A Damped Non-linear Least-squares Computer Program (DALSFEK) for the Evaluation of Equilibrium Constants from Spectrophotometric and Potentiometric Data

By **Roland M. Alcock, Frank R. Hartley,***† and **David E. Rogers,*** Department of Chemistry, The University, Southampton SO9 5NH

A computer program (DALSFEK) has been developed which uses a damped non-linear least-squares iterative method to fit equilibrium constants to spectrophotometric and potentiometric data. The program has been designed in such a way that subroutines that can deal with other types of response variable can easily be added, and the program can utilise several types of observed data simultaneously. An efficient least-squares algorithm due to Marquardt, which combines the method of steepest descent with methods based on the Taylor series in such a way as to exploit the advantages of each, ensures smooth convergence to minimum-variance estimates of the parameters. Facilities are incorporated which allow adjustments to be made to allow for possible sources of systematic errors. The program has been tested on a number of systems and found to be very convenient to use as well as giving satisfactory results.

EQUILIBRIUM constants cannot be measured directly but must be calculated from an observed response function of a fixed (but experimentally adjustable) variable. Quite often the complexity of the system that can be studied experimentally, and the resulting amount of information that can be obtained, is very dependent on the nature of the calculation procedure and the convenience of data handling. Before the widespread availability of electronic computers, most calculation procedures manipulated the data into a form suitable for linear graphical representation.[1] However, only relatively simple systems can be treated in such a linear fashion; for more complex systems involving the presence of three or more different complex species in solution it is first virtually impossible to produce a linear method of data treatment and secondly generally undesirable because of the approximations (and hence errors) inherent in such a treatment.[1] The main limitations of linear (graphical) procedures for the determination of equilibrium constants are: (*i*) it is difficult to obtain reliable estimates of the errors in the parameters; (*ii*) it may be unclear whether the parameters give a reasonable fit to the experimental observables, thus testing different possible models presents difficulties; (*iii*) in some complicated cases only a portion of the data may be used in evaluating a particular parameter (*e.g.* by extrapolation to some point), and the precision of the parameter becomes limited; and (*iv*) considerable effort may be expended in arriving at justifiable assumptions for manipulating the equations into a form that can conveniently be plotted.

---

† *Present address:* Department of Chemistry and Metallurgy, The Royal Military College of Science, Shrivenham, Swindon, Wiltshire SN6 8LA.

[1] F. J. C. Rossotti and H. S. Rossotti, 'The Determination of Stability Constants,' McGraw-Hill, New York, 1961.

The advent of computers has brought non-linear data treatments within the grasp of most research workers. Furthermore, since computer evaluation of either a linear or non-linear system involves preparation of the input data for the program, there is no justification for using a linear treatment in situations in which it is not rigorously applicable. A considerable number of non-linear computer-based methods for the evaluation of equilibrium constants have been described.[2-11] These can very crudely be divided into two groups: (*a*) programs designed for the analysis of one type of data (*e.g.* spectrophotometric only); (*b*) programs designed for the analysis of several types of data. We have already commented on the advantages in terms of reliability of using data obtained from more than one type of physical technique.[12] Computer programs that are suitable for use with several types of data are dominated by Sillén's LETAGROP VRID.[13] However, this is such a large comprehensive program that not only is familiarisation a mighty task but also the program requires a very lengthy input deck for each run. Furthermore, not everyone has access to a computer with sufficient store. The present program was conceived as one that could accept data from as many physical techniques as possible, whilst retaining simplicity of operation.

Background to the Program.—Matrix approach to data handling. In order to facilitate the handling of large amounts of data, a matrix approach based on that of Hamilton[14] was used. Consider a situation in which there are n experimental observations $O_1, O_2, \ldots O_n$, each having an associated random error, $e_1, e_2, \ldots e_n$, and each depending on m unknowns $x_1, x_2, \ldots x_m$. Then we can write equation (1). If the initial estimates of the

$$O_i = f_i(x_1 x_2 \ldots x_m) + e_i \qquad (1)$$

parameter values are $x_1^0, x_2^0 \ldots x_m^0$, then the observables are expressed about the initial point in parameter space according to the Taylor series expansion[15] by (2), and hence (3) in which terms higher than first order

$$O_i = f_i(x_1^0 x_2^0 \ldots x_m^0) + \left(\frac{\partial f_i}{\partial x_i}\right)_0 (x_1 - x_1^0) + \ldots \left(\frac{\partial f_i}{\partial x_m}\right)_0 (x_m - x_m^0) + e_i \qquad (2)$$

$$O_i = f_i(x_1^0 x_2^0 \ldots x_m^0) + \sum_{j=1}^{m} \left(\frac{\partial f_i}{\partial x_j}\right)_0 \Delta x_j + e_i \qquad (3)$$

have been neglected. The change (ΔO_i) in the observable O_i consequent upon making the corrections Δx_j are (4) which may be expressed in matrix notation as

$$\Delta O_i = O_i - f_i(x_1^0 x_2^0 \ldots x_m^0) = \sum_{j=1}^{m} \left(\frac{\partial f_i}{\partial x_j}\right)_0 \Delta x_j + e_i \qquad (4)$$

(5) * in which the problem has been linearised with O,

$$\boldsymbol{O} = \boldsymbol{B}X + E \qquad (5)$$

$\boldsymbol{B}$, and X containing the elements ΔO_i, $(\partial f_i/\partial x_j)_0$, and Δx_j respectively. If a vector $\vec{X}$ represents the 'best-fit' values of X to the regression equation, then a vector of residuals (V) can be defined as in (6). The quantity to

$$V = O - \boldsymbol{B}\vec{X} \qquad (6)$$

be minimised is the sum of the squares of the residuals (S) given by (7) where $\widetilde{V}$ is the transpose of V. If the

$$S = \widetilde{V}V \qquad (7)$$

variances of the residuals are not equal, a weighting matrix can be introduced so that we can write (8) where

$$S = \widetilde{V}\boldsymbol{W}V \qquad (8)$$

$\boldsymbol{W}$ is the inverse of the moment matrix $\boldsymbol{M}$ which contains the variances of the residuals σ_i^2 as diagonal elements and the covariances σ_{ij} as off-diagonal elements. Although the absolute values of these quantities are generally unknown, they are usually known to within a constant scale factor, σ^2, the variance of an observation of unit weight. Hence, it is only necessary to know the relative magnitudes of σ_i^2 and σ_{ij}, and σ_{ij} are usually arranged to be zero, making $\boldsymbol{W}$ a diagonal matrix.

Substituting (6) into (8), we obtain (9) since equation (10) applies. Then we can write (11) and (12), and

$$S = \widetilde{(O - \boldsymbol{B}\vec{X})}\,\boldsymbol{W}(O - \boldsymbol{B}\vec{X}) \qquad (9)$$

$$\widetilde{(\boldsymbol{B}\vec{X})} = \widetilde{\vec{X}}\widetilde{\boldsymbol{B}} \qquad (10)$$

(13) follows. Differentiating with respect to each of the

$$S = \widetilde{O}\boldsymbol{W}O + \widetilde{\vec{X}}\widetilde{\boldsymbol{B}}\boldsymbol{W}\boldsymbol{B}\vec{X} - \widetilde{O}\boldsymbol{W}\boldsymbol{B}\vec{X} - \widetilde{\vec{X}}\widetilde{\boldsymbol{B}}\boldsymbol{W}O \qquad (11)$$

$$\widetilde{O}\boldsymbol{W}\boldsymbol{B}\vec{X} = O\boldsymbol{W}\widetilde{\boldsymbol{B}}\widetilde{\vec{X}} \qquad (12)$$

$$S = \widetilde{O}\boldsymbol{W}O + \widetilde{\vec{X}}\widetilde{\boldsymbol{B}}\boldsymbol{W}\boldsymbol{B}\vec{X} - 2\widetilde{\vec{X}}\widetilde{\boldsymbol{B}}\boldsymbol{W}O \qquad (13)$$

---

* Italic upper case letters represent row or column vectors and italic bold upper case letters are two-dimensional matrices containing elements in two or more columns and rows.

[2] C. W. Childs, P. S. Hallman, and D. D. Perrin, *Talanta*, 1969, **16**, 1119 and refs. therein.

[3] F. J. C. Rossotti, H. S. Rossotti, and R. J. Whewell, *J. Inorg. Nuclear Chem.*, 1971, **33**, 2051 and refs. therein.

[4] P. Gans and H. M. N. H. Irving, *J. Inorg. Nuclear Chem.*, 1972, **34**, 1885.

[5] T. O. Maier and R. S. Drago, *Inorg. Chem.*, 1972, **11**, 1861.

[6] A. Vacca, A. Sabatini, and M. A. Gristina, *Co-ordination Chem. Rev.*, 1972, **8**, 5.

[7] A. Sabatini and A. Vacca, *J.C.S. Dalton*, 1972, 1693.

[8] S. Feldberg, P. Klotz, and L. Newman, *Inorg. Chem.*, 1972, **11**, 2860 and refs. therein.

[9] D. J. Leggett and W. A. E. McBryde, *Analyt. Chem.*, 1975, **47**, 1065.

[10] A. Sabatini, A. Vacca, and P. Gans, *Talanta*, 1974, **21**, 53.

[11] P. Gans, A. Sabatini, and A. Vacca, *Inorg. Chim. Acta*, 1976, **18**, 237.

[12] R. M. Alcock, F. R. Hartley, D. E. Rogers, and J. L. Wagner, *Co-ordination Chem. Rev.*, 1975, **16**, 59.

[13] L. G. Sillén, *Acta Chem. Scand.*, 1964, **18**, 1085.

[14] W. C. Hamilton, 'Statistics in Physical Science,' Ronald Press, New York, 1964.

[15] I. G. Betteley, C. J. Brookes, and S. M. Loxston, 'Mathematics and Statistics for Chemists,' Wiley, New York, 1966.

parameters $\vec{X}$ in turn, and equating to zero we obtain (14) and (15) and hence (16). In a linear situation $\vec{X}$

$$\partial S = 2(\widetilde{\partial \vec{X}})(\widetilde{\boldsymbol{B}}\boldsymbol{W}\boldsymbol{B}\vec{X} - \widetilde{\boldsymbol{B}}\boldsymbol{W}O) \qquad (14)$$

$$(\widetilde{\boldsymbol{B}}\boldsymbol{W}\boldsymbol{B})\vec{X} = \widetilde{\boldsymbol{B}}\boldsymbol{W}O \qquad (15)$$

$$\vec{X} = (\widetilde{\boldsymbol{B}}\boldsymbol{W}\boldsymbol{B})^{-1}\widetilde{\boldsymbol{B}}\boldsymbol{W}O \qquad (16)$$

would give the least-squares estimates of the parameters in one iteration, but in a non-linear case, such as the present, $\vec{X}$ merely gives the correction vector ΔX which (theoretically) produces a lower sum of squares of the errors. This is because the Taylor expansion is not strictly valid, since $\partial f_i/\partial x_j$ varies with x_j. In this way, a vector of improved values of the parameters is generated which is used as the estimates for the next calculation, so that, if the function is well behaved and the starting vector X^0 not too far from the final values, the process will converge to the least-squares estimates in a finite number of iterations.

Estimates of the standard deviations of the parameters $\vec{X}$ can be obtained from the variance–covariance matrix of the parameters[16] $\boldsymbol{M}_x$, defined as in (17) where $\varepsilon[f(x)]$ is a

$$\boldsymbol{M}_x = \varepsilon[(\vec{X} - X)(\vec{X} - X)^{\sim}] \qquad (17)$$

statistical concept known as the ' expected value ' of the function. Therefore if the equality (18) applies, we obtain (19) and (20). But $\boldsymbol{W}$ is the inverse of the

$$\boldsymbol{A} = \widetilde{\boldsymbol{B}}\boldsymbol{W}\boldsymbol{B} \qquad (18)$$

$$\boldsymbol{M}_x = \varepsilon[\boldsymbol{A}^{-1}\widetilde{\boldsymbol{B}}\boldsymbol{W}(O - \boldsymbol{B}X)(O - \boldsymbol{B}X)^{\sim}\boldsymbol{W}\boldsymbol{B}\boldsymbol{A}^{-1}] \qquad (19)$$

$$= \boldsymbol{A}^{-1}\widetilde{\boldsymbol{B}}\boldsymbol{W}\varepsilon[(O - \boldsymbol{B}X)(O - \boldsymbol{B}X)^{\sim}]\boldsymbol{W}\boldsymbol{B}\boldsymbol{A}^{-1} \qquad (20)$$

matrix of the variances of the residuals so that we can write (21) and hence (22). As already noted, the weight

$$\boldsymbol{W}^{-1} = \varepsilon[(O - \boldsymbol{B}X)(O - \boldsymbol{B}X)^{\sim}] \qquad (21)$$

$$\boldsymbol{M}_x = \boldsymbol{A}^{-1}\widetilde{\boldsymbol{B}}\boldsymbol{W}\boldsymbol{W}^{-1}\boldsymbol{W}\boldsymbol{B}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1} = (\widetilde{\boldsymbol{B}}\boldsymbol{W}\boldsymbol{B})^{-1} \qquad (22)$$

matrix is usually only known to within a scale factor of σ^{-2}, so that $\boldsymbol{M}_x$ is given by (23) where $\boldsymbol{W}_0$ contains the

$$\boldsymbol{M}_x = \sigma^2(\widetilde{\boldsymbol{B}}\boldsymbol{W}_0\boldsymbol{B})^{-1} \qquad (23)$$

relative magnitudes of the weights but not their absolute values. An unbiased estimate of σ^2 is obtained from (24) where $n - m$ is the number of degrees of freedom of σ^2 (n = number of observations, m = number of parameters determined).

$$\sigma^2 = \widetilde{\boldsymbol{V}}\boldsymbol{W}_0\boldsymbol{V}/(n - m) \qquad (24)$$

Minimisation of a non-linear function. The most commonly used procedure in the field of equilibrium constants[5,17–23] for the non-linear least-squares estimation of a set of parameters is the Gauss–Newton method,[24] in which the function to be fitted is expanded as a Taylor series truncated at the first-order term. Occasionally, the second-order terms have been taken into account (the Newton–Raphesón approach[24]) in an effort to obtain more efficient convergence properties, but there are conflicting reports[6,7] regarding the efficacy in so doing.

An alternative method is the direct-search[25] or ' pit-mapping ' approach of Sillén and his coworkers,[13,26–29] in which the error-square sum S for an initial set of parameters X is calculated and then S is recalculated with each individual x_i in turn changed by a given h_i. From the S values for $\frac{1}{2}(n + 1)(n + 2)$ systematically chosen points (n = number of equilibrium constants being adjusted), the coefficients of a second-degree surface are found and hence the minimum X_0 of the surface evaluated. This set of values is then used for the start of the next iteration.

The third method,[8,30] known as the steepest-descent method,[31] is conceptually the simplest to envisage. The parameters are adjusted in proportion to the magnitude of the derivatives evaluated at the particular point in parameter space currently occupied. The direction of movement is always ensured to be down the ' slope '. However, the major disadvantage of this type of method is that, after initial rapid progress, further minimisation is painfully slow.[32,33] This is only to be expected, since as the minimum is approached the magnitudes of the derivatives must become much smaller.

Each of these methods has its merits and each is claimed[7,8,27] to have specific advantages over the others, but the main problem encountered by all of them is the reliability of convergence. In order to obtain the corrections with which to calculate better estimates of the parameters, X, so as to be able to commence the next iteration of the Gauss–Newton refinement, equation (16) must be solved. This necessitates inverting the matrix $\widetilde{\boldsymbol{B}}\boldsymbol{W}\boldsymbol{B}$, which can only be achieved if the determinant is not zero. (If it is zero, the matrix is singular and has no inverse, in which case the problem would have no unique

---

[16] Ref. 14, ch. 4, p. 129.

[17] P. J. Lingane and Z. Z. Hugus, *Inorg. Chem.*, 1970, **9**, 757.

[18] R. S. Tobias and M. Yasuda, *Inorg. Chem.*, 1963, **2**, 1307.

[19] W. E. Wentworth, W. Hirsch, and E. Chen, *J. Phys. Chem.*, 1967, **71**, 218.

[20] D. Inman, I. Regan, and B. Girling, *J. Chem. Soc.*, 1964, 349.

[21] J. A. Chopoorian, G. R. Choppin, H. C. Griffith, and R. Chandler, *J. Inorg. Nuclear Chem.*, 1961, **21**, 21.

[22] I. G. Sayce, *Talanta*, 1968, **15**, 1397.

[23] D. D. Perrin and I. G. Sayce, *Talanta*, 1967, **14**, 833.

[24] J. B. Scarborough, ' Numerical Mathematical Analysis,' John Hopkins Press, Baltimore, 1930, p. 187.

[25] J. J. Kinkare, *Analyt. Chem.*, 1970, **42**, 1322.

[26] L. G. Sillén, *Acta Chem. Scand.*, 1962, **16**, 159.

[27] L. G. Sillén, *Acta Chem. Scand.*, 1962, **16**, 173.

[28] L. G. Sillén, *Pure Appl. Chem.*, 1968, **17**, 55.

[29] L. G. Sillén and B. Warnqvist, *Arkiv. Kemi*, 1969, **31**, 315.

[30] L. R. Lieto and C. H. Liu, Report ORNL-TM-2714, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1969; Dissertation, Arizona State University, 1969.

[31] H. A. Spang, *J. Soc. Ind. Appl. Maths. Rev.*, 1962, **4**, 343.

[32] C. G. Wynne, *Proc. Phys. Soc.*, 1959, **73**, 777.

[33] N. R. Draper and H. Smith, 'Applied Regression Analysis,' Wiley, New York, 1966, ch. 10.

solution and the parameters would be indeterminate.) If the determinant, although not zero, is very small the matrix is said to be ill conditioned, a problem that has often been found previously in equilibrium-constant calculations.[34-36]

The physical effect of ill conditioning is that S is subject to haphazard and, in some cases, wild oscillation[37] instead of steady convergence to a minimum value. This is due to the parameter surface (if we momentarily consider three dimensions only) being a narrow, elongated, bent valley, so that the use of equation (16) results in rapid oscillation of direction across the narrow valley. Furthermore, if the starting values of the parameters are too far from the minimum, the Taylor-series approximation is poor and there is a consequent tendency to ' overshoot '. For this reason many workers emphasise the importance of good initial estimates of parameter values in order to ensure convergence.[5,18]

Most of the sophisticated modifications of the Gauss–Newton–Rapheson procedures, together with the various new algorithms for minimisation of a function, have been designed to enhance the speed and reliability of convergence. It is interesting to note how similar they are, being generally classifiable into two groups: (*a*) scaling or optimisation of the parameter corrections is carried out subsequent to the calculations, so as to bring about the maximum decrease in S for that iteration (Hartley's[38] and Greenstadt's[39] methods both use this approach); (*b*) a modification is made to the basic equation prior to the calculation of the corrections which tries to ensure that convergence will eventually be achieved. (Sillén's[13] and Marquardt's methods[40,41] both use this approach, whilst the Fletcher–Powell method[42,43] uses a combination of both approaches.)

In the present program convergence was effected using Marquardt's algorithm[40,41] which combines the method of steepest descent with methods based on the Taylor series expansion in such a way as to exploit the advantages of each. The optimum direction for convergence (Δ) is found by solving equation (25) where $\boldsymbol{I}$

$$(\widetilde{\boldsymbol{B}}\boldsymbol{WB} + c\boldsymbol{I})\Delta = \widetilde{\boldsymbol{B}}\boldsymbol{W}E \qquad (25)$$

is the unit matrix, $\boldsymbol{B}$, $\boldsymbol{W}$, and E have the same significance as before, and c is a quantity added to the diagonal elements of $\widetilde{\boldsymbol{B}}\boldsymbol{WB}$ (which for this procedure to be meaningful must first be normalised so that the diagonal elements are unity). When $c = 0$, Δ is the Taylor-series direction, and as c increases Δ swings towards the steepest-descent direction. Since for many real systems the angle between the Taylor-series and the steepest-descent directions is between 80 and 90°,[41] Marquardt's method attempts a compromise between the two. Within iterations, c is increased until a reduction in the sum of the squares of the residuals (S) is obtained; between iterations, c is reduced to ensure second-order convergence. The basic strategy is therefore as follows.

On the *i*th iteration c_i must be such that $S_i + 1 < S_i$. Let $V > 1$ (V is a scalar). Set c_0 equal to 0.01. Compute $S(c_{i-1})$ and $S(c_{i-1}/V)$.[*] Then: (*i*) if $S(c_{i-1}/V) \leqslant S_i$, let $c_i = c_{i-1}/V$; (*ii*) if $S(c_{i-1}/V) > S_i$ and $S(c_{i-1}) \leqslant S_i$, let $c_i = c_{i-1}$; and (*iii*) if $S(c_{i-1}/V) > S_i$ and $S(c_{i-1}) > S_i$, increase c by successive multiplication by V [†] until for some smallest w,[‡] $S(c_iV^w) \leqslant S_i$, then let $c_i = c_{i-1}V^w$.[§] Although it might appear that this method will involve a large number of calculations because of the repeated inversion of the matrix $(\widetilde{\boldsymbol{B}}\boldsymbol{WB} + c\boldsymbol{I})$, in fact this is not so because $\widetilde{\boldsymbol{B}}\boldsymbol{WB}$ is symmetric and can be diagonalised according to equation (26), where $\boldsymbol{L}$ is a diagonal matrix

$$(\widetilde{\boldsymbol{B}}\boldsymbol{WB})^{-1} = (\widetilde{\boldsymbol{U}}\boldsymbol{LU})^{-1} = \widetilde{\boldsymbol{U}}\boldsymbol{L}^{-1}\boldsymbol{U} \qquad (26)$$

of eigenvalues and $\boldsymbol{U}$ is the corresponding matrix of orthogonal eigenvectors. Hence, since (27) is applicable

$$(\widetilde{\boldsymbol{B}}\boldsymbol{WB})^{-1} = \widetilde{\boldsymbol{U}} \operatorname{diag} l_j^{-1}\boldsymbol{U} \qquad (27)$$

where l_j are the diagonal elements of $\boldsymbol{L}$, we obtain (28).

$$(\widetilde{\boldsymbol{B}}\boldsymbol{WB} + c\boldsymbol{I})^{-1} = \widetilde{\boldsymbol{U}} \operatorname{diag} [1/(l_j + c)]\boldsymbol{U} \qquad (28)$$

Equation (28) affords an alternative view of why Marquardt's algorithm is so successful at ensuring convergence. The addition of c to the diagonal elements of $\widetilde{\boldsymbol{B}}\boldsymbol{WB}$ has helped to overcome the ill conditioning because, if the smallest eigenvalue of $\widetilde{\boldsymbol{B}}\boldsymbol{WB}$ was previously l_0, it will now be $l_0 + c$. Since the determinant cannot now be zero the singularity problem has been removed.

Marquardt's algorithm has previously been used successfully in such diverse areas as force-constant calculations,[44,45] e.s.r. and n.m.r. calculations,[46,47]

---

[*] If c_{i-1} is already $\leqslant 1.0$ compared with the number of significant figures carried, then go on to tests (*ii*) and (*iii*) and ignore comparisons with $S(c_{i-1}/V)$.

[†] V is an arbitrary scalar of value > 1.

[‡] w Is the power to which V is raised to produce a reduction in the sum of the squares of the errors. Thus if $V = 10$, c_{i-1} is multiplied by 10^1, 10^2, 10^3, *etc.*, until $S(c_{i-1}V^w) \leqslant S_i$.

[§] Where parameters are highly correlated, c may increase to unreasonably large values. In such cases a scale factor b_i for the correction vector Δ is used when the angle, determined by c_i, between the correction vector Δ and the steepest-descent direction is less than $\pi/4$.

[34] R. H. Moore and R. K. Zeigler, Los Alamos Scientific Laboratory Report, LA-2367, March 1960.

[35] D. Dyrssen, N. Ingri, and L. G. Sillén, *Acta Chem. Scand.*, 1961, **15**, 694.

[36] J. C. Sullivan, J. Rydberg, and W. F. Miller, *Acta Chem. Scand.*, 1959, **13**, 2023.

[37] D. P. Feder, *Appl. Optics*, 1963, **2**, 1209.

[38] H. O. Hartley, *Technometrics*, 1961, **3**, 269.

[39] J. Greenstadt, *Math. Comput.*, 1967, **21**, 360.

[40] K. Levenberg, *Quart. Appl. Maths.*, 1944, **2**, 164.

[41] D. W. Marquardt, *J. Soc. Ind. Appl. Maths.*, 1963, **11**, 431.

[42] R. Fletcher and M. D. J. Powell, *The Comp. J.*, 1963, **6**, 163.

[43] G. W. Stewart, *J. Assoc. Comp. Machin.*, 1967, **14**, 72.

[44] D. M. Adams and R. G. Churchill, *J. Chem. Soc.* (*A*), 1970, 697.

[45] D. Papousek, S. Toman, and J. Pliva, *J. Mol. Spectroscopy*, 1965, **15**, 502.

[46] D. W. Marquardt, R. G. Bennett, and E. J. Burrell, *J. Mol. Spectroscopy*, 1961, **7**, 269.

[47] J. Pliva, V. Spirko, and S. Toman, *J. Mol. Spectroscopy*, 1966, **21**, 106.

optical-lens design,[32,37,48,49] and fitting of equations to chromatograms.[50] It was therefore adopted in the present work.

Model Testing.—In order to test a series of models to determine which gives the best fit to the observed data it is necessary to have some measure of the 'goodness of fit'. For a non-linear treatment there is no perfect measure [51-53] of the 'goodness of fit' and a number of possibilities have been suggested in the literature. Because different measures are affected differently by truncation of the Taylor series expansion depending on the 'degree of non-linearity', two 'goodness-of-fit' measures are used in the present program, S and R.

S, the sum of the squares of the errors between the observed ($O_{\text{obs.}}$) and calculated ($O_{\text{calc.}}$) observables (absorbance or e.m.f.) is defined by (29).[54] The Hamilton R factor [6,14] is defined by (30) where W_i are the appropriate weighting factors. This value is compared with

$$S = \sum_{i=1}^{N} (O_{\text{obs.}} - O_{\text{calc.}})_i^2 \qquad (29)$$

$$R = \left[\sum_{i=1}^{N} W_i (O_{\text{obs.}} - O_{\text{calc.}})_i^2 \Big/ \sum_{i=1}^{N} W_i (O_{\text{obs.}})_i^2 \right]^{\frac{1}{2}} \qquad (30)$$

the limiting value R_{lim} given by equation (31) where ε_i is the residual in the ith equation calculated from pessimistic estimates of the errors in all the experimental quantities using error-propagation rules.[55] A satisfactory fit may be assumed if $R < R_{\text{lim}}$. If two different models both yield R factors less than R_{lim} then the R-factor ratio test can be applied. Thus if two models H_1 and H_2 give R factors R_1 and R_2 respectively, then model H_2 can be rejected at the α-significance level if equation (32) applies, where m is the number of unknown parameters fitted to the data and $n - m$ is the number of degrees of freedom of the least-squares adjustment. The value of $R_{m,n-m,\alpha}$ is found from statistical tables.[14] If model H_2 cannot be rejected by inequality (32) then the two models are statistically indistinguishable (at significance level α). In this case, however, they may be distinguishable on chemical grounds or because one model gives unacceptable parameter values such as negative absorption coefficients in spectrophotometry.

$$R_{\text{lim}} = \left[\sum_{i=1}^{N} W_i \varepsilon_i^2 \Big/ \sum_{i=1}^{N} W_i (O_{\text{obs.}})_i^2 \right]^{\frac{1}{2}} \qquad (31)$$

$$R_2/R_1 > R_{m,n-m,\alpha} \qquad (32)$$

The Program DALSFEK. The present program, DALSFEK, which incorporates the 'damped least-squares' algorithm of Marquardt,[41] can calculate equilibrium constants from several types of variables. A flow diagram for DALSFEK is available as Supplementary Publication No. SUP 22168 (8 pp.).* To date we have written and successfully used [56,57] subroutines for both absorbance and e.m.f. data, and the program is so structured that if several observables are monitored simultaneously all the data can be used in evaluating the equilibrium constants. The subroutines are as follows.

DALS. DALS is the master routine. It reads in the control parameters such as the number of parameters to be adjusted, the number of fixed parameters, and the number of equilibria. After defining the equilibria it then calls CYCLE.

CYCLE. CYCLE defines the parameter arrays and assigns codes to the types of parameters to be adjusted. It controls the use of the magnetic-tape backing store, so that, as the program cycles over each experiment the total (analytical) concentrations of the components, the experimental observables, and (after entering CONC), the concentrations of each species are stored in a systematic way. The subroutine CONC is called to calculate the concentrations of the species by a Gauss–Newton iteration of the mass-balance and equilibrium equations. The $\boldsymbol{B}$ matrix of differentials is also created here and the matrix $\widetilde{\boldsymbol{B}}\boldsymbol{WB}$ formed. If these differentials can be obtained analytically from the concentrations the program does so, but for the equilibrium constants K the observables are related to these parameters in a non-explicit way. Difference approximations to the derivatives are then calculated using relation (33) (O = observable, K = equilibrium constant) which involves repeated calling of subroutine CONC. The fraction, Δ, by which K is incremented can be varied at will but 0.001 (0.1%) has been found to be satisfactory.

$$\partial O_i / \partial K_i = \Delta O_i / \Delta K_i \qquad (33)$$

When $\widetilde{\boldsymbol{B}}\boldsymbol{WB}$ has been created it is normalised and subroutine LEASTSQ is called to calculate the corrections to the parameters. When the problem has converged, subroutine ERROR is finally called to calculate the standard deviations of the parameters and also some statistical quantities to estimate the 'goodness of fit'. The final refined values and estimates of the standard deviations are then printed out.

CONC. This is a general subroutine which will calculate the concentrations of all the species in a system, given only the equilibrium constants and the mass-balance information. This is done by a Gauss–Newton iteration on some 'guessed' concentrations. Consider m equilibria relating n species in solution. At equilibrium there are m equations of the type (34) where

$$\log_e K_i = a_{i1} \log_e c_1 + \ldots a_{in} \log_e c_n \qquad (34)$$

---

* For details see Notices to Authors No. 7, *J.C.S. Dalton*, 1977, Index issue.

[48] C. G. Wynne and P. M. J. H. Wormell, *Appl. Optics*, 1963, **2**, 1233.

[49] G. H. Spencer, *Appl. Optics*, 1963, **2**, 1257.

[50] S. A. Roberts, D. H. Wilkinson, and L. R. Walker, *Analyt. Chem.*, 1970, **42**, 886.

[51] E. M. L. Beale, *J. Roy. Statistical Soc.*, 1960, **B22**, 41.

[52] I. Guttman and D. A. Meeter, *Technometrics*, 1963, **7**, 623.

[53] H. O. Hartley, *Biometrika*, 1964, **51**, 347.

[54] W. E. Wentworth, *J. Chem. Educ.*, 1965, **42**, 96.

[55] H. H. Ku, *J. Res. Nat. Bur. Stand. Sect. C*, 1966, **70**, 263.

[56] R. M. Alcock, F. R. Hartley, D. E. Rogers, and J. L. Wagner, *J.C.S. Dalton*, 1975, 2189, 2194.

[57] F. R. Hartley, G. W. Searle, R. M. Alcock, and D. E. Rogers, following paper.

$a_{in} > 0$ for a product and <0 for a reactant. This is expressed in matrix notation by (35). Also, consideration of the material balance gives (36) where T_j is the total (analytical) concentration of component j; q_{ji} are integers defining the number of atoms of component j in species i. In matrix notation this becomes (37).

$$\log_e K = \boldsymbol{A} \log_e C \tag{35}$$

$$T_j = q_{j1}c_1 + \ldots q_{jn}c_n \tag{36}$$

$$T = \boldsymbol{Q}C \tag{37}$$

Since $\boldsymbol{A}$ and $\boldsymbol{Q}$ are fixed by the equilibrium definition, then from a knowledge of T and the current values of $\log_e K$ a trial set of C can be adjusted until equations (35) and (37) are satisfied.

The B-matrix elements can be obtained analytically since the equalities (38) are applicable. Probably because of these exact differentials, it has been found that this routine is remarkably well behaved and gives species distributions satisfying the equations from almost any initial starting set of concentrations.

$$\frac{\partial \log_e K_i}{\partial c_n} = \frac{a_{in}}{c_n} \text{ and } \frac{\partial t_j}{\partial c_n} = q_{jn} \tag{38}$$

Table 1 shows an example of the fitting of CONC to some synthetic data for a two-equilibrium model. The (arbitrary) criterion for convergence used to obtain the concentrations in Table 1 was $S < 5.0 \times 10^{-8}$ where S is given by (39) in which nT is the number of chemical components and nK is the number of equilibrium constants. The relative sum of the squares of the residuals must be minimised in this way because the absolute values of the analytical concentrations, T, and the equilibrium constants, K, can vary over many orders of magnitude. For the mass-balance equations this is analogous to minimising the weighted sum of the squares of the residuals, in which the weighting factors are the reciprocals of the squared observables. For the equilibrium equations the comparison is less clear, since the residuals in the least-squares equations are defined in terms of the logarithms of the equilibrium constants. For this reason weighting factors were not incorporated into these least-squares equations.

$$S = \sum_{i=1}^{nT} \left(\frac{T_{\text{obs.}} - T_{\text{calc.}}}{T_{\text{obs.}}} \right)_i^2 + \sum_{j=1}^{nK} \left(\frac{K_{\text{obs.}} - K_{\text{calc.}}}{K_{\text{obs.}}} \right)_j^2 \tag{39}$$

TABLE 1

Application of CONC to synthetic data generated for the equilibria:

$$2[PdCl_4]^{2-} \overset{K_1}{\rightleftharpoons} [Pd_2Cl_6]^{2-} + 2Cl^-$$

$$[Pd_2Cl_6]^{2-} + \text{solvent} \overset{K_2}{\rightleftharpoons} [Pd_2Cl_5(\text{solvent})]^- + Cl^-$$

Analytical chloride concentration ($[Cl]_T$) = 5.60×10^{-3} mol dm^{-3}, analytical palladium concentration ($[Pd]_T$) = 8.054×10^{-4} mol dm^{-3}, $K_1 = 2.818$ mol dm^{-3}, $K_2 = 1 \times 10^{-4}$

Quantity	Initial ' guessed ' value	Final ' calculated ' value
$[PdCl_4^{2-}]$/mol dm^{-3}	1.0×10^{-4}	3.633×10^{-5}
$[Pd_2Cl_6^{2-}]$/mol dm^{-3}	1.0×10^{-4}	3.727×10^{-4}
$[Cl^-]$/mol dm^{-3}	1.0×10^{-2}	3.159×10^{-3}
$[Pd_2Cl_5(\text{solvent})^-]$/mol dm^{-3}	1.0×10^{-5}	1.800×10^{-5}
$[Cl]_t$/mol dm^{-3}		5.600×10^{-3}
$[Pd]_t$/mol dm^{-3}		8.054×10^{-4}
K_1/mol dm^{-3}		2.819
K_2		1.000×10^{-4}

When the concentrations of the species have been found a control parameter is read in, which determines which OBSERV subroutine is called. Thus, if the observables are absorbances, subroutine OBSERV 1 is entered. Control then returns to CYCLE.

OBSERV 1. This subroutine allows stability constants to be calculated from spectrophotometric data. The absorbances of the solutions are calculated from the species concentrations assuming that Beer's law holds. The molar absorption coefficients are read in when this routine is first called. If the molar absorption coefficients are to undergo refinement, then initial ' guessed ' values are read in by OBSERV 1.

VARIAN 1. If it is desired to calculate weights and minimise the weighted sum of the residuals in the absorbances, this subroutine is entered in order to estimate the variances of the residuals. At a particular wavelength λ the residual, r, is given by (40). Assuming that the variance of r depends only on the errors in the experimental quantities, the observed absorbances (A), and the total (analytical) concentrations (T_j), then by error-propagation theory, we can write (41) provided there is no error correlation between A and the T_j. $\sigma_A{}^2$ and $\sigma_{T_j}{}^2$ are the variances in the absorbance readings and the total-concentration values respectively. These quantities may be known from a previous experiment or estimated from repeat values of the same observation.

$$r = A_{\text{obs.}} - A_{\text{calc.}} \tag{40}$$

$$\sigma^2 = \left(\frac{\partial r}{\partial A} \right)^2 \sigma_A{}^2 + \sum_{j=1}^{n} \left(\frac{\partial r}{\partial T_j} \right)^2 \sigma_{T_j}{}^2 \tag{41}$$

OBSERV 2. It is a common procedure in stability-constant determinations to measure free-metal or free-ligand concentrations potentiometrically. If this routine is supplied with information as to which species is being monitored in this way (*e.g.* free metal ion), e.m.f. data are readily calculated for comparison with experimentally observed values. Thus, if the species being experimentally determined in this fashion is numbered as 2, and the concentration of this species in the reference electrode is c_0, then a calculated e.m.f. value may be evaluated from the Nernst equation using c_0 and the concentration of species 2 as obtained from subroutine CONC.

A subroutine similar to that used in OBSERV 1 is used to calculate the variance.

LEASTSQ. When all the observables have been calculated from the initial values of the parameters and the sum of the squares of the errors calculated, this routine is entered from CYCLE to calculate the corrections to the parameters. The matrix $\widetilde{\boldsymbol{B}}\boldsymbol{WB}$ is diagonalised by subroutine HDIAG. By multiplying the eigenvectors into the matrix $\widetilde{\boldsymbol{B}}\boldsymbol{E}$ and subsequently multiplying by the reciprocals of the eigenvalues, the contribution to the correction vector from each eigenvalue is calculated. The corrections are applied to the parameters and subroutine CONC, and subsequently an OBSERV subroutine, are called to recalculate the sum of the squares of the errors for comparison with the previous value. Marquardt's algorithm is then followed to decide whether the corrections must be ' damped ' or whether the new values of the parameters can be carried through to begin the next iteration.

The eigenvalues (roots) and the calculated parameter corrections may be written out in this subroutine. This is useful for observing the conditioning of $\widetilde{\boldsymbol{B}}\boldsymbol{WB}$ and for checking whether any large parameter corrections have consequently been calculated.

HDIAG. This routine calculates the eigenvectors and eigenvalues of a square symmetric matrix by a Gauss–Jordon elimination.

ERROR. This is the last subroutine entered before return to the master routine DALS. It calculates an unbiased estimate of the variance of an observation of unit weight, σ^2, as in equation (24), and the standard deviations of the parameters from the diagonal elements of the variance–covariance matrix as in equation (23). Correlation coefficients are also calculated.[14]

The program was tested by generating synthetic data for three types of problem: (*i*) spectrophotometric data only; (*ii*) potentiometric data only; and (*iii*) a combination of spectrophotometric and potentiometric data. In the course of this work the importance of correctly weighting the data, when the variances of the residuals vary over several orders of magnitude, became apparent. Weighting the data has been emphasised previously,[18,19,26,58] although a statistically correct weighting scheme is only possible if good estimates of the variance–covariance matrix observations are available.[7] A separate subroutine VARIAN to calculate weighting factors was written for both of the OBSERV subroutines, but weighting was much more critical for potentiometric than spectrophotometric data, as found by other workers.[20,59,60]

In addition to experimental errors in the observables, it is important to consider systematic errors, for example in the total analytical concentrations of each of the species present. Sillén showed that such systematic errors should be treated as extra parameters in the computer evaluation of data.[13,26] This is an important technique in defining a ' chemical model ', since often a better fit to the experimental data can be obtained by allowing those parameters liable to systematic error to be varied by the program than by the introduction of an extra species. If, of course, the adjustments made to these parameters are outside their possible experimental error, or if they vary systematically in an implausible way, then it is likely that the ' model ' is incorrect and further species exist. Accordingly, provision was incorporated into DALSFEK for allowing the total analytical concentrations of each species to be varied. We have previously described an example where the application of this facility enabled us to detect a minor component formed when $Na_2[PdCl_4]$ is dissolved in glacial acetic acid.[56]

Applications of DALSFEK.—*Copper*(II)*-ethylenediamine-oxalate system.* DeWitt and Watters[61] determined the first (K_{42}) and second (K_{43}) equilibrium constants for the substitution by oxalate of ethylenediamine (en) bound to Cu^{II} [reactions (42) and (43)]. These workers determined the pH and absorbances at six

$$[Cu(en)_2]^{2+} + [C_2O_4]^{2-} \overset{K_{42}}{\rightleftharpoons} [Cu(en)(C_2O_4)] + en \quad (42)$$

$$[Cu(en)(C_2O_4)] + [C_2O_4]^{2-} \overset{K_{43}}{\rightleftharpoons} [Cu(C_2O_4)_2]^{2-} + en \quad (43)$$

wavelengths for a series of solutions and calculated the equilibrium constants by three different procedures (Table 2). The original data were used as input in DALSFEK, the pH values as independent (error-free) variables held fixed in subroutine CONC, while the species distribution for each solution was calculated. The sum of the squares of the residuals in the absorbances was then minimised to calculate the equilibrium constants. Since ethylenediamine is a fairly strong base, allowance was made for its successive acid-dissociation constants using the same data as the original workers. Oxalic acid was assumed to be a sufficiently strong acid to remain dissociated throughout the pH range of the experiment (5.11—11.12), an assumption also made by the original workers.

TABLE 2

Comparison of calculation procedures for the equilibrium constants of equilibria (42) and (43)

Method of calculation	pK_{42}	Error or range	pK_{43}	Error or range	Ref.
Slope intercept	4.53	$\pm$0.18 [a]	5.88	Not stated	61
Isosbestic points	4.54	(4.41—4.85) [a]	5.60	(5.53—5.67) [a]	61
Determinants	4.51	(4.38—4.74) [a]	5.62	(5.51—5.71) [a]	61
DALSFEK	4.78	0.022 [b]	5.96	0.019 [b]	This work

[a] As recorded in ref. 61. [b] Standard deviations.

The results in Table 2 show fair agreement between the DALSFEK calculations and the original calculations. An even better agreement could probably have been obtained by a suitable weighting procedure, since the pH measurements are, of course, subject to error. However, this is difficult to do without some prior estimate

---

[58] J. Rydberg, *Acta Chem. Scand.*, 1961, **15**, 1723.

[59] R. C. Lansbury, V. E. Price, and A. G. Smeeth, *J. Chem. Soc.*, 1965, 1896.

[60] R. S. Tobias and Z. Z. Hugus, *J. Phys. Chem.*, 1961, **65**, 2165.

[61] R. DeWitt and J. I. Watters, *J. Amer. Chem. Soc.*, 1954, **76**, 3810.

of the variance of the pH readings. Unit weights were therefore used.

Palladium(II)*-chloride-bromide system.* Feldberg *et al.*[8] published a computer evaluation of the successive formation constants of $[PdCl_nBr_{4-n}]^{2-}$ in aqueous solution. These workers obtained very precise absorbance data (± 0.001 absorbance unit) on 61 solutions in

$$[PdCl_nBr_{4-n}]^{2-} + Br^- \overset{K_n}{\rightleftharpoons} [PdCl_{n-1}Br_{5-n}]^{2-} + Cl^- \quad (44)$$

which they varied the $[Br^-] : [Cl^-]$ ratio over a very wide range. Their computer program used the method of steepest descent. Starting from several initial guessed values of the overall formation constants (β_n), the program was allowed to iterate towards the minimum until a particular tolerance limit was satisfied. This resulted in several sets of rapid convergence values. It was found that the sum of the squares of the errors, S, was very insensitive to the rapid convergence value, especially in the case of β_4. This was because the highly correlated parameters produced a broad flat minimum in multidimensional space. Not surprisingly, therefore, the use of the steepest-descent method meant that unless the convergence tolerance was reduced to a very small value, and the program allowed to iterate for a long time, the minimum attained for each set of initial guessed values was different. The solution used by the original workers was to plot out the value of the sum of the squares of the errors as a function of $\log_{10}\beta_4$ in order to find an optimum value of β_4. Finally, the corresponding values of β_1, β_2, and β_3 were calculated by noting that in the vicinity of the minimum equation (45) applies where a_n and b_n are constants, which were evaluated by

$$\beta_n = a_n\beta_4 + b_n \quad (n = 1\text{—}3) \qquad (45)$$

plotting β_n against β_4 for the series of computer-calculated rapid-convergence values. As in so many practical problems, the shape of the parameter surface near the minimum appears to be a skewed flat-bottomed trough and equation (45) therefore offers an alternative to Sillén's twist-matrix method [13] for deciding on the most efficient manner in which to vary β_n.[8] These β values were used as highly refined initial guesses for calculating the final set of formation constants.

TABLE 3

Comparison of the equilibrium constants and molar absorption coefficients for the $[PdCl_nBr_{4-n}]^{2-}$ system calculated by using DALSFEK with those reported in the literature

K_n [a]	DALSFEK	Literature [b]	σ_{K_n}	DALSFEK	Literature [c]
K_1	25.45	25.13	σ_{K_1}	1.88	1.47
K_2	11.38	11.48	σ_{K_2}	0.30	0.522
K_3	5.25	5.248	σ_{K_3}	0.34	0.314
K_4	1.89	1.862	σ_{K_4}	0.10	0.129

	ε/dm^3 mol^{-1} cm^{-1}					
	$[PdCl_3Br]^{2-}$		$[PdCl_2Br_2]^{2-}$		$[PdClBr_3]^{2-}$	
λ/nm	DALSFEK	Literature [d]	DALSFEK	Literature [d]	DALSFEK	Literature [d]
265	1 611	1 700	2 669	2 700	6 911	6 900
275	4 939	5 000	2 482	2 500	2 808	2 900
285	10 320	10 300	4 876	4 900	1 725	1 700
295	11 710	11 700	9 066	9 000	3 431	3 500
305	8 992	8 900	10 700	10 700	7 752	7 800
315	5 019	5 100	9 365	9 400	10 510	10 400
325	2 207	2 000	6 643	6 500	9 810	9 900
335	926	1 000	4 105	4 000	7 545	7 600
345	568	500	2 178	2 000	5 252	5 200
355	601		1 074	1 000	3 428	3 500
365	615		642		2 032	1.900
375	564		472		1 346	

[a] K_n is defined by equilibrium (44). [b] All the data are from ref. 8. The reference quotes $\log_{10}K_n$. The values tabulated here are the antilogarithms of the original values. [c] The original paper quoted $\sigma(\log_{10}K_n)$ values. The quantities tabulated here were obtained using the formula $\sigma_y{}^2 = (\partial y/\partial x)^2\sigma_x{}^2$. [d] The literature values of ε were read from the published spectra since numerical values were not reported.

The original data of Feldberg *et al.* were obtained and analysed using DALSFEK to determine not only the equilibrium constants but also the molar absorbances of the species present. DALSFEK was able, in a reasonable amount of time (*ca.* 200 s on a CDC 7600 *) to duplicate the results of Feldberg *et al.* from almost any set of initial guessed values (see Table 3). Because Marquardt's algorithm calculates the best search direction automatically, the time-consuming graphical processes have been eliminated. In addition, the use of a least-squares procedure has enabled estimates of the standard deviations of the equilibrium constants to be made, whereas Feldberg *et al.* could only obtain these quantities by dividing their data into subsets. In view of the different manner of calculating these deviations, the agreement is surprisingly good.

Other systems. We have previously described the application of DALSFEK to a spectrophotometric determination of the palladium(II)-chloro-olefin system in acetic acid [56] and to a potentiometric study of silver(I)-olefin equilibrium constants in a number of solvents.[57]

---

* Feldberg *et al.* quoted 1 200 s on a CDC 6600. Whilst a 7600 is faster than a 6600 computer it is not six times faster so that the present program is clearly a more efficient one than than that used by the previous workers.

In the latter case DALSFEK was compared with other computational procedures and found to give identical results.

Conclusions.—DALSFEK is a versatile computer program for the determination of equilibrium constants, from either spectrophotometric or potentiometric data, or a combination of both types of data. It is relatively simple to use taking *ca.* 1 h to prepare and punch up the data. The use of Marquardt's algorithm ensures a rapid convergence to the minimum and imposes less restrictions on the precision of the initial guessed values of the parameters than most other programs we have examined.

We thank Dr. L. Newman for the original experimental data on the palladium(II)–chloride–bromide system, the Southampton University Computer Advisory Service, Drs. J. L. Wagner, G. W. Searle, and C. E. Burgess for the provision of experimental data and valuable discussions, and the S.R.C. for the award of a studentship (to R. M. A.).

[7/696 *Received, 25th April,* 1977]