A rapidly convergent iterative method for the solution of the generalised nonlinear least squares problem

D. R. Powell and J. R. Macdonald

Texas Instruments Incorporated, PO Box 5936, Dallas, Texas 75222, USA

A new iterative method is described for the solution of the generalised nonlinear least squares problem: where the model may be nonlinear in its parameters and in the independent variable(s) and all variables are subject to error. The method is described for the case of two arbitrarily related variables; does not require the analytic calculation of derivatives; leads to exceptionally close satisfaction of the least squares conditions; and exhibits especially rapid convergence arising from the use of somewhat unconventional numerical approximations for partial derivatives. Examples are given which compare the results of the method of those of other existing techniques.

(Received October 1970, Revised October 1971)

It is frequently the object of an experiment to find, prove, and/or document some causal relation between experimental variables. Such a relationship may be called a model for the phenomenon investigated. Since a model generally establishes only a small and finite number of connections between several variables and a limited number of parameters, it is always an idealisation of even an isolated part of nature. Once a model has been selected as 'best' by some (preferably) objective criterion from the finite set of ones thought possible for the situation involved, the problem arises of finding the most appropriate values, or estimates, for the parameters involved in the model. Both model selection and parameter determination are often carried out using the method of least squares.

In ordinary least squares analysis it is customary to consider one 'dependent' variable y and l 'independent' variables, x_j, where $j = 1, 2, \ldots, l$. The model is then generally written as $y = f(\mathbf{x}, \boldsymbol{\alpha})$; the components of the vector $\boldsymbol{\alpha}$ are the α_k parameters; and $k = 1, 2, \ldots, m$. Conventionally, dependent and independent variables are more often distinguished not strictly by dependence and independence but by the assumption of no, or negligible, errors in experimental values of the x_j, denoted X_{ji}, and the expected presence of such errors in experimental values of y, denoted Y_i. Here $i = 1, 2, \ldots, n$, and n is the number of (Y_i, X_{ji}) experimentally determined $(l + 1)$-tuple values. Both because it is still the most common case and for simplicity, the rest of this paper will deal only with two variables, y and x, thus taking $l = 1$. Our results may be readily generalised for arbitrary l, however, by well-known methods (Deming, 1943; Wolberg, 1967).

Although there are experimental situations where errors in X are negligible compared to those in Y, only when a variable is intrinsically discrete (because of quantum-mechanical or whole-number reasons) is there a possibility of obtaining exact values of it from experiment. Further, there are many measurement areas where not only are unavoidable random errors present in both Y and X, but those in X are too large to neglect in the usual way. Ordinary least squares, even with weighting of the Y observations, is then inadequate and can lead to bias in estimated parameter and variance values (Macdonald, 1969; Macdonald and Powell, 1971).

Clearly then, a least squares procedure which takes proper account of errors in both variables is frequently needed. Note that such a procedure will be useful whether or not one variable, say X, is 'controlled' (by being set as closely as possible to a given value) or not (Macdonald, 1969). Deming (1943) long ago gave an approximate solution of the problem which has been too little known and used. It has, however, been employed by Wentworth (1965) and, more recently, has been described and extensively applied (Wolberg, 1967; Macdonald, 1969). Southwell (1969) has recently stated, 'There has thus far been no satisfactory solution to the general case of fitting data with errors in both co-ordinates to nonlinear functions.' He then goes on to present what he believes to be such a solution. By 'nonlinear functions' Southwell means functions which are not linear in one or more of the α_k's.

Now Deming's solution does apply to functions nonlinear in the parameters involved and has been used for them (Wentworth, 1965; Wolberg, 1967; Macdonald, 1969). Unfortunately, Southwell gives no comparison with Deming's work on either a theoretical or computational basis. Further, Southwell claims to find exact expressions for the variances of the parameter value estimates, $\hat{a}_k \equiv a_k$. But Draper and Smith (1966), among others, have pointed out that least squares solutions applied to models involving parameters nonlinearly (hereinafter to be termed 'nonlinear models') lead to bias in the usual statistical estimates (see also Macdonald, 1969; Macdonald and Powell, 1971). Finally, perhaps a minor point, but one beloved of the statisticians: Southwell generally ignores the distinction between estimated and exact values and claims to give (nonasymptotic) formulae for exact parameters as well as exact variances. Evidently, Southwell has not tried out his iterative solution on many nonlinear models (he presents quantitative results only for straight line fitting); had he done so, he would have found that even with exceptionally good initial estimates of the parameter values his procedure would not have converged to the *least* squares solution and often would not have converged at all.

The situation appears to us to be as follows. Notwithstanding Southwell's work and claims, up to the present there seems to be available no least squares procedure which will converge to a true least squares solution when applied to a general nonlinear model involving weighting of both Y and X. Further, there has been no complete assessment made yet of the adequacy of Deming's solution for such a problem. In the present work, we present a solution to the above general problem which exhibits exceptional convergence characteristics. While we do not claim invariable convergence for any conceivable model, we do claim, and demonstrate, a *least* squares solution when convergence is achieved. Although the problem of bias in estimates obtained from arbitrary nonlinear least squares solutions is difficult and has not been totally resolved, Box (1971) has recently made an important contribution to the area.

Finally, it is important to emphasise that all models (except $y = \alpha_1 + \alpha_2 x$) are nonlinear in their parameters when the X values are not taken exact but are weighted. Even though the model itself is formally linear in all its parameters, the x-

weighting creates model nonlinearity (Macdonald, 1969). Thus, all polynomials, $y = \sum_{k=1}^{m} \alpha_k x^{k-1}$, of degree higher than one ($m > 2$) become nonlinear in the parameters with x-weighting. Besides the common case of weighting of both variables, x-weighting alone is frequently important for models nonlinear in their *variables*. Suppose that the Y_i's are known so precisely that they may be approximated as exact but that there is uncertainty present in the X_i's. Further, suppose that $y = f(x, \boldsymbol{\alpha})$ is sufficiently nonlinear in x that it cannot be solved directly for x (either because such an explicit solution is mathematically impossible or because it is too complex to be practical). Then ordinary least squares cannot be properly applied to this $y = f(x, \boldsymbol{\alpha})$ situation. Instead, at best one can only weight the Y_i's (actually almost exact) in an effort to compensate for the uncertainty in the X_i's (which must be taken exact). Such improper weighting will, at the least, introduce bias in the parameter standard deviation estimates. The present generalised least squares solution allows one to apply weighting properly in this case and to obtain results for parameter estimates and standard deviation estimates identical to those which would have been obtained had x been solved for directly.

2. Theory

An important contribution to the present problem appeared in early 1969, about the time Southwell's work was submitted for publication: O'Neill, Sinclair, and Smith (1969) gave a rapidly convergent solution of the least squares problem with arbitrary weighting of both X and Y, which, however, applies only to polynomials in x. As we have seen, this is not in general a linear model situation. To reduce computing time and increase accuracy, O'Neill *et al.* make use of expansions of $f(x, \boldsymbol{\alpha})$ in sets of orthogonal polynomials. Many authors (for example, Hall and Canfield, 1967) have used orthonormal polynomials in ordinary least squares analysis. These special techniques are inapplicable for a general model, which may be overtly nonlinear in both its parameters and variables, and are not needed in our present solution for such a model.

O'Neill *et al.* point out two approximations in Deming's solution and show how both may be avoided in their own approach. Further, they apply their method to data involving both straight-line and cubic (in x) models. They compare their numerical results for the cubic, with unity weights of both Y_i and X_i, to results given by Deming's method. Since no parameter standard deviation estimates are presented for either solution, however, the significance of the differences between the two is difficult to assess.

Although O'Neill *et al.* give a much improved iterative solution to the double-weighting problem for polynomials, they find it to be computationally impractical and go on to derive and use a simple approximation to this solution. We shall follow them in using much the same procedure but will further demonstrate that the approximation made, unlike those in Deming's method, can lead to exceptionally close satisfaction of the least squares conditions and thus, when convergence is obtained, to the true *least* squares solution. Further, to vastly expand the generality and applicability of our solution, which must be useful for models nonlinear in both parameters and variables, we shall not use derivatives which must be analytically calculated anew for each specific model, but shall obtain all needed derivatives by numerical approximation. Finally, we shall demonstrate that our iterative method converges much faster, because of the unconventional nature of the approximate numerical derivatives we use, than other procedures which use *exact* partial derivatives!

Let us define the least squares residuals $R_{y_i} \equiv Y_i - y_i$ and $R_{x_i} \equiv X_i - x_i$. Here y_i and x_i are values of y and x calculated as part of the least squares solution. The sum of squares to be minimised is then

$$S \equiv \sum_{i=1}^{n} (w_{y_i} R_{y_i}^2 + w_{x_i} R_{x_i}^2) , \tag{1}$$

where the weights $w_{y_i} \equiv \sigma_{y_i}^{-2}$, $w_{x_i} \equiv \sigma_{x_i}^{-2}$, and the σ's are the standard deviations (or their estimates) of Y_i and X_i, assumed known or measured. Minimisation of S thus minimises the sum of the weighted perpendicular distances between the experimental points (Y_i, X_i) and the calculated curve $y_i = f(x_i, \mathbf{a})$. Clearly, to obtain best parameter-estimate values one should ideally minimise the sum of the perpendicular distances between the true (y, x) points and the curve. Since such points are always unknown in real (as opposed to synthetic) situations, the present procedure seems next best.

The actual minimisation conditions are

$$\frac{\partial S}{\partial \alpha_k} = 0, \tag{2}$$

$$\frac{\partial S}{\partial x_i} = 0. \tag{3}$$

There are thus $(m + n)$ simultaneous (typically nonlinear) equations to be solved to yield the least squares y_i and x_i and corresponding parameter estimates and their estimated standard deviations.

To solve (2) and (3) iteratively, we now define

$$\delta a_k = a_k - a_{k0}, \tag{4}$$

$$\delta x_i = x_i - x_{i0}, \tag{5}$$

where a_{k0} and x_{i0} are starting estimates. We assume δa_k and δx_i to be small and, following O'Neill *et al.*, expand (2) and (3) using Taylor's series to first order, obtaining

$$\left(\frac{\partial S}{\partial \alpha_k}\right)_0 + \sum_{s=1}^{m} \left(\frac{\partial^2 S}{\partial \alpha_k \partial \alpha_s}\right)_0 \delta a_s + \sum_{j=1}^{n} \left(\frac{\partial^2 S}{\partial \alpha_k \partial x_j}\right)_0 \delta x_j = 0 , \tag{6}$$

$$\left(\frac{\partial S}{\partial x_i}\right)_0 + \sum_{s=1}^{m} \left(\frac{\partial^2 S}{\partial x_i \partial \alpha_s}\right)_0 \delta a_s + \sum_{j=1}^{n} \left(\frac{\partial^2 S}{\partial x_i \partial x_j}\right)_0 \delta x_j = 0 . \tag{7}$$

Here the zero subscript denotes derivative evaluation at (x_{i0}, a_{k0}). This truncation is the first approximation introduced. Provided convergence is obtained, it does not cause the *least* squares character of the solution to be lost. We now have $(m + n)$ nonlinear equations to solve for δa_k and δx_i. Using the development and notation of O'Neill, *et al.*, we define

$$g_k^{(1)} \equiv \left(\frac{\partial S}{\partial \alpha_k}\right)_0 , \; g_i^{(2)} \equiv \left(\frac{\partial S}{\partial x_i}\right)_0 ,$$

$$A_{sk} \equiv \left(\frac{\partial^2 S}{\partial \alpha_s \partial \alpha_k}\right)_0 , \; B_{ji} \equiv \left(\frac{\partial^2 S}{\partial x_j \partial x_i}\right)_0 , \tag{8}$$

$$C_{kj} = C'_{jk} \equiv \left(\frac{\partial^2 S}{\partial \alpha_k \partial x_j}\right)_0 ,$$

with $1 \leq k, s \leq m$ and $1 \leq i, j \leq n$.

We can write (6) and (7) in the form of a partitioned matrix,

$$\begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}' & \mathbf{B} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta a} \\ \boldsymbol{\delta x} \end{bmatrix} = - \begin{bmatrix} \mathbf{g}^{(1)} \\ \mathbf{g}^{(2)} \end{bmatrix} , \tag{9}$$

The solution is then given by

$$\begin{bmatrix} \boldsymbol{\delta a} \\ \boldsymbol{\delta x} \end{bmatrix} = -\mathbf{G}^{-1} \begin{bmatrix} \mathbf{g}^{(1)} \\ \mathbf{g}^{(2)} \end{bmatrix} , \tag{10}$$

where $\mathbf{G}$ is the partitioned matrix involving $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$.

It is theoretically possible to calculate all the derivatives in (8), invert the matrix $\mathbf{G}$, calculate the solution vector from (10), adjust the guesses, and repeat the procedure until convergence. There are several drawbacks in this approach, however. For

one thing, the model might be such that obtaining explicit formulae for the derivatives in (8) would be quite difficult. As a matter of fact, the model would not have to be too complex before writing out the second partials without at least one error would be unlikely. Even if the derivatives were obtained numerically, we would not have avoided the second, and usually more serious, objection to the procedure: the inversion of a matrix of order $(m + n)$. As O'Neill *et al.* point out, it is not unusual to have hundreds of data points.

To approach these computational problems, we first note that

$$B_{ij} = 0 \text{ if } i \neq j, \tag{11}$$

so that **B** is a diagonal matrix. Now, if the model were always a linear (in $\boldsymbol{\alpha}$) combination of orthogonal functions, as in the orthogonal polynomial case, then

$$A_{sk} = 0 \text{ if } k \neq s, \tag{12}$$

and **A** would also be a diagonal matrix. In our general case, however, no such simplification is possible.

We do follow O'Neill *et al.* in making the assumption that the elements of the matrix **C**, involving mixed partial derivatives with respect to x_i and α_k, have negligible effects. This assumption allows us to solve (10) and obtain

$$\begin{aligned} \boldsymbol{\delta a} &= -\mathbf{A}^{-1}\mathbf{g}^{(1)} \\ \boldsymbol{\delta x} &= -\mathbf{B}^{-1}\mathbf{g}^{(2)} \end{aligned} \tag{13}$$

Using the expression for S, we find, after some algebra, that

$$\delta x_i = \frac{w_{y_i}(Y_i - y_{i_0})\, y'_{i_0} + w_{x_i}(X_i - x_{i_0})}{w_{y_i}(y'_{i_0})^2 - w_{y_i}(Y_i - y_{i_0})\, y''_{i_0} + w_{x_i}}, \tag{14}$$

where $y_{i_0} = f(x_{i_0}, \mathbf{a})$, $y'_{i_0} = f'(x_{i_0}, \mathbf{a})$, $y''_{i_0} = f''(x_{i_0}, \mathbf{a})$, and the primes denote differentiation with respect to x. This result is formally the same as that given by O'Neill *et al.*, but our actual use of it is different from theirs, as will be seen.

3. The algorithm

In designing an algorithm, several important factors must be considered and balanced against one another. It seems to us that ease of use is very often as important as conservation of machine time. This was the principal guideline for the development of our algorithm. Two fundamental requirements seem called for under such a guideline. First, the allowable models should be completely general instead of polynomials. Second, the derivatives should be numerically estimated using uniquely appropriate difference formulae. Further, the computer program and its execution should involve enough accuracy that all except the most extremely pathological cases will converge to the *least* squares solution without significant round-off or truncation errors.

It would be fairly straightforward to generalise the O'Neill *et al.* method and replace all derivatives by simple difference formulae, but such an approach does not exploit an opportunity to accelerate convergence dramatically. After some numerical experimentation, we found that one can compensate for the omission of the mixed partials C_{kj} in **G** to a significant degree by unconventional substitutions for the derivatives in $\mathbf{g}^{(1)}$ and **A**. By using analytical derivatives, O'Neill *et al.* fail to gain this advantage. In our later examples, we will demonstrate that our general iterative method converges rapidly even with five parameters to be determined.

Our procedure needs both first and second partials to obtain the elements of $\mathbf{g}^{(1)}$ and **A**, and to evaluate (14). We use the following conventional formulae each time (14) is employed in the iteration described below:

$$y'_{i_0} = \frac{1}{2h_{i_0}}\left[f(x_{i_0} + h_{i_0}, \mathbf{a}) - f(x_{i_0} - h_{i_0}, \mathbf{a})\right] + 0(h_{i_0}^2), \tag{15}$$

and

$$y''_{i_0} = \frac{1}{h_{i_0}^2}\left[f(x_{i_0} + h_{i_0}, \mathbf{a}) - 2f(x_{i_0}, \mathbf{a}) + f(x_{i_0} - h_{i_0}, \mathbf{a})\right] + 0(h_{i_0}^2), \tag{16}$$

where $h_{i_0} \equiv \Delta_x x_{i_0}$, and Δ_x is a suitably small input constant, usually in the range of 10^{-4} to 10^{-6}. For utmost accuracy, we generally pick a final value of Δ_x (and the similar Δ_a defined below) in the middle of a region of Δ_x values which all yield essentially the same intermediate and final iterative results. As is well known (Kopal, 1955), increments used in numerical differentiation must be neither too large nor too small in order to balance optimally the effects of truncation and round-off errors.

Now, to complete one step in the iteration, i.e. to obtain the desired increments $\boldsymbol{\delta a}$ and $\boldsymbol{\delta x}$ from (13), we do the following:

1. Using the current $\mathbf{x}_0$ and $\mathbf{a}_0$ vectors, apply formula (14) to get a change in $\mathbf{x}_0$ (call it $\boldsymbol{\delta x}_0^*$). Using Y and $\mathbf{x} = \mathbf{x}_0 + \boldsymbol{\delta x}_0^*$, calculate S from formula (1) and call it $S(0, 0)$.
2. For a given pair of subscripts β and γ, where $1 \leq \beta, \gamma \leq m$, let $h_\beta = a_{\beta_0}\Delta_a$ and $h_\gamma = a_{\gamma_0}\Delta_a$, with Δ_a again a suitably small input constant usually taken equal to Δ_x.
3. Let $a_\beta = a_{\beta_0} + h_\beta$, $a_\gamma = a_{\gamma_0}$, and keep all other parameters unchanged. Apply formula (14) as in (1) and get S. Call it $S(1, 0)$.
4. Let $a_\beta = a_{\beta_0} - h_\beta$, $a_\gamma = a_{\gamma_0}$, apply formula (14), and get $S(-1, 0)$, similar to (1) and (3).
5. Let $a_\beta = a_{\beta_0} + h_\beta$, $a_\gamma = a_{\gamma_0} + h_\gamma$, and using the technique of (1) get $S(1, 1)$.
6. Similarly, let $a_\beta = a_{\beta_0} - h_\beta$, $a_\gamma = a_{\gamma_0} - h_\gamma$, and then get $S(-1, -1)$.
7. Let $a_\beta = a_{\beta_0}$, $a_\gamma = a_{\gamma_0} + h_\gamma$, and get $S(0, 1)$.
8. Let $a_\beta = a_{\beta_0}$, $a_\gamma = a_{\gamma_0} - h_\gamma$, and get $S(0, -1)$.
9. With the notation simplified by defining, e.g. $(\partial S/\partial \alpha)_{a_{\beta_0}}$ as $\partial S/\partial a_{\beta_0}$, then we may write

$$S(1, 0) - S(-1, 0) = 2h_\beta\left(\frac{\partial S}{\partial a_{\beta_0}}\right) + \sum_{j=1}^{n}\left(\frac{\partial S}{\partial x_{0j}}\right)\left[\delta x_{0j}^*(1) - \delta x_{0j}^*(-1)\right] + 0(h_\beta^3),$$

where the increments in **x** are denoted by the earlier convention.

We now embed the x-variation in the differences by the following approximations:

$$\begin{aligned} \frac{\partial S}{\partial a_{\beta_0}} &\cong \frac{1}{2h_\beta}\left[S(1, 0) - S(-1, 0)\right], \\ \frac{\partial S}{\partial a_{\gamma_0}} &\cong \frac{1}{2h_\gamma}\left[S(0, 1) - S(0, -1)\right], \\ \frac{\partial^2 S}{\partial a_{\beta_0}^2} &\cong \frac{1}{h_\beta^2}\left[S(1, 0) - 2S(0, 0) + S(-1, 0)\right], \\ \frac{\partial^2 S}{\partial a_{\gamma_0}^2} &\cong \frac{1}{h_\gamma^2}\left[S(0, 1) - 2S(0, 0) + S(0, -1)\right], \\ \frac{\partial^2 S}{\partial a_{\beta_0}\, \partial a_{\gamma_0}} &\cong \frac{-1}{2h_\beta h_\gamma}\left[S(1, 0) + S(-1, 0) + S(0, 1) + S(0, -1) - 2S(0, 0) - S(1, 1) - S(-1, -1)\right]. \end{aligned} \tag{17}$$

The above partial derivative expressions are all overtly standard (see Davis and Polonsky, 1964) except for the last one, which is a simple generalisation for unequal increments of the usual seven-point mixed partial formula. It may be readily verified by Taylor series expansion. Nevertheless, these expressions are used unconventionally in a most important respect. They have been written as partials for convenience; actually, however, the above description of our procedure

shows that the x_i are not held constant during the calculation of such quantities as $S(1, 0)$, $S(-1, 0)$, $S(1, 1)$, etc. Instead the x_i are changed as described whenever a parameter value is incremented. Thus, the three S's appearing in $(\partial^2 S/\partial a_{\beta_0}^2)$, for example, are all evaluated not only at different points in $\boldsymbol{\alpha}$ space but, implicitly, all at different points in $\mathbf{x}$ space as well. The new $\mathbf{x}$ is, of course, more compatible with the altered parameter value and is thus superior to the usual unaltered $\mathbf{x}$. The above expressions are truly partials with respect to other *parameters* but not with respect to $\mathbf{x}$. It is this crucial approximation which, we shall show, leads to vastly accelerated convergence.

In the actual programming of the algorithm, we found it to be more accurate to store each $S(q, r)$ in terms of its two components from (1), namely

$$S_y(q, r) \equiv \sum_{i=1}^{n} w_{y_i} R_{y_i}^2 ,$$

and

$$S_x(q, r) \equiv \sum_{i=1}^{n} w_{x_i} R_{x_i}^2 , \tag{18}$$

where

$$S(q, r) \equiv S_y(q, r) + S_x(q, r).$$

Then in the calculations of (17), we re-write all the formulas, grouping all the S_x's together and all the S_y's together. For example,

$$\frac{\partial S}{\partial a_{\beta 0}} = \frac{1}{2h_\beta} \{[S_y(1, 0) - S_y(-1, 0)] + [S_x(1, 0) - S_x(-1, 0)]\}.$$

and (19)

$$\frac{\partial^2 S}{\partial a_{\gamma 0}^2} = \frac{1}{h_\gamma^2} \{[S_y(0, 1) - 2S_y(0, 0) + S_y(0, -1)] + [S_x(0, 1) - 2S_x(0, 0) + S_x(0, -1)]\}.$$

This grouping tends to minimise errors resulting from taking differences between large numbers for one sum and then losing the significance of the other sum. We repeat the above procedure for all β and γ, where $1 \leq \beta, \gamma \leq m$. In the computer program we are careful not to compute the same quantity more than once even though the formal description of the procedure allows some unnecessary calculation. (For example, $S(0, 0)$ need be evaluated only once.)

We now have all the elements of $\mathbf{g}^{(1)}$ and $\mathbf{A}$ in (8). Hence, inversion of $\mathbf{A}$ and multiplication by $-\mathbf{g}^{(1)}$ will produce the desired $\boldsymbol{\delta}\mathbf{a}$. The corresponding $\boldsymbol{\delta}\mathbf{x}$ is found as a by-product of the calculation of $S(0, 0)$ above. (It is $\boldsymbol{\delta}\mathbf{x}^*$ for $q = r = 0$.)

The method described above for numerically calculating the 'partial' derivatives is far superior to using either exact derivatives or to a straightforward derivative approximation procedure because it results in local exploration of $(x, \boldsymbol{\alpha})$ space each time an $S(q, r)$ is found. It takes far more iterations for an analytic derivative method to converge in a typical problem. (See, e.g. Section 5, Table 4.) In addition, a straightforward numerical derivative approximation procedure will converge only about as rapidly as the analytic method.

After adjusting the parameters and the 'independent' variable by $\boldsymbol{\delta}\mathbf{a}$ and $\boldsymbol{\delta}\mathbf{x}$, we check to see if there is no significant change in the parameters and if the relative derivatives $\left|a_k \frac{\partial S}{\partial a_k}\right|$ are suitably small. If tests are failed, we perform another step in the iteration. If tests are successful, we consider the least squares solution to be found.

For estimates of the parameter variances we use

$$s_{a_k}^2 = \left(\frac{S}{n - m}\right) A_{kk}^{-1} , \tag{20}$$

which differs implicitly from the usual linearised form, Wolberg, 1967), since our procedure for calculating $\mathbf{A}$ includes the effects of variances in the x estimates.

We define the estimated variance of the least squares fit as

$$s_d^2 = \frac{S - \langle d \rangle^2 n}{(n - m)} \tag{21}$$

where $\langle d \rangle$ is the mean value of the composite weighted residuals (Macdonald 1969). The numerator (not S alone) is distributed according χ^2 with $(n - m)$ degrees of freedom.

4. Starting guesses

We decided to use Deming's iterative method to produce starting parameter guesses for the general procedure. There are several good reasons for this choice. First, Deming's method is easy to program and to use. Second, as will be seen later, Deming's method generally seems to produce parameter estimates close to the *least* squares result. Finally, in cases where $w_{x_i} \gg w_{y_i}$, Deming's method yields the same result as the general method. Note that as $w_{x_i} \to \infty$ the problem reduces to one of ordinary least squares with uncertainty only in the Y_i's. For convenience, we use the same relative increments in the Deming procedure as are later used in the general method.

5. Examples and comparison of methods

We have used the general iterative method on a variety of real problems, many of which involve quite strongly nonlinear models. Also, in order to test the power of the method, we have tried a number of the same cases considered by O'Neill *et al.* One of these cases was also used as an example by Southwell. In all the examples, the results of Deming's method were used as input parameter estimates (iteration 0). We employed $\Delta \equiv \Delta_x = \Delta_a = 10^{-4}$ or 10^{-5} and stopped the iteration when the relative change in each parameter was less than 10^{-7} and $\left|a_k \frac{\partial S}{\partial a_k}\right| < 10^{-7}$ for $1 \leq k \leq m$.

The first example employs a straight-line model, $y = \alpha_1 + \alpha_2 x$. As data, we use the set given by Pearson (1901) and listed in **Table 1.** The weights used are due to York (1966) and are presented in **Table 2.** We show in **Table 3** the results (to four or five significant figures) of the curve fit using the general iterative method for both single precision and double precision calculation*. It is clear from the initial logarithmic partial

Table 1 Pearson's data

i	X_i	Y_i	i	X_i	Y_i
1	0·0	5·9	6	4·4	3·7
2	0·9	5·4	7	5·2	2·8
3	1·8	4·4	8	6·1	2·8
4	2·6	4·6	9	6.5	2·4
5	3·3	3·5	10	7·4	1·5

Table 2 York's weights

i	w_{x_i}	w_{y_i}	i	w_{x_i}	w_{y_i}
1	1,000	1·0	6	80	20
2	1,000	1·8	7	60	70
3	500	4·0	8	20	70
4	800	8·0	9	1·8	100
5	200	20·0	10	1·0	500

*We carried out our calculations on a CDC 6400 computer with a maximum of 14 significant digits in single precision and a maximum of 29 significant digits in double precision.

Table 3 Results of general method using Pearson's data with York's weights
$(y = a_1 + a_2 x)$

(1) Single Precision ($\Delta = 10^{-4}$)

Iteration Number	a_1	a_2	S	$a_1 \frac{\partial S}{\partial a_1}$	$a_2 \frac{\partial S}{\partial a_2}$
0	5·3961	−0·46345	11·9564	$4{\cdot}2 \times 10^{-5}$	$1{\cdot}1 \times 10^{1}$
1	5·3982	−0·46388	11·9520	$9{\cdot}3 \times 10^{-4}$	$1{\cdot}0 \times 10^{1}$
2	5·4775	−0·47998	11·8665	$1{\cdot}0 \times 10^{-1}$	$8{\cdot}3 \times 10^{-1}$
3	5·4799	−0·48053	11·8664	$1{\cdot}8 \times 10^{-4}$	$1{\cdot}4 \times 10^{-3}$
4	5·4799	−0·48053	11·8664	$1{\cdot}3 \times 10^{-9}$	$4{\cdot}4 \times 10^{-10}$

(2) Double Precision ($\Delta = 10^{-4}$)

Iteration Number	a_1	a_2	S	$a_1 \frac{\partial S}{\partial a_1}$	$a_2 \frac{\partial S}{\partial a_2}$
0	5·3961	−0·46345	11·9564	$4{\cdot}2 \times 10^{-5}$	$1{\cdot}1 \times 10^{1}$
1	5·4774	−0·47995	11·8665	$1{\cdot}1 \times 10^{-1}$	$8{\cdot}8 \times 10^{-1}$
2	5·4799	−0·48053	11·8664	$2{\cdot}0 \times 10^{-4}$	$1{\cdot}5 \times 10^{-3}$
3	5·4799	−0·48053	11·8664	$5{\cdot}0 \times 10^{-10}$	$4{\cdot}7 \times 10^{-9}$

(3) Estimates of the standard deviations

Deming	$s_{a_1} = 0{\cdot}361$	$s_{a_2} = 0{\cdot}0707$	$s_d = 1{\cdot}222$
General Method	$s_{a_1} = 0{\cdot}252$	$s_{a_2} = 0{\cdot}0496$	$s_d = 1{\cdot}218$

derivative values of S given in the table that the least squares minimisation conditions are poorly satisfied by the Deming results. Note the $\sim 40\%$ decrease in s_{a_k} in going from Deming's parameter standard deviation estimates to our present estimates.

Table 4 indicates the number of 'iterations' used by various procedures to achieve the least squares solution. We place the word 'iteration' in quotes as a caution in the examination of Table 4, since each separate method uses a different iteration scheme. The table gives only a subjective comparison of convergence speeds. Southwell's iterative method does not converge at all in this case because his formula for δx was derived without the use of the necessary second partials. He does obtain the correct solution in his own work, however, when he eliminates x_i from (1) by solving $\frac{\partial S}{\partial x_i} = 0$ exactly. This can be done only when one has a model which is linear in the independent variable (the case Southwell considered).

Table 4 'Iterations' required to achieve minimum sum of squares, S, for Pearson's data with York's weights
$(y = a_1 + a_2 x)$

METHOD	NO. 'ITERATIONS'
Southwell	Did not converge
O'Neill	148
O'Neill (Aitken)	13
General Method with Exact Derivatives (Single Precision)	148
General Method (Single Precision)	4
General Method (Double Precision)	3
Minimum $S =$	11·866353
$a_1 =$	5·4799
$a_2 =$	−0·48053

In order to compare our method directly with that of O'Neill *et al.*, we programmed their algorithm (with the omission of their orthogonal polynomial expansion) for the CDC 6400. For the present straight-line model, the O'Neill algorithm run on the 6400 gave the same parameter estimates and sum of squares as those quoted by O'Neill *et al.* (for the ICT 1905 computer) at 125 iterations and also at 148, where convergence was reached. Thus, it appears likely that the orthogonal polynomial modification actually used by O'Neill *et al.*, makes no difference in the number of iterations required for convergence, and it seems clear that there were no important errors in either the original numerical work of O'Neill *et al.* or in our realisation of their method for the CDC 6400. Using Aitken's accelerative process, O'Neill *et al.* get convergence with 13 iterations. In our tests of the O'Neill *et al.* method, we found that the use of the Epsilon Algorithm (Macdonald, 1964) after every five steps produced convergence with only 10 iterations. Our present method does not need accelerative techniques since convergence typically occurs very rapidly.

Table 4 shows not only that the present general algorithm exhibits exceptionally rapid convergence but also that when it is used with exact analytical derivatives rather than with the special approximations of equations (17), it requires exactly the same large number of iterations as does the O'Neill *et al.* method. Further, even when conventional approximate derivatives (without the **x** incrementation implicit in our use of equations (17)) are used in the general method, one finds that, for a judicious choice of $\Delta_x = \Delta_a$, again 148 iterations are required for convergence. These results make it clear that indeed the gain in convergence speed in the general method arises completely from our unconventional derivative approximations. A full theoretical analysis and justification of our approach has recently been carried out (Jones), which completely supports our present claims. As already stated, the exploration of $(\mathbf{x}, \boldsymbol{\alpha})$ space in the present method compared to

Table 5 Results of general method using Pearson's data with unity weights on both X and Y
$(y = \alpha_1 + \alpha_2 x + \alpha_3 x^2 + \alpha_4 x^3)$

(1) Single Precision ($\Delta = 10^{-5}$)

Iteration Number	a_1	a_2	a_3	a_4	S
0	5·9988	−1·0050	0·15706	−0·01372	0·48677
1	6·0304	−1·0049	0·15255	−0·01319	0·48539
2	6·0152	−1·0000	0·15258	−0·01325	0·48515
3	6·0153	−0·9998	0·15247	−0·01324	0·48515

(2) Double Precision ($\Delta = 10^{-5}$)

Number Iteration	a_1	a_2	a_3	a_4	S
0	5·9988	−1·0050	0·15706	−0·01372	0·48677
1	6·0151	−0·9997	0·15242	−0·01324	0·48515
2	6·0153	−0·9998	0·15247	−0·01324	0·48515

(3) Estimates of the standard deviations

Deming	$s_{a_1} = 0{\cdot}361$	$s_{a_2} = 0{\cdot}404$	$s_{a_3} = 0{\cdot}128$	$s_{a_4} = 0{\cdot}0113$	$s_d = 0{\cdot}2848$
General Method	$s_{a_1} = 0{\cdot}265$	$s_{a_2} = 0{\cdot}298$	$s_{a_3} = 0{\cdot}0918$	$s_{a_4} = 0{\cdot}0080$	$s_d = 0{\cdot}2844$

Table 6 Results of general method using Pearson's data with unity weights on both X and Y
$(y = \alpha_1 + \alpha_2 x + \alpha_3 x^2 + \alpha_4 x^3 + \alpha_5 x^4 + \alpha_6 x^5)$

Iteration Number	a_1	a_2	a_3	a_4	a_5	a_6	S
0	5·924	−0·7407	0·02688	$-3{\cdot}324 \times 10^{-3}$	$2{\cdot}692 \times 10^{-3}$	$-3{\cdot}208 \times 10^{-4}$	0·45300
1	5·911	−0·5813	−0·1014	0·03377	$-1{\cdot}925 \times 10^{-3}$	$-1{\cdot}102 \times 10^{-4}$	0·45035
2	5·916	−0·6168	−0·06605	0·02114	$-5{\cdot}272 \times 10^{-5}$	$-2{\cdot}083 \times 10^{-4}$	0·45034
3	5·916	−0·6116	−0·07134	0·02302	$-3{\cdot}298 \times 10^{-4}$	$-1{\cdot}939 \times 10^{-4}$	0·45033
4	5·915	−0·6019	−0·08157	0·02677	$-8{\cdot}950 \times 10^{-4}$	$-1{\cdot}640 \times 10^{-4}$	0·45033
5	5·915	−0·6038	−0·07962	0·02607	$-7{\cdot}899 \times 10^{-4}$	$-1{\cdot}695 \times 10^{-4}$	0·45033
6	5·915	−0·6034	−0·08011	0·02624	$-8{\cdot}163 \times 10^{-4}$	$-1{\cdot}681 \times 10^{-4}$	0·45033
7	5·915	−0·6035	−0·08001	0·02621	$-8{\cdot}110 \times 10^{-4}$	$-1{\cdot}684 \times 10^{-4}$	0·45033
8	5·915	−0·6034	−0·08003	0·02622	$-8{\cdot}121 \times 10^{-4}$	$-1{\cdot}683 \times 10^{-4}$	0·45033
9	5·915	−0·6034	−0·08002	0·02621	$-8{\cdot}118 \times 10^{-4}$	$-1{\cdot}683 \times 10^{-4}$	0·45033
10	5·915	−0·6034	−0·08003	0·02622	$-8{\cdot}119 \times 10^{-4}$	$-1{\cdot}683 \times 10^{-4}$	0·45033

Estimates of standard deviations:

	s_{a_1}	s_{a_2}	s_{a_3}	s_{a_4}	s_{a_5}	s_{a_6}	s_d
Deming	0·416	1·38	1·30	0·465	0·0714	0·0038	0·3366
General Method	0·288	1·20	1·17	0·417	0·0623	0·0033	0·3355

only $\boldsymbol{\alpha}$ space in all other approaches, leads to a compensation of most of the error (which is *not* actually negligible, in general, as originally assumed) arising from the omission of the matrix **C** from the basic partitioned matrix **G**. This conclusion is made particularly plausible by the character of **C**, which itself directly involves $(\mathbf{x}, \boldsymbol{\alpha})$ space and is the only object in the analysis to do so except our unconventional derivatives. Crudely, what we omit when we ignore **C** we restore when we use generalised $(\mathbf{x}, \boldsymbol{\alpha})$ derivatives.

Table 5 presents the results of applying our method to a cubic model using Pearson's data with unit weights of both the X_i and Y_i. For the same problem, O'Neill *et al.* needed nine iterations for approximate convergence using their mixed

Table 7 *P-V* **data**

i	X_i	Y_i	i	X_i	Y_i
1	1·0	26·38	8	8·0	23·50
2	2·0	25·79	9	9·0	23·24
3	3·0	25·29	10	10·0	23·00
4	4·0	24·86	11	11·0	22·78
5	5·0	24·46	12	12·0	22·58
6	6·0	24·10	13	13·0	22·39
7	7·0	23·78	14	14·0	22·22

Table 8 Results of general method using *P-V* data and ME_1 model (Macdonald, 1969; Macdonald and Powell, 1971) $(y = \alpha_1(1 + \alpha_2\alpha_3 x)^{-1/\alpha_3})$

	Iteration Number	a_1	a_2	a_3	S
1. $w_{y_i} = 1{\cdot}0$, $w_{x_i} = 0$	(Ordinary least squares)				
	0	27·1125	33·7661	6·60017	0·0012872
2. $w_{y_i} = 1{\cdot}0$, $w_{x_i} = 1{\cdot}0$,	Single Precision, $\Delta = 10^{-4}$				
	0	27·1167	33·6446	6·62096	0·0011444
	1	27·1167	33·6427	6·62122	0·0011444
3. $w_{y_i} = 0{\cdot}02$, $w_{x_i} = 1{\cdot}0$,	Single Precision, $\Delta = 10^{-4}$				
	0	27·1544	33·5720	6·80419	0·012639
	1	27·1549	32·5607	6·80542	0·012616
	2	27·1549	32·5620	6·80524	0·012615
	3	27·1549	32·5618	6·80526	0·012615
4. $w_{y_i} = 0$, $w_{x_i} = 1{\cdot}0$,	Single Precision, $\Delta = 10^{-4}$				
	0	27·1546	32·5663	6·80517	0·012708
	1	27·1555	32·5478	6·80732	0·012685
	2	27·1555	32·5481	6·80729	0·012684
5. Estimates of standard deviations		s_{a_1}	s_{a_2}	s_{a_3}	s_d
$w_{y_i} = 1{\cdot}0$, $w_{x_i} = 0$	Deming	0·0178	0·511	0·0949	0·0108
	*GM	0·0125	0·359	0·0666	0·0108
$w_{y_i} = 1{\cdot}0$, $w_{x_i} = 1{\cdot}0$	Deming	0·0193	0·535	0·0966	0·01020
	*GM	0·0136	0·377	0·0680	0·01020
$w_{y_i} = 0{\cdot}02$, $w_{x_i} = 1{\cdot}0$	Deming	0·0296	0·667	0·0997	0·03390
	*GM	0·0223	0·503	0·0747	0·03387
$w_{y_i} = 0$, $w_{x_i} = 1{\cdot}0$	Deming	0·0297	0·667	0·0997	0·03399
	*GM	0·0214	0·483	0·0720	0·03396

*General Method

method (direct plus Aitken). The present method with exact or conventional approximate derivatives required 142 iterations instead of the two or three required with unconventional derivative approximations. Table 5 again shows a decrease of 35 to 40% in the s_{a_k}'s in going from Deming's estimates to our present estimates. In this example and in most cases we have examined, however, the actual parameter change, $\Delta a_k \equiv |(a_k)| - (a_k)$ General Method has been found to be less than (s_{a_k}) General Method. Thus, although our procedures assure us of a *least* squares solution, the results obtained are generally not very significantly different from Deming-method results if linearised s_{a_k} values can be trusted. With this last proviso, we may also conclude that a cubic is a poor model for Pearson's data. For $k = 3$ and 4, $|s_{a_k}/a_k| > 0{\cdot}6$, showing that a_3 and a_4 are ill-defined and not very significant.

Table 6, a quintic model, indicates that although the sum of squares, S, may stabilise quickly, occasionally it takes several more iterations to produce stability in the parameters. For this model, however, $|s_{a_2}/a_2| \cong 2$ and similar ratios for $k > 2$ are even much higher. Thus, here only one parameter is significant, and a quintic model is an extremely poor choice. The increase in s_d from 0·284 for the cubic model to 0·336 for the present quintic is further confirmation of this conclusion. Although a quintic would not be an acceptable model for the data used, the results are included to demonstrate convergence.

Table 7 gives some actual pressure-volume data. This is a preliminary, heretofore unpublished data set on *Kr* obtained by C. A. Swenson and M. Anderson (private communication) In **Table 8,** we present the results of assuming a three-parameter nonlinear model and using the data of Table 7 with four different weightings. Here all three parameter estimates appear to be highly significant. We have presented them to considerably more figures than the s_{a_k}'s warrant to show how the iterations proceed. Again the Δa_k's are insignificant, but we see that different weighting can result in very significant changes in parameter estimates.

6. Summary and conclusions

Curve fitting using data with errors in several measured variables frequently must be performed, often with strongly nonlinear models. We have presented a computer-oriented general iterative method designed to find least squares solutions without the necessity of user-supplied derivatives. In practice, we find extremely rapid convergence whenever ordinary nonlinear least squares (using, for example, the Marquardt algorithm) converges. As always in nonlinear situations, it is necessary sometimes to use starting guesses for parameters which are so far removed from the least-squares results that either convergence is not observed or it goes to a local minimum of the response surface. We recommend that in difficult cases the problem first be solved by ordinary nonlinear least squares (with $w_{x_i} \equiv 0$ and w_{y_i} arbitrary or unity), then the resulting estimated parameter values be used as starting guesses in the Deming algorithm (with w_{x_i} and w_{y_i} specified from the experi-

mental data), and finally the Deming results be used as starting guesses for the present algorithm. Further, in some experiments with synthetic, highly accurate data, we found that the values of Δ_x and Δ_a in the calculation of derivatives had to be chosen extremely small (e.g. $\sim 10^{-10}$) to obtain convergence. Using real data, however, we have not encountered such problems.

Our method is somewhat heuristic because of the great difficulty in performing numerically what 'exact' theory would demand. Comparison with previously reported general least squares procedures leads us to believe that the present method is more than competitive. We concur with O'Neill *et al.* in their conclusion that Deming's method will often provide adequate results without further calculation. Our iterative scheme, however, almost always produces a reduction of 10 to 40% or more in parameter standard deviation estimates when compared with ordinary linearized estimates, along with a lower sum of squares, guaranteed to be at least a relative minimum when our convergence criteria are satisfied. Even though the bias of nonlinear equation estimates may not always be negligible, we still feel that such reduction is useful. Whether or not an experimenter wishes more precision than Deming's method provides will depend upon his unique situation. Perhaps most important, he cannot be assured of a *least* squares solution with Deming's method, and there may be situations where significant differences arise in parameter estimates obtained by the two methods.

Acknowledgement

We wish to thank Dr. Eric L. Jones for his most valuable suggestions and discussions of this work. We much appreciate permission to use the preliminary *Kr* data of C. A. Swenson and M. Anderson.

References

Box, M. J. (1971). Bias in nonlinear estimation, *J. Roy. Stat. Soc., Series B*, Vol. 33, pp. 171-190.

Davis, P. J., and Polonsky, I. (1964). Numerical interpolation, differentiation, and integration, in *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series No. 55, US Government Printing Office, pp. 875-924.

Deming, W. E. (1943). *Statistical Adjustment of Data*, New York: Wiley.

Draper, N. R., and Smith, H. (1966). *Applied Regression Analysis*, New York: Wiley.

Hall, K. R., and Canfield, F. B. (1967). Optimal recovery of virial coefficients from experimental compressibility data, *Physica*, Vol. 33, pp. 481-502.

Jones, E. L. An efficient, second-order algorithm for the generalised nonlinear least squares problem, *The Computer Journal*, to be submitted.

Kopal, Z. (1955). *Numerical Analysis*, London: Chapman and Hall; see pp. 104-110.

Macdonald, J. R. (1964). Accelerated convergence, divergence, iteration, extrapolation, and curve fitting, *J. Appl. Phys.*, Vol. 35, pp. 3034-3041.

Macdonald, J. R. (1969). Review of some experimental and analytical equations of state, *Rev. Mod. Phys.*, Vol. 41, pp. 306-349.

Macdonald, J. R., and Powell, D. R. (1971). Discrimination between equations of state, *J. Res. Natl. Bureau of Standards, Series A*, Vol. 75A, pp. 441-453.

O'Neill, M., Sinclair, I. G., and Smith, F. J. (1969). Polynomial curve fitting when abscissas and ordinates are both subject to error, *The Computer Journal*, Vol. 12, pp. 52-56.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space, *Phil. Mag.*, Vol. 2, pp. 559-572.

Southwell, W. H. (1969). Fitting experimental data, *J. Comp. Phys.*, Vol. 4, pp. 465-474.

Wentworth, W. E. (1965). Rigorous least squares adjustment, *J. Chem. Ed.*, Vol. 42, pp. 96-103; 162-167.

Wolberg, J. R. (1967). *Prediction Analysis*, New Jersey: Van Nostrand.

York, D. (1966). Least-square fitting of a straight line, *Can. J. of Phys.*, Vol. 44, pp. 1079-1086.

Book review

Numerical Initial Value Problems in Ordinary Differential Equations, by C. William Gear, 1971; 253 pages. (*Prentice-Hall*, £6·50)

This text has 12 chapters on: 1, Introduction (and Euler Method); 2, Higher order one-step methods; 3, Systems of equations and equations of order greater than one; 4, Convergence, error bounds, and error estimates for one-step methods; 5, The choice of step size and order; 6, Extrapolation methods (of Bulirsch and Stoer, etc.); 7, Multivalue or Multistep methods—introduction; 8, General multistep methods, order and stability; 9, Multivalue methods; 10, Existence, convergence, and error estimates for multivalue methods; 11, Special methods for special problems (mainly stiff equations); 12, Choosing a method.

The treatment is in general good especially in the more practical Chapters 5, 6, 11 and 12. Complete FORTRAN routines are described and given for fourth order Runge Kutta with automatic step-length control, and for a variable order, variable step-length multivalue method with optional provision for stiff equations. A FORTRAN version of Bulirsch and Stoer's ALGOL procedure is also given. These large programs are photographically reproduced from computer printout. Smaller programs are type-set and much easier to read.

The discussion of one-step methods is shorter than that of Henrici (1962), but all the essential theoretical points are covered.

Multistep methods are treated from the less familiar multi*value* point of view (Gear, 1967), although this is scarcely used in Chapters 6 and 7. These multi*value* methods are exemplified by the finite difference and Nordsieck forms which are equivalent to the Adams-Bashforth-Moulton method. The stored quantities at each step are linear combinations of the usual y_r and $f(y_r, t_r)$ but the same approximating polynomial is used for all equivalent methods. Possible advantages are economy of arithmetic, especially when changing step length, and that the method may depend on a smaller number of previous steps, since the k stored values may approximate y_r and $f(y_r, t_r)$ for $0 \leqslant r \leqslant ((k-1)/2)$. Methods investigated include those using a fixed number of corrector iterations, as well as iteration to convergence. The only particular multistep or multivalue methods discussed are those of Milne and Adams-Bashforth-Moulton.

Chapter 10 includes a host of theorems relating the root condition, stability, consistency, convergence and asymptotic error form. The Dahlquist theory on the maximal order of stable multistep methods is also given. I did not find this theory easy to understand, partly because the author treats systems of pth order equations involving $(p-q)$ other derivatives, necessitating norms involving these two suffices, and partly because he does not always make clear exactly how the steps in the proof follow from the given hypotheses.

Rather than the usual one used by Gear, I prefer the (to me) more obvious definition of consistency of order r, for a pth order equation:

$$\Sigma \alpha_i z(t_{n-i})/h^p - \Sigma \beta_i f(z(t_{n-i})\ t_{n-i}) = z^{(p)}(t_n) - f(z(t_n), t_n) + O(h^r).$$

The requirement that $\Sigma \beta_i = 1$ and the index of the order h term are then automatic.

Very many stability concepts—asymptotic stability, absolute and relative stability regions, stiff A stability, and several others—are introduced and clearly explained. Figures 11.2 and 11.3 illustrate the difficulty of solving stiff equations, and the advantage of the backward Euler method very well.

I noticed rather more than the usual number of misprints, and in

Continued on page 169