JOURNAL OF COMPUTATIONAL PHYSICS 9, 273–289 (1972)

A Diverse Strategy Method of Minimization*

GEORGE A. BAKER, JR. AND D. L. HUNTER

Department of Applied Mathematics, Brookhaven National Laboratory
Upton, Long Island, New York

We break the problem of finding the minimum of a function down according to the type of surface encountered locally. Based on this surface classification we adopt that minimization strategy which ought to be most effective. We give a modified version of Krylov's eigenvector method, with a starting vector derived by a theorem of Lagrange and Beltrami. A convenient lower bound on the condition number of a positive definite matrix is obtained. We give a number of numerical examples of our procedure. The methods presented depend on the availability of analytic expressions for the first and second derivatives.

1. INTRODUCTION AND SUMMARY

The problem of finding the minimum point of a function of several variables is a classic computational problem. It occurs over and over again in computation problems of physics. To cite only a few examples, it occurs in nonlinear least-squares analysis, such as exponential decay curves over uneven intervals and phase shift analysis, and also the problem of the solution of a system of nonlinear equations can be reduced to the solution of a minimization problem. The difficulty of this problem is well known for, as Hamming [1] observed: "The more parameters that occur nonlinearly, the very much more the computation required to find the least squares fit. Experience shows that when the number of nonlinear parameters reaches four or five the process can be exceedingly painful and slow."

The classical methods for minimization fall into two broad categories. The first is fitting a simple form (usually quadratic) to the local surface, extrapolating a location of the minimum and jumping to that point. This procedure works well effectively doubling the accuracy at every step as long as one is "sufficiently close" to the minimum; however, the difficulty is that frequently it is extremely hard to get "sufficiently close" for this procedure to operate efficiently, or for that matter even to converge. The other category is traveling over the surface seeking at every step to reduce the norm of the error. Mesh stepping procedures, gradient methods,

* This work performed under the auspices of the U. S. Atomic Energy Commission.

and the refined problem parameter variation method of Davidenko [2] where by parameter changes a problem with a known solution is continuously distorted into the sought problem, belong to this other category. These methods have two principal disadvantages: they are not as efficient as the first category of methods near the minimum, and second, in common with the first category of methods, they become stuck on saddle points, where the surface is flat, that is zero slope, but decreases in some directions at a slower rate than linearly.

For the sort of problems we are interested in, refinement of an approximate solution is not a problem. The Newton–Raphson method of fitting the surface with a quadratic form and solving for the "bottom of the bowl" is perfectly adequate. The main problem is in getting "sufficiently close" (frequently one part in ten thousand is required) for Newton's method to be effective. Our idea is to employ a diverse strategy, and not to limit our action to any one fixed method or procedure. We analyze the local nature of the surface and act accordingly. The main distinction we make is based on whether the matrix of second derivatives is positive definite or indefinite.

In the second section we describe the various methods of minimization which we employ. These include several Newton–Raphson procedures, gradient methods and an eigenvector method. An improved version of Krylov's method is used to solve for the eigenvector and a theorem due to Lagrange and Beltrami is used to provide a suitable starting vector for Krylov's method. The reason why Runge–Kutta procedures are not helpful here is also discussed. Search procedures along a path over the surface, and extrapolative and interpolative minimization by interval halving are discussed.

In the third section we classify different types of surfaces which may be encountered and select a pattern of procedures which should be most likely to be successful in error reduction. We also prove a convenient lower bound on the condition number of a positive definite matrix.

In the final section we describe briefly the minimum problem we are interested in and discuss a range of examples. Included among these examples are two classical ones previously studied by other workers.

2. Basic Methods Employed

In this section we give a list of the basic methods of minimization which are used. An analysis of when they are used is deferred to the next section. We shall be concerned with that class of problems where the surface is twice continuously differentiable and these derivatives can be obtained analytically. The problems we are concerned with and the numerical examples we consider all have this property. All the procedures we discuss depend on this continuous differentiability

and minimization of a surface composed of a myriad of, say, plane-faces is not contemplated.

We define our surface by the function $S(\mathbf{x})$ where $\mathbf{x}$ is the position vector in N dimensional space with components x_i, $i = 1,..., N$.

The first method we employ is the Newton–Raphson procedure. That is, if one expands

$$S(\mathbf{x}) = S(\mathbf{y}) + \sum_{i=1}^{N} \frac{\partial S(\mathbf{y})}{\partial x_i}(x_i - y_i) + \frac{1}{2}\sum_{j=1}^{N} \frac{\partial^2 S(\mathbf{y})}{\partial x_i \, \partial x_j}(x_i - y_i)(x_j - y_j) + O(|\mathbf{x} - \mathbf{y}|^3), \tag{2.1}$$

then in this approximation one may solve for the increment $(\mathbf{x} - \mathbf{y})$ which moves us to the point where the gradient vanishes. That is to say

$$\sum_{j=1}^{N} \frac{\partial^2 S(\mathbf{y})}{\partial x_i \, \partial x_j}(x_j - y_j) = -\frac{\partial S(\mathbf{y})}{\partial x_i}. \tag{2.2}$$

In the approximation (2.1), the projected value at the Newton–Raphson point is

$$S(\mathbf{x}_{NR}) = S(\mathbf{y}) + \frac{1}{2}\sum_{i=1}^{N} \frac{\partial S(\mathbf{y})}{\partial x_i}(x_i - y_i) + O(|\mathbf{x} - \mathbf{y}|^3). \tag{2.3}$$

Consequently, this procedure is not apt to be useful in finding a minimum if the dot product of the increment vector with the gradient vector indicates a step in the uphill direction. We have solved Eq. (2.2) by pivotal Gaussian elimination. A more refined procedure was not thought worthwhile for two reasons. First, we will check the condition of the second derivative matrix before deciding on what procedure to use and in the case of poorly conditioned matrices will take other action in the first instance. Furthermore, it will be the length of the projection of the increment vector along the near zero-eigenvalue eigenvector(s) which will be ill-defined. In all but the simplest case of a purely quadratic surface, one would hardly suppose that such a relatively long step would, even if accurately found, provide a good indication of the location of the minimum.

In addition to the Newton–Raphson or bottom-of-the-bowl method, we have found restricted Newton–Raphson procedures to be useful. That is, we apply the Newton–Raphson procedure not in the full N-dimension position vector space, but in a subspace. This procedure is a combination of mesh-stepping techniques and jumping to a (restricted) minimum. The idea is, if one cannot achieve a minimum in all variables at one stroke, perhaps if some are held fixed, one can minimize with respect to the remainder by solving a problem of reduced dimensionality (and complexity). The following iteration would then move off in

a new direction after a minimum in the subspace had been located. In the particular class of problems we are concerned with, we found the use of two subspace restrictions advantageous as we had three natural classes of variables. The smallest subset is composed of one-third of the variables and the surface function is a pure quadratic form in terms of them.

In all the Newton–Raphson techniques, and indeed in all the procedures which we will employ, we are not content to simply try the computed step. First we test it against the initial value of the error for that iteration, and insist that the error norm be reduced; that is, we will not take a step which increases the value of $S(\mathbf{x})$. Secondly, as for our problems the minimum value of S is zero, we ask whether the value of S has been reduced by an arbitrary factor, say, two. If it has not, we institute a search procedure. We take further steps in the same direction equal to the original step until the error stops getting smaller. If more than ten steps are required, we increase the step length by a factor of ten at each step until we have passed the minimum. We then interpolate by interval halving until the relative minimum in that direction is found with the desired precision. If the value at the midpoint of the interpolation interval is greater than the smaller endpoint by more than one-fourth the amount the largest endpoint is, then quadratic interpolation projects a minimum outside the interpolation interval. We then compute the function value one half step beyond the small endpoint and if that value is smaller than that at the midpoint, we move to the extended half interval as our next smaller interpolation interval. In order to reduce the number of function evaluations required in the interpolation procedure, the convergence of the second difference approximation to the second derivative in that direction is monitored and a jump to the probable minimum is made as soon as the predetermined convergence criterion is met.

We have also employed what we call the second-order gradient method. The ordinary gradient method steps off down the gradient [Householder, 3] and searches for a minimum in that direction. However, if we wish to truely follow a path of steepest descents, we must solve the system of differential equations

$$\frac{dx_i}{dt} = -\frac{\partial S(\mathbf{x})}{\partial x_i}, \qquad i = 1,..., N. \tag{2.4}$$

If one solves these equations to low order in t, one can obtain second-order accuracy from the gradient and second-order derivative matrix, as

$$\Delta x_i = -\frac{\partial S}{\partial x_i} t + \frac{1}{2} \sum_j \frac{\partial^2 S}{\partial x_i \partial x_j} \frac{\partial S}{\partial x_j} t^2 + O(t^3), \qquad i = 1,..., N. \tag{2.5}$$

The second-order gradient method then consists of searching along the *curve* (2.5) for the minimum value of S. The initial step length is chosen as the positive real t

for which (2.5) substituted in (2.1) and carried to second order only in t leads to $S = 0$. If there is no such t, then (2.1) is carried only to first order in t. The search is carried out as described above, with due modifications to take account of the curved path of search.

We have also investigated use of the ordinary gradient procedure, but in those circumstances in which it is useful the second-order gradient is generally a more effective method, and is very little lengthier in time, given that we already have the matrix of second derivatives available. Further, we have considered the direct integration of (2.4) by a Runge–Kutta method in an effort to follow the path of steepest descents. While such a procedure is possible, it has, in our experience, been a relatively long and inefficient method compared to the other procedures we discribe. Furthermore, if (2.4) is used directly on a parabolic region $t \to \infty$ is required to reach the minimum. That difficulty can be overcome by using arc length in position vector space instead of t as an independent variable. Nevertheless on sample cases we have tried, the step size required to keep the integration going downhill instead of uphill was much too small to be competitive.

The final method we have employed we call the eigenvector method. In order to explain the rationale behind this method suppose that we can approximate the surface locally by a quadratic plus a linear form. If we have one (or more) negative eigenvalue of the matrix of second derivatives, then not only can we go down hill by proceeding in any direction making an angle less than 90 degrees with the negative gradient, but if we go in the direction of an eigenvector with negative eigenvalue then the slope becomes steeper as we proceed. If the surface is globally quadratic, the function value in that direction must decrease asymptotically to minus infinity, not only in the downhill eigenvector direction but also in the uphill eigenvector direction as well. Greenstadt [12] has previously suggested the use of eigenvector directions. Fiacco and McCormick [4] discuss under the heading of "Modified Newton Method" the employment of approximate eigenvector steps in the case of an indefinite second-derivative matrix. Pearson [5] in his fifth algorithm investigates essentially the same procedure. The type of approximate eigenvectors they use are very closely related to those we discuss below at Eqs. (2.18)–(2.21). We have elected to compute accurately the eigenvector corresponding to the minimum eigenvalue. For many problems this procedure may be unduly elaborate and could, for those problems, be dispensed with; however, for our problems we have found that the increase in the magnitude of the negative eigenvalue, relative to that of the abovementioned approximate eigenvector, was not uncommonly by a factor of 10–100 with a corresponding improvement in the rate of error reduction.

In order to solve the eigenvector problem we have used Krylov's method [6] with a special iterative improvement method. This method uses the characteristic polynomial and, as we develop it, will be reliable only for the minimum eigenvalue and eigenvector and not for other eigenvalues or eigenvectors. Even complex roots

of our characteristic polynomial can appear for symmetric matrices, though they are, of course, spurious. These shortcomings are of no concern to us as they only effect aspects irrelevant to our needs. Other eigenvector methods may be superior, but this one is the best of those we have tried for this problem. The mathematics of Krylov's method is extremely similar to that for Padé approximants [see, Ref. 7 for a review]. The method is briefly as follows. Suppose we seek the eigenvector for **B**. We define

$$\mathbf{A} = \mathbf{B} - 0.6u\mathbf{I}, \tag{2.6}$$

where

$$u = \left[\sum_{i,j} | b_{ij} |^2\right]^{1/2} \tag{2.7}$$

is a well-known upper bound to the eigenvalues of **B** and **I** is the identity matrix. This has the effect of making the sought eigenvalue the one of largest magnitude. Define a sequence of vectors by the relation

$$\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n . \tag{2.8}$$

If we expand

$$\mathbf{x}_1 = \sum_{i=1}^{N} \alpha_i \boldsymbol{\xi}_i , \tag{2.9}$$

where $\boldsymbol{\xi}_i$ are eigenvectors of **A** with eigenvalues λ_i, then

$$\mathbf{x}_{n+1} = \sum_{i=1}^{N} \alpha_i \lambda_i{}^n \boldsymbol{\xi}_i . \tag{2.10}$$

If we now define the coefficients

$$c_n = \mathbf{x}_{n+1} \cdot \mathbf{x}_1 = \sum_{i=1}^{N} \alpha_i{}^2 \lambda_i{}^n, \tag{2.11}$$

then the auxiliary function

$$C(z) = \sum_{n=0}^{\infty} c_n z^n \tag{2.12}$$

sums to

$$C(z) = \sum_{i=1}^{N} \frac{\alpha_i{}^2}{1 - \lambda_i z} , \tag{2.13}$$

which is of the form of the ratio of a polynomial of degree $N-1$ over one of degree N. This form is exactly the $[N, N-1]$ Padé approximant to (2.12). The denominator polynomial which is the characteristic polynomial is then given by the standard formula [7]

$$Q(z) = \det \begin{vmatrix} c_0 & c_1 & \cdots & c_N \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ c_{N-1} & c_N & \cdots & c_{2N-1} \\ z^N & z^{N-1} & \cdots & 1 \end{vmatrix}. \tag{2.14}$$

(The polynomial in practice is more conveniently evaluated by the solution of an equivalent set of linear equations or through recursion relations [7].) The roots of $Q(z)$ are found by Newton–Raphson and Bairstow iterations, the smallest one, λ_s selected, and

$$\mathscr{E}_s(z) = \frac{Q(z)}{1 - \lambda_s z} = \sum_{j=0}^{N-1} e_j z^j \tag{2.15}$$

formed. Now as $\mathscr{E}_s(\lambda_i^{-1}) = 0$ for $\lambda_i \neq \lambda_s$, we must have

$$\sum_{j=0}^{N-1} e_{N-1-j}\mathbf{x}_{j+1} = \alpha_s \lambda_s{}^N \mathscr{E}(\lambda_s^{-1})\, \xi_s \tag{2.16}$$

as that linear combination of the first N $\mathbf{x}$'s which points in the direction of ξ_s. In order to check the accuracy, the vector computed in (2.16) is tested in the Rayleigh quotient

$$\lambda_s = \langle \xi_s \mid \mathbf{A} \mid \xi_s \rangle / \langle \xi_s \mid \xi_s \rangle \tag{2.17}$$

and if the value of λ_s from (2.17) agrees with the root of $Q(z)$, then it is accepted. Otherwise the process is repeated using ξ_s of (2.16) as a new $\mathbf{x}_1$. This procedure increases the prominence of the smallest eigenvalue and although by making α_s much larger than the other α's we may decrease their accuracy, we are unconcerned about this occurrence as we seek only the minimum. A slightly annoying aspect occurs when one of the poorly determined roots moves onto the negative real axis below the real singularities. We counter this difficulty by checking (for negative eigenvalues of **B**) all the real roots between $-1.6u$ and $-0.60u + v$ (u from 2.7 and v = minimum of zero and 90% of the smallest Rayleigh quotient so far obtained) and selecting that eigenvector with the smallest Rayleigh quotient as ξ_s for the next starting vector. We pick the above range of roots as we are only interested in finding negative or "zero" eigenvalues of **B**. In practice, one or two applications of Krylov's method is usually sufficient.

Having determined the eigenvector, we must now choose one of two possible directions and a step length. In the case that the smallest eigenvalue of **B** is negative, we compute two steps, one forward and one backward by the condition that S vanish in quadratic approximation and a search (in each direction) is carried out as before. Note that this procedure represents a possible departure from strict steepest descents as it can conceivably look over the brow of a hill to see a better region on the other side. If the smallest eigenvalue of **B** is positive, we pick a step length which would make S vanish in linear approximation. This procedure and it alone of those we have tried is very effective in escaping from saddle points. Saddle point trapping is a characteristic difficulty of all methods whose step length or direction is calculated from the gradient.

In the above discussion of Krylov's method, the choice of the initial vector was not specified. Plainly a vector more parallel to the sought vector than perpendicular to it is desirable. The surfaces we have investigated have produced particularly difficult matrices in the saddle shaped regions. The sort of matrix encountered has one very small, relative to the largest positive eigenvalue, negative eigenvalue plus several small positive eigenvalues. The other eigenvalues are distributed up the real axis to the largest. This eigenvalue distribution makes the determination of a vector which produces a negative Rayleigh quotient very difficult by many of the usual methods. We have adopted a procedure, based on the Lagrange–Beltrami Theorem [8], which produces a vector whose Rayleigh quotient has the same sign as that of the smallest eigenvalue of **B**. The Lagrange–Beltrami Theorem states that if none of the determinants

$$\det | \mathbf{B}_k | = \det \begin{vmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{k1} & \cdots & b_{kk} \end{vmatrix}, \tag{2.18}$$

$k = 1, 2, \ldots, n-1$, are zero then the quadratic form

$$Q(x) = \sum_{i,j}^{n} x_i b_{ij} x_j \tag{2.19}$$

can be expressed as

$$Q(x) = \sum_{k=1}^{n} \frac{\det | \mathbf{B}_k |}{\det | \mathbf{B}_{k-1} |} y_k^2, \tag{2.20}$$

where $\det | \mathbf{B}_0 | = 1$ and

$$y_k = x_k + \sum_{j=k+1}^{n} a_{kj} x_j, \qquad k = 1, 2, \ldots, n. \tag{2.21}$$

The ratio of determinants in (2.20) are just the diagonal elements of the triangularized form of **B**. This result can be seen from the facts that the operations involved in triangularization leave the determinant unchanged and that the coefficient of $b_{kk}y_k^2$ in (2.20) is unity. Now from (2.20) it follows at once that it is necessary and sufficient for $Q(x)$ to be positive definite, that is to have all positive eigenvalues, for all the determinants in (2.18) to be positive. Consequently, if there are negative eigenvalues, at least one determinant of (2.18) must be negative. Thus our procedure is to triangularize **B**, and find the minimum diagonal element (less than zero when there is a negative eigenvalue). Say this is the coefficient of y_l^2 in (2.20). We then construct by Gram–Schmidt orthogonalization that vector in the l dimensional subspace spanned by $x_1, ..., x_l$ which is perpendicular to $y_1, y_2, ..., y_{l-1}$; that is, equivalently, to the first $l - 1$ rows of either **B**, or the triangularized version of **B**. This vector necessarily has a Rayleigh quotient of the same sign as the sign of the minimum eigenvalue. We use it as a starting vector for the Krylov procedure.

3. Classification of Surface Types

In order to select an appropriate procedure or sequence of procedures from those described in the previous section, we first classify the nature of the surface into twelve logical categories (six of which cannot occur for the problems with which we are most concerned) and describe the action to be taken in each case. We base the classification scheme on the first two derivatives. The first derivative can be either (α) of zero magnitude or (β) of nonzero magnitude. By "zero" we mean the larger of the limit set by round-off in its calculation or small enough to predict a Newton Raphson step of length less than a preassigned error tolerance. The second derivative matrix can be (a) strictly positive definite—that is, all eigenvalues are greater than zero; (b) nonnegative definite—that is, all eigenvalues are greater than zero except there is at least one zero eigenvalue; (c) Saddle-point—that is, some eigenvalues are positive and some are negative; (d) Flat—that is, all eigenvalues are zero; (e) nonpositive definite—that is, all eigenvalues are negative except that at least one is zero; (f) strictly negative-definite—that is all eigenvalues are negative. Taking all possible combinations of types of first and second derivatives leads to 12 logical categories for the local nature of the surface. For the cases we will be interested in $\partial^2 S/\partial x_i^2 > 0$, and hence the sum of the eigenvalues is greater than zero and so cases (d)–(f) cannot arise. However, we will discuss them anyway for completeness.

By a zero eigenvalue we mean one which is small in magnitude in comparison to the eigenvalue of largest magnitude. We use as a criterion, the expected error in the Newton–Raphson method. This error can be related to the problem of inverting the second derivative matrix. We require that enough accuracy remain to reproduce

the sign and magnitude of the change in S due to a Newton–Raphson step. We use [Marus, 9] the result

$$\Lambda(\mathbf{A}(\mathbf{A}^{-1}) - I) \leqslant 14.24\, P(\mathbf{A})\, n^2 \beta^{-s}, \tag{3.1}$$

where $(\mathbf{A}^{-1})$ is the approximate inverse to the $n \times n$ matrix $\mathbf{A}$ generated by Gaussian elimination using s places on a scale of β, $P(\mathbf{A}) = |\lambda_{\max}|/|\lambda_{\min}|$ is a condition number for $\mathbf{A}$, and $\Lambda(\mathbf{X})$ is the largest eigenvalue of $\mathbf{X}$. If we require our estimate of Λ to be less than 1/2 [from (2.1)], say 1/7, then we would say that $\mathbf{A}$ has for our purposes a "zero" eigenvalue, when

$$P(\mathbf{A}) \geqslant \frac{\beta^s}{100n^2}. \tag{3.2}$$

It is convenient for our purposes to have a method to estimate $P(\mathbf{A})$, the condition number $\lambda_1(\mathbf{A})/\lambda_n(\mathbf{A})$ of an $n \times n$, positive-definite matrix $\mathbf{A}$ without having to solve for all the eigenvalues or invert $\mathbf{A}$, where we define

$$\lambda_1(\mathbf{A}) \geqslant \lambda_2(\mathbf{A}) \geqslant \cdots \geqslant \lambda_n(\mathbf{A}) > 0. \tag{3.3}$$

To this end we will establish an estimate which is a lower bound for $P(\mathbf{A})$. First let us note two known inequalities [Beckenbach and Bellman, 8]. First, by the Cauchy–Poincaré Separation Theorem

$$\lambda_s(\mathbf{A}) \geqslant \lambda_s(\mathbf{B}) \geqslant \lambda_{s+r}(\mathbf{A}), \qquad s = 1,\ldots, n - r, \tag{3.4}$$

where $\mathbf{B}$ is the projection of $\mathbf{A}$ on any subspace of dimension $n - r$. Secondly, by "An Inequality Concerning Minors" [8]

$$\det |\, \mathbf{A}_{1n} \,| \leqslant \det |\, \mathbf{A}_{1k} \,| \det |\, \mathbf{A}_{k+1,n} \,|, \tag{3.5}$$

where the determinants are defined by

$$\det |\, \mathbf{A}_{rs} \,| = \det \begin{vmatrix} a_{rr} & \cdots & a_{rs} \\ \vdots & \ddots & \vdots \\ a_{sr} & \cdots & a_{ss} \end{vmatrix}. \tag{3.6}$$

Since a determinant is equal to the product of the eigenvalues

$$\frac{\det |\, \mathbf{A}_{1,k+1} \,|}{\det |\, \mathbf{A}_{1,k} \,|} = \frac{\prod_{s=1}^{k+1} \lambda_s(\mathbf{A}_{1,k+1})}{\prod_{s=1}^{k} \lambda_s(\mathbf{A}_{1,k})} = \left\{ \prod_{s=1}^{k} \frac{\lambda_s(\mathbf{A}_{1,k+1})}{\lambda_s(\mathbf{A}_{1,k})} \right\} \lambda_{k+1}(\mathbf{A}_{1,k+1})$$

$$= \lambda_1(\mathbf{A}_{1,k+1}) \left\{ \prod_{s=1}^{k} \frac{\lambda_{s+1}(\mathbf{A}_{1,k+1})}{\lambda_s(\mathbf{A}_{1,k})} \right\}. \tag{3.7}$$

Thus, by the Cauchy–Poincaré separation theorem

$$\lambda_1(\mathbf{A}) \geqslant \lambda_1(\mathbf{A}_{1,k+1}) \geqslant \frac{\det \mid \mathbf{A}_{1,k+1} \mid}{\det \mid \mathbf{A}_{1,k} \mid} \geqslant \lambda_{k+1}(\mathbf{A}_{1,k+1}) \geqslant \lambda_n(\mathbf{A}). \tag{3.8}$$

Thus by taking the ratio of the maximum to the minimum of the determinant ratios, we can obtain a lower bound for P. We can improve the lower bound for $\lambda_1(\mathbf{A})$ because by (3.4) and (3.5)

$$\lambda_1(\mathbf{A}) \geqslant a_{jj} \geqslant \frac{\det \mid \mathbf{A}_{1,j} \mid}{\det \mid \mathbf{A}_{1,j-1} \mid}, \qquad j = 2,..., n. \tag{3.9}$$

Thus the maximum of the a_{jj} is a better lower bound for $\lambda_1(\mathbf{A})$. Now if we triangularize a matrix by adding multiples of rows to other rows, the determinant is unchanged. Thus it follows easily that if $\mathbf{T}$ = triangularization of $\mathbf{A}$, its diagonal elements are given by

$$t_{11} = a_{11}, \qquad t_{jj} = \frac{\det \mid \mathbf{A}_{1,j} \mid}{\det \mid \mathbf{A}_{1,j-1} \mid}, \qquad j = 2,..., n. \tag{3.10}$$

Hence we have

$$P(\mathbf{A}) = \frac{\lambda_1(\mathbf{A})}{\lambda_n(\mathbf{A})} \geqslant \frac{\max_{1 \leqslant j \leqslant n}[a_{jj}]}{\min_{1 \leqslant j \leqslant n}[t_{jj}]} \tag{3.11}$$

for a positive definite matrix. It is to be noted that in this estimate the actual calculations are nearly the same as in the trial vector for starting our eigenvector procedure.

We tabulate in Table I the 12 logical categories and what seems to us the best initial procedure. The categories 7–12 cannot occur for our problems but we have listed them for completeness.

On the basis of Table I we have simplified the classification of the surface to three types, based on the second derivative alone. They are: (1) positive definite; (2) indeterminate, including nonnegative definite and flat; and (3) indefinite, if there is at least one negative eigenvalue.

If the surface is positive definite, a Newton–Raphson procedure is tried. If the error is not reduced by a minimum amount (we choose by a factor of two), a search along the Newton–Raphson direction is tried. If the error is still not reduced by a factor of two, the two, subspace, Newton–Raphson procedures (with search as required, except for the purely quadratic subspace) are tried. Finally, the second-order gradient method is tried, if the error has still not been reduced by a factor of two. That procedure is then chosen which was most successful in reducing the error.

TABLE I

Surface Type Strategy Table

Category			Surface type	Probable best procedure
1	α	*a*	zero slope, pos. def.	solution, quit
2	β	*a*	nonzero slope, pos. def.	Newton–Raphson
3	α	*b*	zero slope, nonneg. def.	If no improvement possible with eigenvector step, must be a solution.
4	β	*b*	nonzero slope, nonneg. def.	Newton-Raphson procedure should fail on zero determinant, so use an eigenvector step or second order gradient step.
5	α	*c*	zero slope, saddle point	eigenvector step
6	β	*c*	nonzero slope, saddle point	eigenvector step or second order gradient step
7	α	*d*	zero slope, flat	Higher order method required
8	β	*d*	nonzero slope, flat	second order gradient
9	α	*e*	zero slope, nonpos. def.	eigenvector step
10	β	*e*	nonzero slope, nonpos. def.	eigenvector step
11	α	*f*	zero slope, neg. def.	eigenvector step
12	β	*f*	nonzero slope, neg. def.	eigenvector step

If the surface is indefinite, the eigenvector procedure is tried first. If the error is not reduced by a factor of two, the second-order gradient procedure is tried, and finally the three Newton–Raphson procedures. If the error has not been reduced by a factor of two, the most successful procedure at error reduction is selected.

If the surface is indeterminate, the same sequence of steps is followed as when it is indefinite, except that the iteration procedure can now terminate on a full-space Newton–Raphson step whereas it cannot terminate unless the error has been reduced to the roundoff limit if the surface is indefinite.

4. Numerical Examples

The physical and mathematical problem in which we are mainly interested arises in the statistical mechanics of the critical point and is a problem of approximate analytic continuation. At the critical point, many of the thermodynamic properties become singular. These physical singularities are reflected in singularities in the

mathematical functions which describe these properties. Extensive, exact power series expansions are known in many instances for these functions. The problem which presents itself is to determine A_i, y_i, γ_i, $i = 1,\ldots, N$, such that a prescribed set of power series coefficients of the function

$$g(x) = g_0 + \sum_{i=1}^{N} A_i[(1 + xy_i)^{\gamma_i} - 1]/\gamma_i = \sum_{j=0}^{\infty} g_j x^j \tag{4.1}$$

agree with those of the known expansion $f(x)$. We have chosen for our surface function

$$S(x) = \sum_{j=k+1}^{k+3N} (g_j - f_j)^2, \tag{4.2}$$

where the f_j are given and the g_j can be given explicitly as

$$g_j = \sum_{i=1}^{N} \frac{A_i}{\gamma_i} \binom{\gamma_i}{j} y_i^j, \tag{4.3}$$

where $\binom{a}{b}$ is the usual binomial coefficient. The resulting minimum problem turns out to be a very difficult one in the sense that the surface changes very rapidly in terms of the parameters and condition numbers of 10^9 are quite common for the matrix of second derivatives.

We can conveniently label our test cases by the given functions $f(x)$ or $S(x)$. The surface function $S(x)$ for the first 6 tests is derived from the $f(x)$ quoted by means of (4.2). They are:

Test Series No. 1:

$$f(x) = (1 - x)^{-2} + 2(1 - 2x)^{-1}. \tag{4.4}$$

Test Series No. 2:

$$f(x) = (1 - 2x)^{-1.1} + (1 + 2.5x)^{-1}. \tag{4.5}$$

Test Series No. 3:

$$f(x) = (1 - 2x)^{-1.1} + (1 + 2.5x)^{-1} + 0.8(1 - 2.5x)^{-1}. \tag{4.6}$$

Test Series No. 4:

$$f(x) = (1 - x)^{-1} + \tfrac{1}{2}(1 - x)^{-0.5}. \tag{4.7}$$

Test Series No. 5:

$$f(x) = (1 + 2ix)^{-1} + (1 - 2ix)^{-1}. \tag{4.8}$$

TABLE II

Numerical Results

Test	Starting point	Iterations
1.1	0.1 %	20
1.2	1.0 %	47
1.3	1.0 %	61
1.4	100 %	54[a]
2.1	1.0 %	11
2.2	10. %	22
3.1	0.5 %	608
3.2	0.1 %	178
4.1	1.0 %	264
4.2	0.1 %	91
5.1	0.5 − 1.0 %	9
5.2	1.0 %	11
5.3	1.0 %	9
5.4	10. %	16
6.1	0.1 %	5
6.2	1.0 %	10
6.3	5.0 %	30
7.1	1.0, −1.2, +5.0	18
8.1	15.0, −2.0, +5.0	6[b]
8.2	15.0, 2.2302, +5.0	7[b]
8.3	15.0, 3.0, 5.0	7
8.4	15.0, 7.0, 5.0	9

[a] Converges to a different solution.
[b] Converges to the relative minimum.

Test Series No. 6:

$$f(x) = 2(1 - 0.9x)^{-1.25} + (1 + \omega x)^{-1} + (1 + \omega^2 x)^{-1} - 4(1 + x)^{-0.5}, \quad (4.9)$$

where ω is one of the cube roots of unity.

Test No. 7 (Rosenbrock's notorious, banana-shaped valley, Rosenbrock [10]):

$$S(A, y, \gamma) = 100(A - y^2)^2 + (1 - y)^2 + (1 - 4\gamma + \gamma^2)^2. \quad (4.10)$$

The last term is a decoupled extra term to make the number of variables divisible by 3 and thereby avoid extensive modification of our computer program.

Test No. 8 (Polynomial Equation of Freudenstein and Roth [11]):

$$\begin{aligned} S(A, y, \gamma) = {} & [-13 + A + ((-y + 5)y - 2)y]^2 \\ & + [-29 + A + ((y + 1)y - 14)y]^2 + (1 - 4\gamma + \gamma^2)^2. \end{aligned} \quad (4.11)$$

The same remark applies about the last term in (4.11) as about the last term in (4.10).

In Table II we have tabulated the results of our experience with these procedures. They were programmed for the Brookhaven CDC 6600, a 60-bit machine. The first digit in the test number refers to the series being tested. The starting points differ from the exact solution by the quoted percentage in every parameter. The behavior for the first series is roughly as follows. Initially the surface is indefinite and a variety of procedures are employed and after a number of iterations a positive definite region is reached in which Newton–Raphson with search is the main procedure. Progress continues at a modest rate until the parameters are about 0.01 % off the solution at which point the bowl shaped region about the solution is entered and the Newton–Raphson procedure behaves in textbook fashion dramatically decreasing the error with every application. In test 1.4, a far-off start leads to a different solution

$$\begin{array}{lll} A_1 = 3.903004 & y_1 = -0.52121305 & \gamma_1 = -0.86745371 \\ A_2 = -3.7493483 & y_2 = -2.1428515 & \gamma_2 = -0.39286949 \end{array} \quad (4.12)$$

thereby proving that these equations do not necessarily have a unique solution.

The behavior on test series 2 is very similar, except that it is a much easier case, presumably because neither singularity completely dominates the other in low order. We stop our iterations if the Newton–Raphson step is less than 10^{-7} or the error has been reduced by twenty-seven orders of magnitude below that for all $A_i = 0$. This stopping rule is used for all the cases reported.

The third test series is a particularly difficult one as over 90% of the iterations are spent with a condition No. $\geqslant 10^9$. Whether the number of iterations could be cut substantially by the use of more significant figures is unknown. Again, a variety of procedures is used, the Newton–Raphson with search being the most frequently selected procedure in the indeterminate region. The final error of the nine parameters was at worst one in the sixth figure when iteration was terminated according to either criterion.

The fourth test series has two confluent singularities. Most of the iterations were spend in an indefinite region using a variety of procedures. The solution, bowl-shaped region was finally found when the parameters were only off by a few parts in 10^6.

Test series 5 has two complex conjugate singularities and shows that the same procedures described can be applied to real series with complex singularities with appropriate complex conjugates in the formulas. All of the procedures used must maintain a complex conjugate pair of singularities as complex conjugates; however, due to round-off we have found it desirable to continually enforce this condition so as not to get off course.

Test series 6 shows that there must be other considerations than simply the number of singularities as there are 12 parameters (8 nonlinear) in this case and speedy convergence was obtained.

Test 7, Rosenbrock's banana shaped valley was started at the classic starting point, and simply followed the curving valley to its minimum point at $A = y = 1.0$.

In Test 8, minimum points of the surface corresponding to the polynomial equations of Freudenstein and Roth were easily obtained by our procedures. In test 8.1 the relative minimum at

$$A = 11.412778987, \qquad y = -0.89680525327 \tag{4.13}$$

instead of the absolute minimum at

$$A = 5.0, \qquad y = 4.0, \qquad (\gamma = 3.732050876) \tag{4.14}$$

was found. In Test 8.2, the program demonstrated the power of the eigenvector procedure to find a steeper slope by looking over the brow of a hill. This test surface has a saddle point at $y = \frac{1}{8}(2 + \sqrt{22}) \approx 2.23014$, and A corresponding. By starting with $y = 2.2302$ and minimizing with respect to A first we arrive on the solution side of a saddle point. The eigenvector procedure is able to look over the saddle point and make a greater initial improvement by going down the steeper slope toward the relative minimum than by going along the more gradual slope to the real minimum so convergence is obtained to the relative minimum.

REFERENCES

1. R. W. HAMMING, "Numerical Methods for Scientists and Engineers," p. 248, McGraw-Hill, New York, 1962.
2. D. F. DAVIDENKO, On a new method of numerical solution of systems of nonlinear equations. *Doklady Akad. Nauk SSSR (N. S.)* **88** (1953), 601–602.
3. A. S. HOUSEHOLDER, "Principles of Numerical Analysis," McGraw-Hill, New York, 1953.
4. A. V. FIACCO AND G. P. MCCORMICK, "Nonlinear Programming: Sequential Unconstrained Minimization Techniques," John Wiley and Sons, New York, 1968.
5. J. D. PEARSON, Variable metric methods of minimisation, *Computer J.* **12** (1969), 171–178.
6. V. N. FADDEEVA, "Computational Methods of Linear Algebra," Dover Publications, New York, 1959.
7. G. A. BAKER, JR. AND J. L. GAMMEL (Eds.), "The Padé Approximant in Theoretical Physics," Academic Press, New York, 1970.
8. E. F. BECKENBACH AND R. BELLMAN, "Inequalities," Springer–Verlag, New York, 1965.
9. M. MARCUS, "Basic Theorems in Matrix Theory," Nat. Bur. Stand. Appl. Math. Ser., **57**, 1960.
10. H. H. ROSENBROCK, An automatic method for finding the greatest or least value of a function, *Computer J.* **3** (1960), 175–184.
11. F. FREUDENSTEIN AND B. ROTH, Numerical solution of systems of nonlinear equations, *J. Assoc. Comput. Mach.* **10** (1963), 550–556.
12. J. L. GREENSTADT, On the relative efficiencies of gradient methods, *Math. Comput.* **21** (1967), 360–367.