may fail, but the variable projection method may be more reliable.

Izatt et al. compared pit mapping with Davidon for entropy titration data. They concluded that Davidon was more reliable than pit mapping but a little slower [21].

Sabatini and Vacca give data which show that pit mapping (simplified LETAGROP) is faster than SCOGS (numerical Jacobian), but showed that, when the problem was reformulated so that an exact Jacobian is obtained, Gauss—Newton with shift optimisation was faster still [99]. The different conclusions reached by Tobias and Yasuda probably result from a different parameter scaling [133].

Thus, when the function cannot be differentiated the same considerations apply as given above. If Powell 65 requires too much store, then perhaps Peckam [140] may be useful.

K. SOME PROBLEMS ARISING

Many difficulties in data fitting have a chemical as well as a computational source. A few of these can be considered here.

(i) Multiple solutions

If the data are insufficiently precise or extensive they may possibly be fitted by two or more parameter vectors. This happens in the determination of pseudo-symmetric crystal structures [204,205], but a false minimum may even be found in a well-defined structure such as adamantane [206]. Multiple minima occur unpredictably in many areas such as in electron diffraction work [207], in fitting chromatography peaks [95] and MCD spectra [208] and in kinetic data [42]. Two sets of force constants fit the (overdetermined) vibrational data for methyl fluoride [91]. Some say that formation constants are occasionally ambiguous [73,209,210], though this suggestion is dismissed by others [15]. Entropy titration data may give two minima [21].

The chemist must always be alert to the possibility that a computed solution is not unique, though no general method exists to assist him here. Closely spaced minima will cause computation difficulties. This is especially true in spectral band analysis where two overlapping bands may be "simulated" by one broader band [211].

(ii) Scaling the parameters

The parameters should be so chosen as to make the function as nearly as possible quadratic [212]. Also, a conflicting requirement is that the elements g_j of the gradient should be comparable in magnitude. For example, the scaling $x = \ln \alpha$ is useful for orbital exponents [153] and β-decay data [213]. For formation constants $x = \log \beta$ put all the parameters on a pK scale and appears attractive [132,133]. However, if the β's appear linearly in the least squares equations this has a disastrous effect on the function, which is roughly quadratic in β[182].

The transformation $x = \ln \alpha$ ensures that α is positive. The transformation $x = \sqrt{\alpha}$ also keeps α, the new variable, positive, since $\alpha = x^2$, but also dictates that the function has two minima, at $\pm x$.

Scaling of the least squares normal equations is desirable, [74] and can sometimes be achieved by parameter scaling [17].

(iii) Constrained minimisation

This is discussed by Box et al. [4] and a decision for algorithms is given by Fletcher [212]. The generalised least squares method can be easily adapted for simple constraints [214]. Equality constraints, such as apply to band intensities in resonance spectra, are simply accommodated by working with a reduced parameter set [37,94,215]. Constraints may be treated as weighted subsidiary equations in crystallography [216].

In chemistry parameters are often subject, for physical reasons, to non-negativity constraints, i.e. $x_j > 0$. This can be achieved by a change of variable [4], though not without attendant dangers (vide supra). For example, in resolving mass spectra, non-negativity constraints tended to introduce a bias against small parameters [217]. For these reasons it is probably better to allow parameters to assume negative values during the refinement, while ensuring that the function is continuous, or to eliminate them completely [218]. An inappropriately negative parameter implies either that the model is wrong or that the data are inadequate; the remedy should therefore be chemical rather than computational.

(iv) Initial parameter estimates

The need for "good" initial parameter estimates to avoid non-convergence has been frequently stressed [6,23,31,94,97,203,209,219,220]. It applies with the most powerful refinement methods, though with less force. Direct searching for initial parameter estimates has already been discussed. Simulation [187] or other analogue devices, e.g. a curve analyzer, are sometimes used.

Differentiation gives an indication of peak positions in experimental curves. First derivatives [220,221] give poor resolution; second derivatives give better resolution [158,222]. However, because experimental noise is magnified by the differentiation process [96] some type of smoothing is essential. The convolution technique (vide supra) may perform smoothing and differentiation simultaneously [85]. Morrey has suggested that third and fourth differentials give the best resolution, but spurious peaks may be introduced [223]. A method for fourth derivatives based on simple differences [224] cannot be recommended, though it may work on closely spaced data [225].

Curve differentiation locates peaks regardless of the functional form of the bands. It can therefore provide a valuable check on the validity of a resolution, as well as providing initial parameter estimates.

(v) Model testing

The use of data fitting to determine the "best" chemical model is mainly an exercise in statistics [66]. Regression analysis for this purpose is discussed in ref. 1. Multiple linear regression was used to find spurious components in γ-ray spectra [226]. Otherwise model selection is not usually carried out systematically; an exception is provided by Varga et al. who investigated up to 128 models in a solvent extraction experiment [227].

An attempt was made to assign NMR spectra automatically by comparison of all reasonable assignments [228]. An automatic selector of species present in solution equilibria is discussed by Sillén and Warnqvist [186] and has been tentatively applied [229,230]. Generally, the computational effort required for rigorous model testing is prohibitive as each model tested involves a complete refinement. However, it is important if such a project is undertaken that all refinements converge satisfactorily.

The number of species required to generate a set of absorption spectra obtained by changing an independent variable can be found by moment analysis [231]. In this way methyl red was shown to exist in three forms, by varying pH [231].

Band resolution may be performed with simultaneous determination of the optimal band shapes (spline functions) [87]. This remarkable technique permits an envelope to be resolved into a minimum number of bands and avoids the difficulty that even well-resolved bands cannot be properly fitted when using analytical shape functions. For example, skewed bands do not require special treatment.

L. ALGORITHMS

Published algorithms have been indicated in the text. Some excellent private libraries exist, such as that of the Nottingham Algorithms Group (NAG), but are not generally accessible. For information about the Harwell library write to Mr. S. Marlow, Building 8.12, A.E.R.E., Harwell, Berkshire (U.K.). This is useful for containing many of Powell's and Fletcher's original codings.

A suite of programs for absorption spectrophotometry is described by Jones [232]. These include scale changes, peak and valley location, smoothing and differentiation and a variety of band separation techniques. The texts of all programs (FORTRAN) are available in the bulletins, refs. 15, 16 and 17, quoted by Jones [232].

REFERENCES

1 J. Kowalik and M.R. Osborne, Methods for Unconstrained Minimisation, Elsevier, New York, 1968.

2 L.C.W. Dixon, Non-linear Optimisation, English Universities Press Ltd., London, 1972.
3 U. Hoffmann and H. Hoffmann, Einführung in die Optimierung, Weinheim, Verlag Chemie, 1971.
4 M.J. Box, D. Davies and W.H. Swann, Non-linear Optimisation Techniques, Oliver and Boyd, Edinburgh, 1969.
5 S.M. Goldfield and R.E. Quant, Non-linear Methods in Econometrics, North Holland, Amsterdam, 1972.
6 E. Bernal G., J. Chem. Phys., 55 (1971) 2538.
7 D.P. Graddon and E.C. Watton, J. Inorg. Nucl. Chem., 21 (1961) 49.
8 M.E. Lamm and D.M. Neville, J. Phys. Chem., 69 (1965) 3872.
9 T.M. Shepherd, J. Chem. Soc. Dalton, (1972) 813.
10 A. Yingst, R.M. Izatt and J.J. Christensen, J. Chem. Soc. Dalton, (1972) 1199.
11 J.J. Christensen, D.P. Wrathall and R.M. Izatt, Anal. Chem., 40 (1968) 175.
12 P. Gans and H.M.N.H. Irving, J. Inorg. Nucl. Chem., 34 (1972) 1885.
13 J.G. Gregory, B. Evans and P.C. Weston (Eds.), Proc. Int. Solvent Extraction Conf., S.C.I., London, 1971, p. 1025.
14 W.E. Wentworth, W. Hirsch and E. Chen, J. Chem. Phys., 71 (1967) 218.
15 L.G. Sillén and B. Warnqvist, Ark. Kemi, 31 (1970) 315.
16 M.J.D. Powell, Comput. J., 7 (1964) 155.
17 A. Sabatini, A. Vacca and P. Gans, Talanta, 21 (1974) 53.
18 W.C. Davidon, USAEC Rep., No. ANL-5990, 1959.
19 R. Fletcher and M.J.D. Powell, Comput. J., 6 (1963) 163.
20 J.G. Becsey L. Berke and J.R. Callan, J. Chem. Educ., 45 (1968) 728.
21 R.M. Izatt, D. Eatough, R.L. Snow and J.J. Christensen, J. Phys. Chem., 72 (1968) 1208.
22 R. Klenk, F. Grass, F. Girsig and O. Brandstädter, Nucl. Instrum. Meth., 94 (1971) 357.
23 J.P. Carver, E. Schecht and E.R. Blout, J. Amer. Chem. Soc., 88 (1966) 2550.
24 J. Langlet and H. van der Meer, J. Mol. Struct., 10 (1971) 91.
25 K. Nagano and D.E. Metzler, J. Amer. Chem. Soc., 89 (1967) 2891.
26 A. Vacca, personal communication.
27 R. Arnek, L.G. Sillén and O. Wahlberg, Ark. Kemi, 31 (1970) 353.
28 T. Shimanouchi and I. Suzuki, J. Chem. Phys., 42 (1965) 296.
29 R. Hooke and T.A. Jeeves, J. Assoc. Comput. Machinery, 8 (1961) 212.
30 M. Bell and M.C. Pike, Commun. Assoc. Comput. Machinery, 9 (1966) 684.
31 P.J. Burchill and J.A. McRae, Aust. J. Chem., 24 (1971) 187.
32 J.G. Gaultieri and G.P. de Lhery, J. Chem. Phys., 55 (1971) 1541.
33 A.C. Breti and T.W. Dingle, Theor. Chim. Acta, 24 (1972) 111.
34 D.M. Bishop, J.-C. Leclerc and Y. Takahata, J. Mol. Spectrosc., 39 (1971) 79.
35 D.M. Bishop and A.-J.A.Wu, J. Chem. Phys., 54 (1971) 2917.
36 R.G. Clark and E.T. Stewart, Mol. Phys., 22 (1971) 341.
37 D. Ditchfield, W.J. Here and J.A. Pople, J. Chem. Phys., 54 (1971) 724, 5001.
38 H.H. Rosenbrock, Comput. J., 3 (1960) 175.
39 R.C. Dobson, D.M. Hayes and R. Hoffmann, J. Amer. Chem. Soc., 93 (1971) 6188
40 L. Schuffenecker, D. Balesdent and J. Houriez, Bull. Soc. Chim. Fr., (1970) 1239.
41 W.H. Swann, Report on the Development of a New Direct Search Method of Optimisation, I.C.I. Ltd., Central Instrument Laboratory Research Note 64/3, 1964.
42 R.N.F. Thornley, A.G. Sykes and P. Gans, J. Chem. Soc. A, (1971) 1494.
43 J.A. Ramsay, J.A. Ladd and W.J. Orville-Thomas, J. Chem. Soc. Faraday I, 68 (1972) 193.
44 G.J. Ford, L.D. Pettit and C. Sherrington, J. Inorg. Nucl. Chem., 33 (1971) 4119.
45 D.S. Barnes, G.J. Ford, L.D. Pettit and C. Sherrington, J. Chem. Soc. A, (1971) 2883.
46 G. Beech and K. Miller, J. Chem. Soc. Dalton, (1972) 861.
47 J.C. Ruckman, R.T.W. Morrison and R.H. Buck, J. Chem. Soc. Dalton, (1972) 426.

48 D.W. Calvin and T.M. Reed, J. Chem. Phys., 54 (1971) 3733.
49 K. O-ohata, H. Taketa and S. Huzinaga, J. Phys. Soc. Japan, 21 (1966) 2306.
50 J.A. Dye and V.A. Nicely, J. Chem. Educ., 48 (1971) 443.
51 I.D. Faux and A.B. Lidiard, Z. Naturforsch. A, 26 (1971) 62.
52 M.J. Norgett and R. Fletcher, J. Phys. C, 3 (1970) L190.
53 J.W. McIver and A. Komornicki, Chem. Phys. Lett., 10 (1971) 303.
54 D.F. Brailsford and B. Ford, Chem. Phys. Lett., 12 (1971) 60.
55 P. Th. van Duijnen and G.B. Cook, Mol. Phys., 21 (1971) 475; 22 (1972) 637.
56 W. Spendley, G.R. Hext and F.R. Himsworth, Technometrics, 4 (1962) 441.
57 J.A. Nelder and R. Mead, Comput. J., 7 (1965) 308; R. O'Neill, Appl. Stat., 20 (1971) 338.
58 R. Aveyard and B.J. Briscoe, J. Chem. Soc. Faraday I, 68 (1972) 478; Trans. Faraday Soc., 66 (1970) 2911.
59 R.L. D'Arcy and K. Watt, Trans. Faraday Soc., 66 (1970) 1236.
60 L.D. Kispert and M.T. Rogers, J. Chem. Phys., 54 (1971) 3326.
61 D.F. Ewing, B. Hudson, D.E. Webster and P.B. Wells, J. Chem. Soc. Dalton, (1972) 1287.
62 N. Bodor and M.J.S. Dewar, J. Amer. Chem. Soc., 93 (1971) 6685; M.J.S. Dewar and D.H. Lo, J. Amer. Chem. Soc., 93 (1971) 7201.
63 E. Haselbach, Helv. Chim. Acta, 54 (1971) 2257.
64 S.K. Gangwal, R.R. Hudgins, A.W. Bryson and P.L. Silveston, Can. J. Chem. Eng., 49 (1971) 113.
65 F. Cajori, History of Mathematics, Macmillan, New York, 1919.
66 W.C. Hamilton, Statistics in Physical Science, Ronald Press, New York, 1964.
67 J.B. Scarborough, Numerical Mathematical Analysis, 3rd edn., John Hopkins, Baltimore, 1955.
68 W.E. Deming, Statistical Adjustment of Data, Wiley, New York, 1943.
69 R.N. Roy, N. Roy, A.L.M. Bothwell, J. Gibbons and W. Vernon, J. Chem. Soc. Dalton, (1972) 530.
70 P.E.M. Allen, A.E. Byers and R.M. Lough, J. Chem. Soc. Dalton, (1972) 479.
71 P. Quittner, Anal. Chem., 41 (1969) 1504.
72 D.C. Robinson, Nucl. Instrum. Meth., 78 (1969) 120.
73 A.M. Bond, Coord. Chem. Rev., 6 (1971) 377.
74 G.E. Forsythe and C.B. Moler, Computer Solution of Linear Algebraic Systems, Prentice-Hall, Englewood Cliffs, N.J., 1967.
75 P.K. Karmalkar, Z. Phys. Chem., 218 (1961) 189.
76 T. Groth, Nucl. Instrum. Meth., 82 (1970) 93.
77 A.D. Jones and D.R. Williams, J. Chem. Soc. A, (1970) 3138.
78 J.A. Blackburn (Ed.), Spectral Analysis: Methods and Techniques, Dekker, New York, 1970.
79 D.L. Singleton and F.E. Stafford, Inorg. Chem., 11 (1972) 1209.
80 W.H. Lawton and E.A. Sylvestre, Technometrics, 13 (1971) 617.
81 P. Kruger, Principles of Activation Analysis, Wiley, New York, 1971.
82 N.D. Eckhoff, Nucl. Instrum. Meth., 74 (1969) 77.
83 C.R. Crawford and J.D. Morrison, Anal. Chem., 40 (1968) 1464.
84 G.L. Cumming, J.S. Rollet, F.J.C. Rossotti and R.J. Whewell, J. Chem. Soc. Dalton, (1972) 2652.
85 A. Savitzky and M.J.E. Golay, Anal. Chem., 36 (1964) 1627.
86 J.G. Hayes (Ed.), Numerical Approximations to Functions and Data, Athlone Press, London 1970.
87 W.H. Lawton, E.A. Sylvestre and M.S. Maggio, Technometrics, 14 (1972) 513.
88 T.R. Dickson, The Computer in Chemistry, W.H. Freeman, San Francisco, 1968.
89 G. Golub, Numer. Math., 7 (1965) 206.
90 O. Chacón and P. Matzke, J. Mol. Struct., 9 (1971) 243.

91 J. Aldous and I.M. Mills, Spectrochim. Acta, 18 (1962) 1073.
92 V.J. Law and R.V. Bailey, Chem. Eng. Sci., 18 (1963) 189.
93 S.W. Marshall, J.A. Nelson, R.M. Wilenzuk, Commun. Assoc. Comput. Machinery, 8 (1965) 313.
94 B.J. Duke and T.C. Gibb, J. Chem. Soc. A, (1967) 1478.
95 A.H. Anderson, T.C. Gibb and A.B. Littlewood, Anal. Chem., 42 (1970) 434.
96 A.B. Littlewood, T.C. Gibb and A.H. Anderson, in C.L.A. Harbourne (Ed.), Gas Chromatography 1968, The Institute of Petroleum, London, 1969.
97 L.M. Schwartz, Anal. Chem., 43 (1971) 1336.
98 A.J. Bard and D.M. King, J. Chem. Educ., 42 (1965) 127.
99 A. Sabatini and A. Vacca, J. Chem. Soc. Dalton, (1972) 1693.
100 H.O. Hartley, Technometrics, 3 (1961) 269.
101 K. Levenberg, Quart. Appl. Math., 2 (1944) 164.
102 A. Girard, Rev. Opt., 37 (1958) 225, 397.
103 C.G. Wynne, Proc. Phys. Soc., London, 73 (1959) 777.
104 D.D. Morrison, JPL Seminar Proceedings, 1960.
105 D.W. Marquart, J. Soc. Ind. Appl. Math., 11 (1963) 431.
106 R. Fletcher, A.E.R.E. Rep., No. R-6799, 1971.
107 R.R. Meyer and P.M. Roth, J. Inst. Math. Appl., 9 (1972) 218.
108 D.M. Adams and R.G. Churchill, J. Chem. Soc. A (1970) 697.
109 D. Papoušek, S. Toman and J. Plíva, J. Mol. Spectrosc., 15 (1965) 502.
110 J.M. Freeman and T. Henshall, J. Mol. Spectrosc., 25 (1968) 101.
111 F.J. Boerio and J.L. Koenig, J. Chem. Phys., 52 (1970) 4826.
112 V. Spirko and J. Morávek, J. Mol. Spectrosc., 33 (1970) 368.
113 J. Plíva, V. Spirko and S. Toman, J. Mol. Spectrosc., 21 (1966) 106.
114 Z. Cihla and A. Chedin, J. Mol. Spectrosc., 40 (1971) 337.
115 R.A. Pinkowitz and T.J. Swift, J. Chem. Phys., 54 (1971) 2858.
116 S.M. Roberts, D.H. Wilkinson and L.R. Walker, Anal. Chem., 42 (1970) 886.
117 L.H. Scharpen, J.S. Muenter and V.W. Laurie, J. Chem. Phys., 53 (1970) 2513.
118 R.S. Freund, T.A. Miller, D. de Santis and A. Lurio, J. Chem. Phys., 53 (1970) 2290.
119 T.A. Miller, J. Chem. Phys., 54 (1971) 330.
120 P.C. Kelly and W.R. Harris, Anal. Chem.. 43 (1971) 1184.
121 J.W. Roddy and C.F. Coleman, J. Inorg. Nucl. Chem., 32 (1970) 3956.
122 R.Y. Koo and H.W. Hsu, J. Chem. Phys., 52 (1970) 2392.
123 H.M. Lin and R.L. Robinson, J. Chem. Phys., 54 (1971) 52.
124 I.T. Steinberger, C. Atluri and O. Schnepp, J. Chem. Phys., 52 (1970) 2723.
125 W.J. Jones and B. Musulin, J. Mol. Spectrosc., 40 (1971) 424.
126 J. Prasad and N.C. Peterson, Inorg. Chem., 10 (1971) 88.
127 R. Klingbeil, J. Chem. Phys., 56 (1972) 132.
128 W. Davis, J. Mrochek and P.R. Judkins, J. Inorg. Nucl. Chem., 32 (1970) 1689.
129 J.W. Roddy, C.F. Coleman and S. Arai, J. Inorg. Nucl. Chem., 33 (1971) 1099.
130 N.O. Lemcoff and R.E. Cunningham, J. Catal., 23 (1971) 81.
131 C. Norris and F.R. Nordmeyer, J. Amer. Chem. Soc., 93 (1971) 4044.
132 I.G. Sayce, Talanta, 15 (1968) 1397; 18 (1971) 653; 19 (1972) 831.
133 R.S. Tobias and M. Yasuda, Inorg. Chem., 2 (1963) 1307.
134 C.W. Childs and D.D. Perrin, J. Chem. Soc. A, (1969) 1039.
135 L.J. Zompa, Inorg. Chem., 10 (1971) 2647.
136 M.J.D. Powell, Comput. J., 7 (1965) 303.
137 H. Rinneberg and J.A. Weil, J. Chem. Phys., 56 (1972) 2012.
138 M.J. Cooper, Acta Crystallogr., Sect. A, 26 (1970) 208.
139 P.E. Siska, J.M. Parson, T.P. Schafer and Y.T. Lee, J. Chem. Phys., 55 (1971) 5762.
140 G. Peckam, Comput. J., 13 (1970) 418.
141 M.J.D. Powell, A.E.R.E. Rep., No. R-5947, 1968.

142 W. Murray (Ed.), Numerical Methods for Unconstrained Optimisation, Academic Press, London, 1972.
143 W.E. Wentworth, J. Chem. Educ., 42 (1965) 96, 162.
144 J. Lingane and Z.Z. Hugus, Inorg. Chem., 9 (1970) 757.
145 D.R. Powell and J.R. MacDonald, Comput. J., 15 (1972) 148.
146 A. Jones, Comput. J., 13 (1970) 301.
147 B. Gellai and G. Jancso, J. Mol. Struct., 12 (1972) 478.
148 R.P. Tewarson, Comput. J., 12 (1969) 388.
149 G. Peters and J.H. Wilkinson, Comput. J., 13 (1970) 309.
150 L.I. Hodgson and J.S. Rollet, Acta Crystallogr., 6 (1963) 329.
151 J.J. Christensen, J. Ruckman, D.J. Eatough and R.M. Izatt, Thermochim. Acta, 3 (1972) 203; D.J. Eatough, J.J. Christensen and R.M. Izatt, Thermochim. Acta, 3 (1972) 219; D.J. Eatough, R.M. Izatt and J.J. Christensen, Thermochim. Acta, 3 (1972) 233.
152 W.H. Lawton and E.A. Sylvestre, Technometrics, 13 (1971) 461; see also pp. 477 and 480 for discussion.
153 R. Fletcher, Mol. Phys., 19 (1970) 55.
153 (a) G.H. Golub and V. Pereyra, Soc. Ind. Appl. Maths. J. Numer. Anal., 10 (1973) 413.
154 P.L. Orioli and A. Vacca, Acta Crystallogr., Sect. A, 29 (1973) 89.
155 T.G. Strand, D.A. Kohl and R.A. Bonham, J. Chem. Phys., 39 (1963) 1307.
156 M.O. Faltens and D.A. Shirley, J. Chem. Phys., 53 (1970) 4249.
157 C.L. Gardner and M.A. Weinberger, Can. J. Chem., 48 (1970) 1317.
158 J.T. Routti and S.G. Prussin, Nucl. Instrum. Meth., 72 (1969) 125.
159 E.W. Baumann, J. Inorg. Nucl. Chem., 31 (1969) 3155.
160 M. Wells, Commun. Assoc. Comput. Machinery, 8 (1965) 169; R. Fletcher, Commun. Assoc. Comput. Machinery, 9 (1966) 686; P.A. Hamilton and J. Boothroyd, Commun. Assoc. Comput. Machinery, 12 (1969) 512.
161 Y. Bard, Math. Comput., 22 (1968) 665.
162 M.J.D. Powell, J. Inst. Math. Appl., 7 (1971) 21.
163 J.E. Bailey and F. Horn, Ber. Bunsenges., 73 (1969) 274.
164 J.R. Hullett and W. Wilson, J. Chem. Soc. A, (1968) 2569.
165 R.D. Brown, F.R. Burden and B.T. Hart, Theor. Chim. Acta, 22 (1971) 214.
166 R. Kari and B.I. Sutcliffe, Chem. Phys. Lett., 7 (1970) 149.
167 G.D. Zeiss and M.A. Whitehead, J. Chem. Soc. A, (1971) 1727.
168 T.A. Claxton and N.A. Smith, Theor. Chim. Acta, 22 (1971) 399.
169 R. Fletcher, Comput. J., 13 (1970) 317.
170 R. Fletcher, A.E.R.E. Rep., No. R-7125, 1972.
171 B.A. Murtagh and R.W.H. Sargent, Comput. J., 13 (1970) 185.
172 M.C. Biggs, J. Inst. Math. Appl., 8 (1971) 315.
173 P.E. Gill and W. Murray, J. Inst. Math. Appl., 9 (1972) 91.
174 P.E. Gill, W. Murray and R.A. Pitfield, N.P.L. Rep., No. NAC-11, 1972.
175 J. Greenstadt, Math. Comput., 21 (1967) 360.
176 P.D. Dacre and M. Elder, Mol. Phys., 22 (1971) 593.
177 S.M. Goldfield, R.E. Quant and H.F. Trotter, Econometrica, 34 (1966) 541.
178 N.C. Pyper, personal communication.
179 G.W. Stewart, J. Assoc. Comput. Machinery, 14 (1967) 72.
180 S.A. Lill, Comput. J., 13 (1970) 111; 14 (1971) 106.
181 G.J. Ford, P. Gans, L.D. Pettit and C. Sherrington, J. Chem. Soc. Dalton, (1972) 1763.
182 P. Gans and A. Vacca, Talanta, 21 (1974) 45.
183 D. Dryssen, N. Ingri and L.G. Sillén, Acta Chem. Scand., 15 (1961) 694.
184 N. Ingri and L.G. Sillén, Ark. Kemi, 23 (1964) 47.
185 D.R. Williams, J. Chem. Soc. Dalton, (1972) 790.
186 L.G. Sillén and B. Warnqvist, Ark. Kemi, 31 (1970) 341.
187 S. Forsen, S. Okermark and T. Alm, Acta Chem. Scand., 18 (1964) 2313.

188 L. Ödberg, A. Lofvenberg, E. Högfeld and F. Fredlund, J. Inorg. Nucl. Chem., 34 (1972) 2065.
189 R. Barbucci, L. Fabrizzi and P. Paoletti, J. Chem. Soc. Dalton, (1972) 1099.
190 L.G. Sillén and B. Warnqvist, Ark. Kemi, 31 (1970) 377.
191 P. Paoletti, A. Vacca and D. Arenare, J. Phys. Chem., 70 (1966) 193; Coord. Chem. Rev., 1 (1966) 280.
192 R. Fletcher and C.M. Reeves, Comput. J., 7 (1964) 149.
193 D.H. Sleeman, Theor. Chim. Acta, 11 (1968) 135.
194 S. Margulies, Rev. Sci. Instrum., 39 (1968) 478.
195 N. Ingri, W. Kalotowicz, L.G. Sillén and B. Warnqvist, Talanta, 14 (1967) 1261.
196 D.D. Perrin and I.G. Sayce, Talanta, 14 (1967) 833.
197 P. Rabinowitz, Numerical Methods for Non-linear Algebraic Equations, Gordon and Breach, London, 1970.
198 M.J.D. Powell, A.E.R.E. Rep., No. T.P. 439, 1971.
199 C.G. Broyden, Math. Comput., 19 (1965) 577.
200 P. Gans, J. Chem. Soc. A, (1971) 2017.
201 P. Gans, J. Mol. Struct., 12 (1972) 411.
202 Y. Bard, Soc. Ind. Appl. Math. J. Numer. Anal., 7 (1970) 157.
203 J. Pitha and R.N. Jones, Can. J. Chem., 44 (1966) 3031.
204 V. Gramlich and W.M. Meier, Z. Kristallogr., 133 (1971) 134.
205 W.S. McDonald and D.W.J. Cruickshank, Acta Crystallogr., Sect. B, 27 (1971) 1315.
206 J. Donohue and S.H. Goodman, Acta Crystallogr., 22 (1967) 952.
207 P.E.J. Arnold, E.A.V. Ebsworth and H.F. Jessop, J. Chem. Soc. Dalton, (1972) 1681.
208 P.N. Schatz, A.J. McCaffery, W. Suetka, G.N. Henning, A.B. Ritchie and P.J. Stephens, J. Chem. Phys., 45 (1966) 722.
209 D.D. Perrin and I.G. Sayce, J. Chem. Soc. A, (1967) 82.
210 D.R. Rosseinsky and H. Kellawi, J. Chem. Soc. A, (1969) 1207.
211 D.W. Marquart, R.G. Bennett and E.J. Burrel, J. Mol. Spectrosc., 17 (1961) 269.
212 R. Fletcher, A.E.R.E. Rep., No. TP-432, 1970.
213 D.C. Robinson, Nucl. Instrum. Meth., 79 (1970) 65.
214 J. Aldous and I.M. Mills, Spectrochim. Acta, 19 (1963) 1567, 1585.
215 R.C. Fergusen and D.W. Marquart, J. Chem. Phys., 41 (1964) 2087.
216 J. Waser, Acta Crystallogr., 16 (1963) 1091.
217 D.G. Luenberger and V.E. Dennis, Anal. Chem., 38 (1966) 715.
218 L.G. Sillén, Acta Chem. Scand., 18 (1964) 1085.
219 K. Momoki, H. Sato and H. Ogawa, Anal. Chem., 39 (1967) 1072.
220 K. Hedberg and M. Iwasaki, Acta Crystallogr., 17 (1964) 529; M. Iwasaki, F.N. Futch and K. Hedberg, Acta Crystallogr., 17 (1964) 538.
221 H.P. Yule, Anal. Chem., 40 (1968) 1480.
222 A. Savitzky, Anal. Chem., 33 (1961) 25A.
223 J.R. Morrey, Anal. Chem., 40 (1968) 905.
224 W.L. Butler and D.W. Hopkins, Photochem. Photobiol., 12 (1970) 439, 451.
225 T.R. Griffiths and P.J. Potts, personal communication.
226 M.A. Hogan, S. Yamamoto and D.F. Corell, Nucl. Instrum. Meth., 80 (1970) 61.
227 L.P. Varga, W.D. Wakley, L.S. Nicolson, M.L. Madden and J. Patterson, Anal. Chem., 37 (1965) 1003.
228 W.D. Keller, T.R. Lusebrink and C.H. Sederholm, J. Chem. Phys., 44 (1966) 782.
229 L. Kuča and E. Högfeld, Acta Chem. Scand., 22 (1968) 183.
230 S. Hietanen and L.G. Sillén, Acta Chem. Scand., 22 (1968) 265.
231 R.M. Wallace and S.M. Katz, J. Phys. Chem., 68 (1964) 3890.
232 R.N. Jones, Appl. Opt., 8 (1969) 597.