

# Widely Convergent Method for Finding Multiple Solutions of Simultaneous Nonlinear Equations\*

**Abstract:** A new method has been developed for solving a system of nonlinear equations  $g(x) = 0$ . This method is based on solving the related system of differential equations  $dg/dt \pm g(x) = 0$  wherein the sign is changed whenever the corresponding trajectory  $x(t)$  encounters a change in sign of the Jacobian determinant or arrives at a solution point of  $g(x) = 0$ . This procedure endows the method with a much wider region of convergence than other methods (occasionally, even global convergence) and enables it to find multiple solutions of  $g(x) = 0$  one after the other. The principal limitations of the method relate to the extraneous singularities of the differential equation. The role of these singularities is illustrated by several examples. In addition, the extension of the method to the problem of finding multiple extrema of a function of  $N$  variables is explained and some examples are given.

## Introduction

A new method has been developed for solving systems of nonlinear equations [1]. The method, based on integrating a related system of differential equations, has a wide region of convergence and may occasionally be globally convergent. Moreover, it is capable of finding multiple solutions (sometimes all of them) one after the other without requiring deflation of the original equations. The method does have significant limitations, however, and these will be discussed below.

This method evolved as a result of unsuccessful efforts to solve a certain nonlinear electrical problem, the "tunnel diode" problem of Fig. 1, which frustrated the usual methods of solution. The basic difficulty in this problem—a difficulty which is common to many other nonlinear systems—is the fact that the Jacobian matrix may become singular. When this happens, Newton's method and its variants cease to function. However, a simple and effective way of circumventing this difficulty has been found. As a result, the method has a much larger region of convergence than those methods to which a Jacobian singularity presents an impasse.

By a straightforward extension of this method, it is possible to find multiple extrema of a scalar function of a vector variable and under favorable conditions to find the global maximum and/or global minimum. This extension, which is based on finding multiple zeroes of the gradient of the function, is described in a companion paper [2] and in condensed form in Appendix C. The application of the method to the case of complex zeroes will be treated in a subsequent paper.

During the course of this development the author has made extensive use of visual displays generated by an IBM 1130 computer driving an IBM 2250-4 display unit. All of the drawings in this paper were computer-generated, first being presented on the display unit and then being copied by an IBM 1627 plotter (only the labeling has been added for purposes of publication). The value of these visual displays in guiding the development and in providing insight to the workings of the new method can scarcely be overemphasized.

## Genesis of the method

Let the system of nonlinear equations be written in the form

$$g(x) = 0, \quad (1)$$

where  $g(x)$  is an  $n$ -dimensional vector-valued function of an  $n$ -dimensional vector  $x$ . One variant of Newton's method which the author has studied is based on the recursion formula

$$x_{i+1} = x_i - h_i \left( \frac{\partial g}{\partial x} \right)_{x_i}^{-1} g(x_i), \quad (2)$$

where  $x_i$  is the  $i$ th approximation to the solution vector and  $(\partial g / \partial x)$  is the Jacobian matrix. As Broyden shows [3], divergence of this iteration process can be avoided by choosing the step size  $h_i$  so as to minimize  $\|g(x_{i+1})\|$ ,

\*The method reported in this paper was presented orally at the Conference on Numerical Methods for Nonlinear Optimization, University of Dundee, Scotland, June, 1971.

the magnitude of the new *residual* or *error* vector—or at least so as to make

$$\|g(x_{i+1})\| \leq \|g(x_i)\|. \quad (3)$$

But avoiding divergence does not guarantee convergence. For in the case of the tunnel diode problem of Fig. 1, the foregoing algorithm fails to converge from any starting point to the left of the vertical line through point B. What happens is the following.

For this problem, the equations are

$$\begin{bmatrix} g_1(x_1, x_2) \\ g_2(x_1, x_2) \end{bmatrix} = \begin{bmatrix} f(x_1) - x_2 \\ x_1 + Rx_2 - E \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (4)$$

where  $f(x_1)$  is the function describing the nonlinear tunnel diode curve,  $R$  is a resistance value, and  $E$  a battery voltage. For any given values of  $x_1$  and  $x_2$ , as shown in Fig. 1(b),  $g_1(x_1, x_2)$  is the vertical distance from point  $P_1$  to the point  $(x_1, x_2)$  while  $g_2(x_1, x_2)$  is the horizontal distance from  $(x_1, x_2)$  to the point  $P_2$ . Thus the line  $P_1P_2$  represents the  $g$  vector in this problem.

For any starting point to the right of the vertical line through B, the foregoing minimization algorithm will force both  $P_1$  and  $P_2$  towards the solution point C, causing the  $g$  vector to shrink to zero. But for any starting point to the left of this line, the algorithm drives  $P_1$  toward point A and  $P_2$  towards point A', where the length of the  $g$  vector is minimum—but not zero. The algorithm is then unable to proceed further because at every point on the vertical line through A, the Jacobian matrix,

$$J = \frac{\partial g}{\partial x} = \begin{bmatrix} df/dx_1 & -1 \\ 1 & R \end{bmatrix} \quad (5)$$

becomes singular since  $df/dx_1 = -1/R$ . (This also happens everywhere on the vertical line through B.) Thus,  $\det J = 0$  and so the inverse Jacobian matrix  $J^{-1}$  required by Eq. (2) no longer exists.

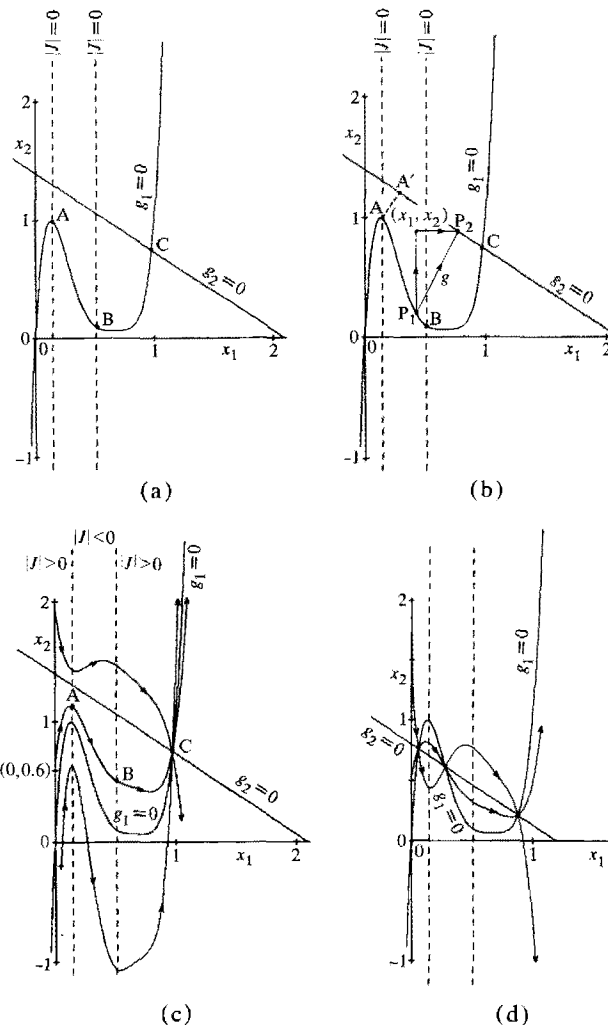
Now it is well known in circuit analysis that difficult non-linear problems, such as the one just described, can often be solved by integrating a related system of differential equations until a steady-state solution is obtained. This observation suggested the use of a synthetic differential equation such as

$$\frac{dx}{dt} + g(x) = 0, \quad (6)$$

which provides the desired solution to Eq. (1) when  $dx/dt = 0$ . This approach works well for the tunnel diode problem but fails on other problems and has been abandoned in favor of the equation

$$\frac{dg}{dt} + g(x) = 0 \quad (7)$$

suggested by L. E. Kugel [4].



**Figure 1** The tunnel diode problem. (a) Curves defined in Eq. (4); (b) graphical interpretation; (c) trajectories through a single solution; (d) trajectories through three solutions.

Equation (7), which was found later to have been studied extensively in connection with the methods of "continuation" and "parameter imbedding," [5-11] proves to be effective as long as the Jacobian matrix is nonsingular. For it is obvious from the analytic solution of Eq. (7), namely

$$g[x(t)] = g[x(0)]e^{-t} \quad (8)$$

that Eq. (1) will be satisfied when  $t \rightarrow \infty$ .

Since  $g(x)$  is not an explicit function of  $t$ , it follows from the chain rule that

$$\frac{dg}{dt} = \left( \frac{\partial g}{\partial x} \right) \frac{dx}{dt}, \quad (9)$$

so that Eq. (7) can be recast in the form

$$\frac{dx}{dt} = - \left( \frac{\partial g}{\partial x} \right)^{-1} g(x) = -J^{-1}g(x) \quad (10)$$

as long as the Jacobian matrix remains nonsingular. However, when  $\det J = 0$ , Eqs. (7) and (10) cease to exist and instead we have

$$\frac{dg}{dt} = \left( \frac{\partial g}{\partial x} \right) \frac{dx}{dt} = 0 \quad (11)$$

with  $dx/dt \neq 0$ . Accordingly, this approach fails in the tunnel diode problem when  $x_1 = a$  or  $x_1 = b$ .

The way to remove the impasse presented by the Jacobian singularity is simply to allow a change of sign in Eq. (7) so that we have

$$\frac{dg}{dt} \pm g(x) = 0 \quad (12)$$

with solution

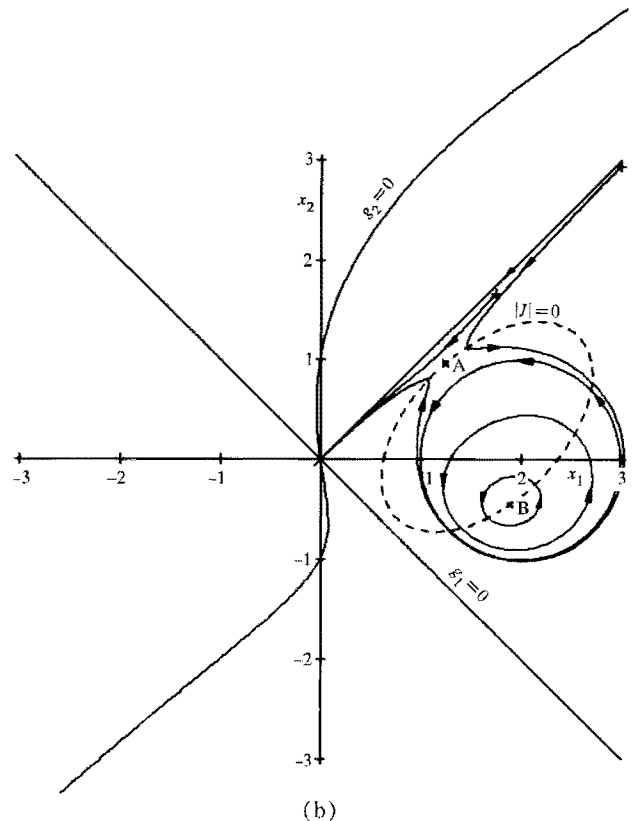
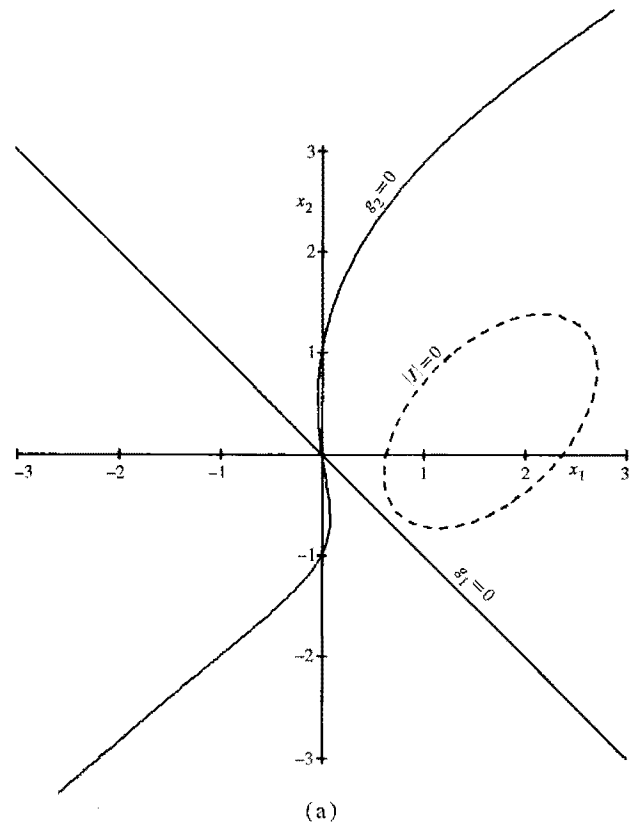
$$g[x(t)] = g[x(0)]e^{\mp t}. \quad (13)$$

Thus, when  $\det J$  changes sign (after passing through zero), the sign in Eq. (12) is reversed so that the integration process may be continued. Equation (10), therefore, becomes

$$\frac{dx}{dt} = \mp J^{-1}g(x). \quad (14)$$

In the case of the tunnel diode problem, assume that the starting point is  $(0, 0.6)$ , as shown in Fig. 1(c), and that initially the positive sign is taken in Eq. (12) since  $\det J > 0$ . The integration path (or trajectory) proceeds as shown up to the Jacobian singularity at point A while the  $g$  vector decreases exponentially in magnitude (*without* changing direction) according to Eq. (13). If then the negative sign is taken in Eq. (12), the trajectory proceeds to the next Jacobian singularity at point B. During this interval,  $\det J < 0$  and the  $g$  vector is increasing exponentially—as it obviously must if the trajectory is to get “over the hump” en route to the solution at point C. Finally, the positive sign is restored and the integration path proceeds directly to the solution at point C. If the sign is reversed once more, the integration path beyond point C diverges to infinity. (Two other trajectories are also shown in the drawing.)

Now if the value of the constant  $E$  in Eq. (4) is reduced sufficiently, the tunnel diode problem exhibits three solutions as shown in Fig. 1(d). All three of these solutions can be found in sequence by appropriate use of the foregoing rules for changing the sign in Eq. (12)—namely, reversing sign whenever  $\det J$  changes sign and whenever a solution point has been reached. Moreover, the algorithm is globally convergent for this problem and for certain other problems that have been studied. However, the method is not globally convergent in general, as explained in the following section.



**Figure 2** Brent's problem. (a) Curves defined by Eq. (15); (b) trajectories in the presence of a saddle point and a vortex point.

### Theoretical basis of the method

R. Brent [12] proposed the following problem as an example in which the present method is not globally convergent.

$$g_1(x_1, x_2) = a(x_1 + x_2) = 0 \quad (15a)$$

$$g_2(x_1, x_2) = a(x_1 + x_2) + (x_1 - x_2)[(x_1 - b)^2 + x_2^2 - c] = 0. \quad (15b)$$

These functions, for the values  $a = 4$ ,  $b = 2$ ,  $c = 1$ , are shown in Fig. 2(a), the locus defined by  $\det J = 0$  being an ellipse.

Brent correctly predicted that any trajectory starting from a point within the disc defined by

$$(x_1 - 2)^2 + x_2^2 = 1 \quad (16)$$

would fail to converge on the solution point at (0,0). Indeed, the region of nonconvergence is slightly larger than this disc and is characterized by the fact that all trajectories within it are closed curves. Outside this region convergence is always obtained.

Now it is obvious from the trajectories shown in Fig. 2(b) that point A, which lies on the ellipse  $\det J = 0$ , has all the earmarks of a *saddle point* while point B, which also lies on this ellipse, resembles a *vortex point*. Initially, these points were thought to be somehow different from the usual singular points of a second order differential equation because Eq. (14) implies the existence of a *conjugate pair* of (vector) differential equations.

These special points, and the trajectories in their neighborhoods, were surmised to be phenomena peculiar to the interaction between this conjugate pair of differential equations across the boundary,  $\det J = 0$ , separating their respective domains. However, further study has shown that these points are indeed true singular points. Moreover, every solution of  $g(x) = 0$  is also a singular point—specifically a *degenerate node* that may be either stable or unstable, depending on the choice of sign in Eq. (14).

To explain these features of the method, let us recast Eq. (14) in the form of a single vector differential equation instead of a conjugate pair. Using Cramer's rule for the inverse of a matrix in terms of its adjoint we have

$$\frac{dx}{dt} = \frac{\text{adj } J}{|\det J|} g(x). \quad (17)$$

Here the rule for changing sign when  $\det J$  changes sign is automatically incorporated by using the absolute magnitude of  $\det J$ . Thus, whenever the right hand side of Eq. (17) is a null vector, a singular point occurs. In particular, all solutions of  $g(x) = 0$  correspond to singular points.

To determine the type of singular point in any instance, it is sufficient to determine the eigenvalues of the coefficient

matrix of the variational equation [13] related to Eq. (17). In the present case, this matrix may be called a *secondary Jacobian* and is defined in terms of Eq. (14) by the expression

$$\begin{aligned} \frac{\partial \dot{x}}{\partial x} &= \mp \left[ \left( \frac{\partial J^{-1}}{\partial x} \right) g(x) + J^{-1} \left( \frac{\partial g}{\partial x} \right) \right] \\ &= \mp \left[ \left( \frac{\partial J^{-1}}{\partial x} \right) g(x) + I \right], \end{aligned} \quad (18)$$

where  $\dot{x} = dx/dt$ .

Whenever  $g(x) = 0$ , this secondary Jacobian matrix reduces to  $\mp I$  with eigenvalues  $\mp 1$ . Hence, each solution point of  $g(x) = 0$  corresponds to a *degenerate node* [14] which is stable if the negative sign is used in Eq. (14) and unstable if the positive sign is used. Since the solutions of  $g(x) = 0$  are of primary interest, we shall call them *essential* singular points of Eq. (14).

*Extraneous* singular points may also occur when  $g(x) \neq 0$  provided that the  $g$  vector satisfies the homogeneous equation

$$(\text{adj } J)g = 0, \quad (19)$$

for this condition will make the right-hand side of Eq. (17) vanish. Equation (19) also implies that  $\text{adj } J$  is singular. Moreover, since Cramer's rule requires that

$$J \cdot \text{adj } J = |J| \cdot I, \quad (20)$$

the singularity of  $\text{adj } J$  implies that  $|J| = 0$ . Hence,  $J$  itself is singular.

Now it is well established that  $\text{adj } J$  is a null matrix if the rank of  $J$  is  $n - 2$  or less; but if  $\text{rank}(J) = n - 1$ , then  $\text{rank}(\text{adj } J) = 1$ . In this case,  $\text{adj } J$  can be expressed in dyadic form as [15]

$$\text{adj } J = pq^t \quad (21)$$

where  $p$  is a column vector and  $q^t$  a row vector. (An explicit method for computing  $p$  and  $q$  is given in Appendix A.)

Equation (19) can now be written in the form

$$(\text{adj } J)g = (pq^t)g = p(q^t g) = 0 \quad (22)$$

or, equivalently,

$$q^t g = 0. \quad (23)$$

Thus, if  $\text{rank}(J) = n - 1$ , every  $g$  vector in the  $(n - 1)$ -dimensional subspace orthogonal to the  $q$  vector of Eq. (21) will satisfy Eq. (19); if  $\text{rank}(J) = n - 2$ , then every  $g$  vector in the entire  $n$ -space satisfies Eq. (19). In summary then, extraneous singular points of Eq. (14) occur whenever any point on the locus defined by  $\det J = 0$  corresponds to a *nonzero*  $g$  that satisfies Eq. (19).

The theoretical basis of the present method for solving  $g(x) = 0$  is thus intimately tied to the nature of the

singular points of Eq. (14) and their influence on the trajectories defined by this equation. Although no definitive explanation can yet be offered, it is evident from experimental results that the existence of one or more extraneous singular points may, and often will, preclude the method from being globally convergent and may also prevent some trajectories from traversing all solutions.

So far, the only types of extraneous singular points observed have been saddle points and vortex points. Indeed, it seems unlikely that spiral points (foci) will occur since they represent points of stable equilibrium, which are to be expected only when  $g(x) = 0$ . However, neither experimental nor theoretical evidence can be given to justify this conjecture.

Unfortunately, saddle points and vortex points are not only extraneous but troublesome. On the one hand, every vortex point is surrounded by a region of nonconvergence. On the other hand, every saddle point will repel any trajectory that comes within its range of influence. This may occasionally give rise to a region of nonconvergence. A more frequent effect of these saddle points is to deflect certain trajectories so that they fail to traverse zeroes of  $g(x)$  that neighboring trajectories will find. Sometimes this deflecting action of saddle points has no harmful effects at all. But this is more likely to be the exception than the rule.

In the absence of extraneous singular points, it seems plausible to expect not only that the method will be globally convergent but also that every trajectory will pass through all solution points of  $g(x) = 0$ . This conjecture is based on the behavior of the algorithm in two problems where extraneous singular points are known to be absent. It is also supported by the observation that the nodal points act like electric charges and the trajectories like electrostatic field lines that either terminate on charges or else pass to infinity. A sound theoretical treatment of the role of both essential and extraneous singular points is needed, however, to establish definitive conditions for global convergence and the ability to find all solutions of  $g(x) = 0$ .

### Computational techniques

Although numerical integration of Eq. (14) may be used to obtain any trajectory  $x(t)$  in  $x$ -space, the existence of an analytic solution for the *image* trajectory in  $g$ -space, given by Eq. (13), provides a unique and distinct advantage. For as Eq. (13) shows, every trajectory in  $x$ -space must be such that its image under the transformation  $g(x)$  lies on a straight line determined by  $g[x(0)]$ . Recognition of this important property of the method can lead to significant improvements in the numerical solution of Eq. (14) as we shall show.

In particular, Euler's method, which is the simplest and least accurate of the commonly used integration

schemes, can be supplemented by the technique of "steering" to provide very accurate solutions of Eq. (14). Euler's method, applied to Eq. (14) yields the recursion formula

$$x_{i+1} = x_i \mp h_i J_i^{-1} g(x_i), \quad (24)$$

which is identical to Eq. (2) for the variable-step Newton method except for the alteration of sign.

According to Eq. (14), the "Newton vector"  $\mp J^{-1}g$  is tangent to the trajectory at each point and may be regarded as the *velocity*, if  $t$  is taken to represent time. Equation (24) then shows that the trajectory may be approximated by taking a finite step along this velocity vector. Obviously, the new point  $x_{i+1}$  will be displaced more or less from the actual trajectory, depending on its curvature. Consequently, even if the point  $x_i$  is on the actual trajectory so that  $g(x_i)$  satisfies Eq. (13), the image of the new point, namely  $g(x_{i+1})$ , cannot, in general, satisfy Eq. (13). The idea of "steering," then, is to modify or correct the direction of the velocity vector somewhat so as to make  $g(x_{i+1})$  satisfy Eq. (13) more closely than it otherwise could.

The actual algorithm, which amounts to a predictor-corrector scheme, is as follows: First, rather than using the Newton vector  $\mp J^{-1}g$ , which may vary considerably in magnitude, we actually normalize it. The effect of this normalization is to replace the independent variable  $t$  by the variable  $s$  representing distance along the trajectory. Specifically, we use

$$\frac{dx}{dt} = \frac{dx}{ds} \cdot \frac{ds}{dt} \quad (25)$$

and

$$\frac{ds}{dt} = \left[ \left( \frac{dx}{dt} \right)^t \left( \frac{dx}{dt} \right) \right]^{\frac{1}{2}} \quad (26)$$

so that Eq. (17) can be written as

$$\frac{dx}{ds} = \frac{\text{adj } J}{|\det J|} \cdot \frac{|\det J|}{[g^t(\text{adj } J)^t(\text{adj } J)g]^{\frac{1}{2}}} \cdot g(x) \quad (27)$$

or, using the Euclidean norm,

$$\frac{dx}{ds} = \frac{(\text{adj } J)g}{\|(\text{adj } J)g\|}. \quad (28)$$

Next, Eq. (24) is replaced by

$$x_{i+1} = x_i + h_i \frac{(\text{adj } J)g}{\|(\text{adj } J)g\|}, \quad (29)$$

where  $h_i$  now corresponds to a fixed-length step along the trajectory. Equation (29) is used to predict the new point  $x_{i+1}$  and then  $g(x_{i+1})$ , or more concisely  $g_{i+1}$ , is evaluated. The component of  $g_{i+1}$  *orthogonal* to the vector  $g[x(0)]$ , or  $g_0$ , is then computed according to the relation

$$v = g_{i+1} - k g_0, \quad (30)$$

where  $k g_0$  is the projection of  $g_{i+1}$  along  $g_0$  and  $k = g_{i+1}^t g_0 / g_0^t g_0$ .

Annihilating this  $v$  vector by adding the appropriate correction vector to  $x_{i+1}$  will insure that the image of the corrected point satisfies Eq. (13). To first order, this correction vector is computed by the relation

$$\delta x = -J^{-1}v \quad (31)$$

where the same inverse Jacobian matrix (or better still, its  $LU$  factors [16]) already computed for Eq. (14) can be used unchanged. The corrected new point is

$$\bar{x}_{i+1} = x_{i+1} + \delta x \quad (32)$$

and if  $g(\bar{x}_{i+1})$  does not have a sufficiently small orthogonal component, the corrector equations can be applied repeatedly.

The trajectories shown in most of the examples in the paper were obtained by combining Euler's predictor and the "steering" corrector formulas just described. In this way, very accurate trajectories were obtained at minimal computational cost. Of course, the basic objective of the method is to find the solutions of  $g(x) = 0$  so that the actual *shape* of these trajectories is of no real consequence as long as all the zeroes of  $g(x)$  can be reliably found. This will be a governing objective in developing an efficient computational algorithm. But for the present purposes of highlighting the underlying mathematical character of the method, accurate trajectories have been thought desirable.

Recently Broyden [17] has proposed a new method of solving non-linear equations which has a noteworthy resemblance to the present method. Broyden's new method is based on a differential equation of the form

$$\frac{dx}{dt} = -J^{-1}g_0 \quad (33)$$

which, like Eq. (10), is limited to problems in which the Jacobian matrix is nonsingular. However, Broyden uses the *initial*  $g$  vector,  $g_0$ , or  $g[x(0)]$ , in place of the *current*  $g$  vector,  $g[x(t)]$ , on the righthand side. He then solves Eq. (33) by a method akin to Euler's method of integration, but with the step size computed according to the curvature of the trajectory. The net result is to replace our Eq. (2) by the relation

$$x_{i+1} = x_i + b_i J_i^{-1} g_0, \quad (34)$$

where  $b_i$  is used to represent Broyden's step size.

Now Broyden's equation incorporates somewhat the consequences, though not the basic idea, of "steering" in the following sense. When the steering correction is applied as described above, the current  $g$  vector,  $g(x_i)$  in Eq. (24), is made to be codirectional with  $g_0$ . There-

fore, we may set  $g(x_i) = k_i g_0$  so that Eq. (24) can be written in the form

$$x_{i+1} = x_i + h_i k_i J_i^{-1} g_0. \quad (35)$$

By amending Broyden's equation to include our sign change, we see that the expression

$$b_i = h_i k_i \quad (36)$$

for the step sizes correlates Broyden's equation with ours.

It must be recognized, however, that even with equivalent step sizes chosen according to Eq. (36), these two methods will not give identical results. For although Broyden's predictor will give the same new point as Eq. (35) on the *first* step (when  $k_0 = 1$ ), all subsequent points will differ. (This is because Broyden's predictor alone cannot satisfy Eq. (13), whereas our steering technique can.) Moreover, the cumulative effects of this discrepancy will cause the successive Jacobian matrices in Broyden's method and ours to become increasingly dissimilar. Thus the two trajectories will be noticeably different, especially with large step sizes.

This observation has been confirmed experimentally and our predictor-corrector scheme has been found to produce trajectories that "home in" on the zeroes of  $g(x)$  considerably better than trajectories produced by Broyden's method. In fact, convergence to the single solution of the tunnel diode problem of Fig. 1(c) has been obtained from the points  $(0.75, \pm 100)$  and  $(0.8, \pm 100)$  with only two or three "Newton steps" each followed by one or more steering corrections. Thus, the steering principle alone is a powerful adjunct to existing methods of solving nonlinear equations.

### Experimental results

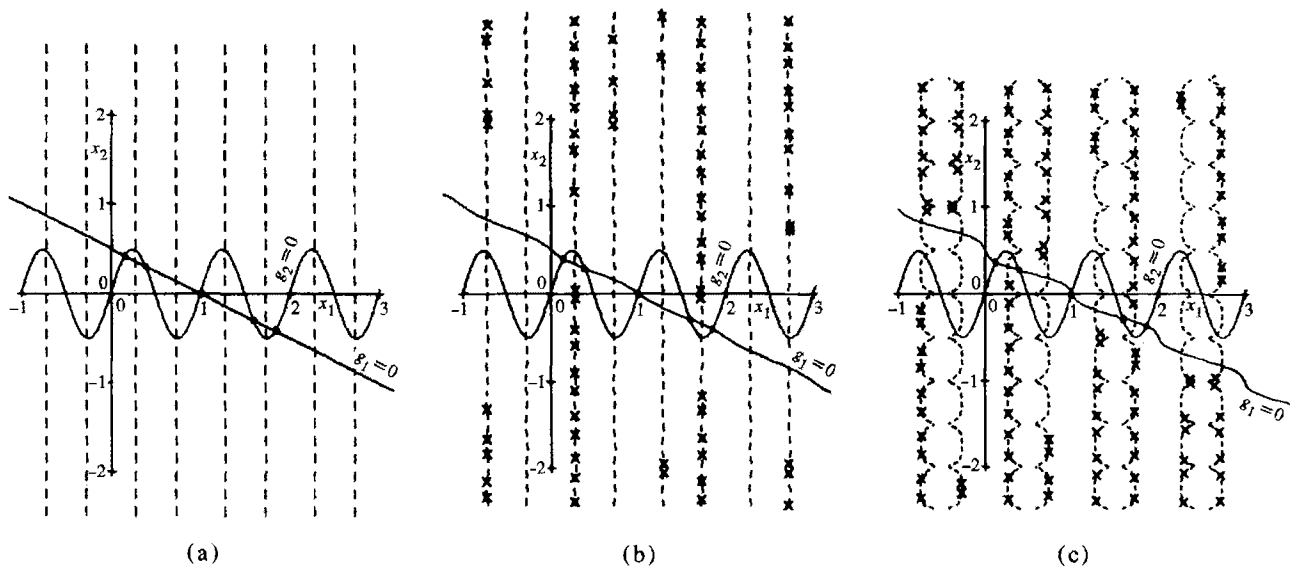
Several two-dimensional problems have been studied in some detail. In order to discuss their extraneous singular points, we first display explicitly the singular adjoint matrix for the  $2 \times 2$  case. According to Eq. (A10) of Appendix A, we have

$$\text{adj } J = \frac{1}{J_{11}} \begin{bmatrix} -J_{12} & -J_{21} \\ J_{11} & J_{11} \end{bmatrix} \quad (37)$$

so that the  $q^t$  vector of Eq. (21) is  $[-J_{21} \ J_{11}]$ . Hence, any point on the locus  $\det J = 0$  which has as its image a *nonzero*  $g$  vector orthogonal to  $q$  will correspond to an extraneous singular point.

Since in the case of a singular  $2 \times 2$  matrix the rows (and columns) are proportional, it follows that  $q^t$  may also be represented by  $[J_{22} \ -J_{12}]$ . Therefore, the necessary orthogonality condition is either

$$-J_{21}g_1 + J_{11}g_2 = 0 \quad (38)$$



**Figure 3** Trigonometric problem. (a) Curves defined by Eq. (41) for  $c = 0$ ; no extraneous singularities present; (b)  $c = 0.05$ ; extraneous singularities far from solution points; (c)  $c = 0.13$ ; extraneous singularities closer to solution points.

or

$$J_{22}g_1 - J_{12}g_2 = 0. \quad (39)$$

In the tunnel diode problem, discussed earlier in terms of Eqs. (4) and (5), it will be recalled that the loci defined by  $\det J = 0$  are the vertical lines through points A and B of Fig. 1(a). Hence,  $\det J = 0$  implies that  $x_1 = a$  or  $x_1 = b$ . But the orthogonality condition of Eqs. (38) and/or (39) reduces to

$$f(x_1) = (E - x_1)/R, \quad (40)$$

which must hold simultaneously with  $x_1 = a$  or  $x_1 = b$ .

Now if the variable  $x_2$  is eliminated from Eq. (4), the result is Eq. (40)—which must therefore correspond to a solution of Eq. (4). That is, the null vector,  $g = 0$ , is the only vector that satisfies the orthogonality condition of this problem. Obviously, then, no extraneous singular points can occur in the tunnel diode problem.

Clearly, the situation just described corresponds to the tangency of the curves  $g_1(x_1, x_2) = 0$  and  $g_2(x_1, x_2) = 0$  at point A or at point B in Fig. 1(a). It is natural therefore, to consider the effect of such tangencies on the present algorithm. Since exact tangency cannot be achieved on a digital computer, the curves must either approach very closely to each other without touching, or else they must intersect in two distinct, but very close, points. In the first instance, the main problem is that of deciding on an appropriately small convergence criterion that the  $g$  vector must satisfy. But in the second instance, computational difficulties may arise during the process of projecting the trajectory through the first solution point and continuing the search for additional

solutions. Neither situation has been studied in detail, however.

A problem somewhat like the tunnel diode problem is that of solving the equations

$$g_1(x_1, x_2) = a - bx_2 + c \sin(dx_2) - x_1 = 0 \quad (41a)$$

$$g_2(x_1, x_2) = x_2 - e \sin(fx_1) = 0. \quad (41b)$$

In Fig. 3(a), we show these two curves for the case where  $c = 0$ , so that  $g_1 = 0$  is a straight line, and where  $a = 1$ ,  $b = 2$ ,  $d = 4\pi$ ,  $e = 0.5$ ,  $f = 2\pi$ . In this instance, the loci for  $\det J = 0$  are straight lines, satisfying the equation

$$\cos(fx_1) = -1/bef, \quad (42)$$

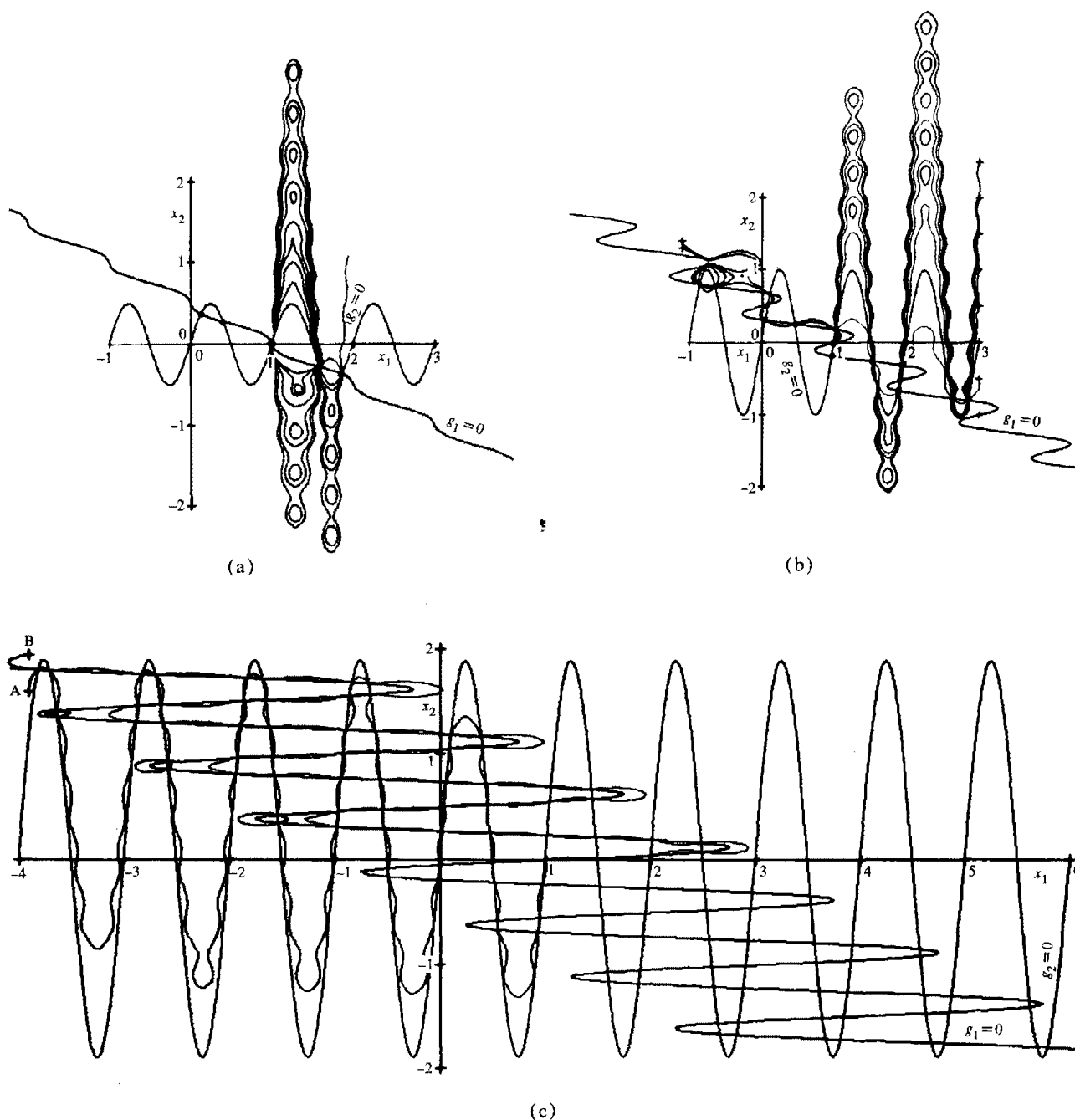
and the orthogonality condition reduces to

$$\sin(fx_1) = (a - x_1)/be. \quad (43)$$

But Eq. (43) is identical with the result obtained from Eqs. (41a) and (41b) by eliminating the variable  $x_2$ . Thus, the null vector,  $g = 0$ , alone satisfies the orthogonality condition so that no extraneous singularities exist, just as in the tunnel diode problem.

However, when  $c \neq 0$ , extraneous singularities come into the picture with a vengeance, as Figs. 3(b) (for  $c = 0.05$ ) and 3(c) (for  $c = 0.13$ ) clearly demonstrate. (The extraneous singular points are designated by cross marks.) Note that as  $c$  increases, these singularities approach closer to the solution points. Note also that because the functions in Eq. (41) are periodic, there is an infinite number of extraneous singular points.

In Figs. 4(a), (b) and (c), several different trajectories are shown, many of which pass through all of the solution



**Figure 4** Trajectories for trigonometric problem. (a)  $c = 0.13$  ; convergent and cyclic trajectories; (b)  $c = 0.5$  and  $e = 1.0$  ; cyclic trajectories without vortex point; (c)  $c = 2$  and  $e = 1.9$  ; showing 119 solution points.

points. (To reduce clutter in these drawings, we have taken advantage of the symmetry of the problem relative to the point  $(1,0)$  by showing only half of the centrosymmetric trajectories.)

Figure 4(a) (for  $c = 0.13$ ) seems to indicate that each saddle point is paired with a vortex point. But this is not always so. In Fig. 4(b) (for  $c = 0.5$  and  $e = 1.0$ ) the saddle point at  $(-0.28, 0.91)$  is not associated with a vortex point

at all. Instead, there is a neighboring family of closed trajectories that pass through the two left-most solution points. Many cyclic trajectories are, however, shown to be associated with vortex points.

In Fig. 4(c), the trajectories starting at  $x_1 = 3$  and  $x_2 = 2.5, 2.0, 1.5, 1.0, 0.5, 0.0, -0.5$ , and  $-1.0$  pass through all of the solution points. (Incidentally, the starting points  $(3,0)$  and  $(3,-1)$  lie on the curves  $g_2 = 0$  and  $g_1 = 0$ .



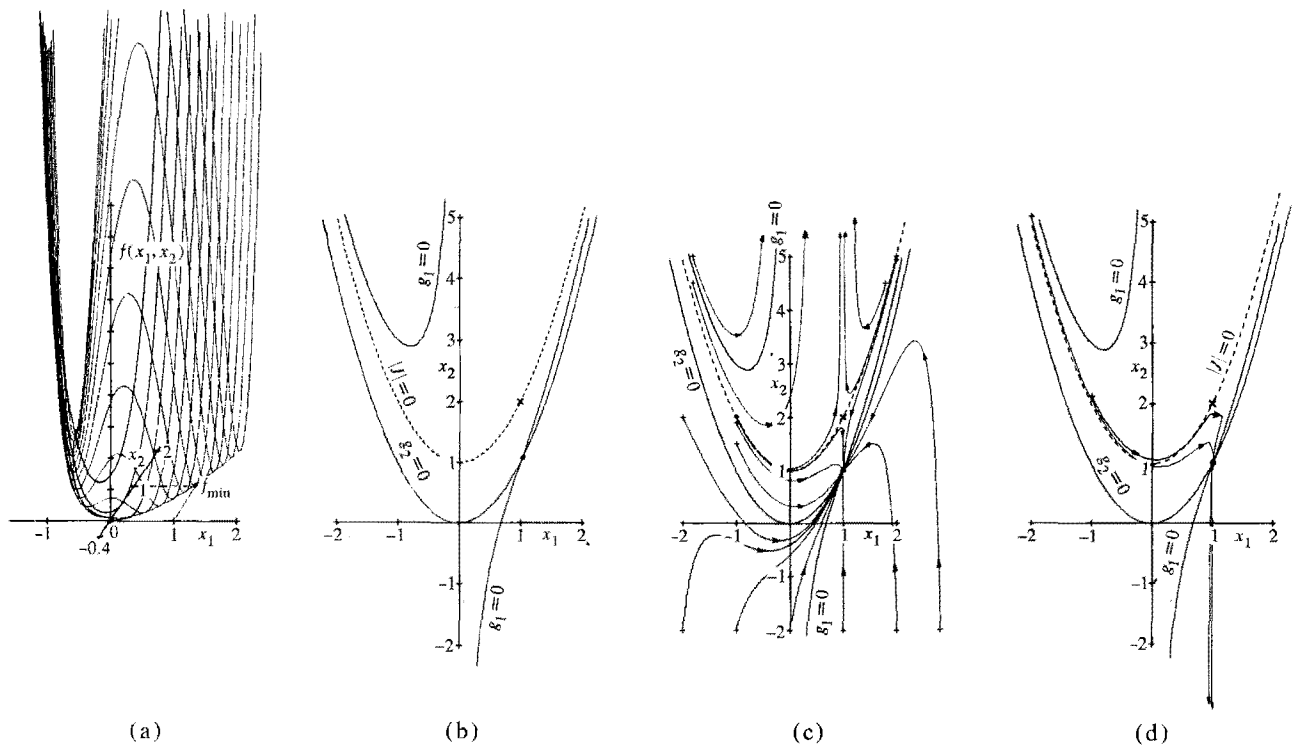


Figure 5 Rosenbrock's function defined by Eq. (44). (a)  $a = 100$  and  $b = 1$ ; (b) gradient equations, Eq. (45);  $a = 0.5$  and  $b = 1$ ; (c) solution showing region of nonconvergence; (d) solution showing escape path from region of nonconvergence.

But these curves are always valid solutions of Eq. (14) anyway, so that the trajectories need not be computed.) The trajectory starting at  $(-1.1, 1.3)$  also passes through all solutions; but its neighbor, starting at  $(-1.1, 1.4)$  does not. It is deflected by the saddle point at  $(-0.28, 0.91)$  so that it entirely misses the two left-most solutions. In Fig. 4(c), where  $c = 2$  and  $e = 1.9$ , there are 119 solutions and the trajectory starting from the point  $(-3.9, 1.6)$  passes through all of them. But the neighboring trajectory, starting at the point  $(-3.9, 1.9)$  fails to find three pairs of solutions. For each of these pairs, incidentally, there is a family of cyclic trajectories passing through them, as shown in the drawing.

A problem that is often used to test optimization procedures is to find the minimum of Rosenbrock's function [18]

$$f(x_1, x_2) = a(x_1^2 - x_2)^2 + (x_1 - b)^2, \quad (44)$$

where the usual values given are  $a = 100$  and  $b = 1$ . This function is displayed in Fig. 5(a). Its minimum, which occurs at  $x_1 = x_2 = 1$ , can be found by computing the simultaneous solution of the two gradient functions

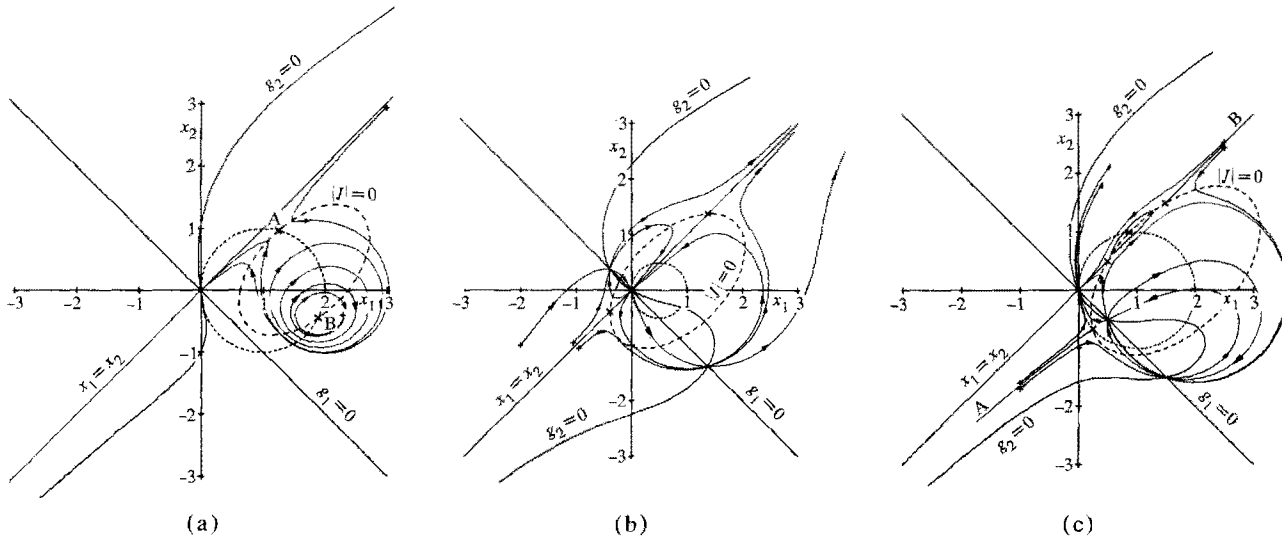
$$\frac{\partial f}{\partial x_1} = g_1(x_1, x_2) = 4ax_1(x_1^2 - x_2) + 2(x_1 - b) = 0 \quad (45a)$$

and

$$\frac{\partial f}{\partial x_2} = g_2(x_1, x_2) = -2a(x_1^2 - x_2) = 0. \quad (45b)$$

In Fig. 5(b), we have shown the loci of  $g_1 = 0$  and  $g_2 = 0$  for the case  $a = 0.5$  and  $b = 1$ , these values being chosen to make the solution point more readily discernible. The locus of  $g_2 = 0$  is the parabola  $x_2 = x_1^2$  while the locus of  $g_1 = 0$  is a curve with two branches. The locus of  $\det J = 0$  turns out to be the parabola  $x_2 = x_1^2 + \frac{1}{2}a$ , shown as a dashed line. The solution point is at  $x_1 = x_2 = 1$ .

Applying the orthogonality condition to Eq. (45), we obtain the result  $x_1 = b$  which, together with the condition  $\det J = 0$  shows that the only extraneous singularity is at the point  $(b, b^2 + \frac{1}{2}a)$ . As Fig. 5(c) shows, this singularity turns out to be a saddle point which has the remarkable property of deflecting away from the solution point every trajectory that lies above the locus  $\det J = 0$ . This locus, then, acts as a separatrix between the region of convergence and the region of convergence. It is interesting to observe in passing that Broyden [17], in using Rosenbrock's problem to test his method, chose a starting point inside the region of nonconvergence just described. We have not tried to see if his method would actually converge from a point within our region of nonconvergence.



**Figure 6** Brent's problem. (a) solution showing escape path from region of nonconvergence; (b)  $a = 4$ ,  $b = 1$ , and  $c = 2.05$ ; solution shows two saddle points, but convergence is global except along  $x_1 = x_2$ ; (c)  $a = 4$ ,  $b = 2$ , and  $c = 2.5$ ; three saddle points and one vortex point.

However, by suppressing our steering correction and using Euler's method of integration with large step size, trajectories starting from points not too far inside the region of nonconvergence can be made to converge. Two examples are shown in Fig. 5(d). "Sloppy" integration also enables trajectories to escape from the region of nonconvergence in Brent's problem as illustrated in Fig. 6(a). In this case, the escape path is a spiral, as would be expected, and the smaller the step size, the tighter the spiral.

In Fig. 6(a), a step size of 0.05 was used in generating the escape path illustrated. In addition, a step size of 0.01 was used to form an almost complete cycle to show how imperfect the closure of this cycle is without the use of the steering correction. With steering, perfect closure is readily attained with a step size as large as 0.1.

At this point, it is difficult to predict whether sloppy integration will provide more than a stand-by strategy for getting out of regions of nonconvergence. Obviously, some strategy must be evolved to handle this problem; but sloppy integration seems to lack the force that is needed to deal with a situation as intricate as this one.

Returning to Brent's problem, we find that the orthogonality condition in this case leads to two expressions, one for the straight line

$$x_1 = x_2 \quad (46a)$$

and the other for the circle

$$(x_1 - b)^2 + x_2^2 = b^2/4, \quad (46b)$$

either of which will satisfy Eq. (38) or Eq. (39). The intersections of these curves with the ellipse  $\det J = 0$

determine the locations of the extraneous singular points shown in Figs. 6(a), (b), and (c).

In Fig. 6(b), for which  $a = 4$ ,  $b = 1$ , and  $c = 2.05$ , there are two saddle points caused by the intersection of  $\det J = 0$  with the line  $x_1 = x_2$ . In this instance, our method turns out to be globally convergent and every trajectory passes through all three solutions.

But in Fig. 6(c), where  $a = 4$ ,  $b = 2$ , and  $c = 2.5$ , a much different situation arises. Here,  $\det J = 0$  and the line  $x_1 = x_2$  determine two saddle points while the circle and  $\det J = 0$  determine one saddle point and one vortex point, the latter being surrounded by the only region of nonconvergence in the problem. Three types of trajectory occur in this case. Trajectories starting from points below the separatrix marked B in the first quadrant or below the separatrix marked A in the third quadrant, pass through all three solutions. But if the starting points are above these separatrices, the corresponding trajectories pass through only the one solution point at the origin. Finally, there is a family of trajectories, cyclic in nature, that pass through the other two solutions.

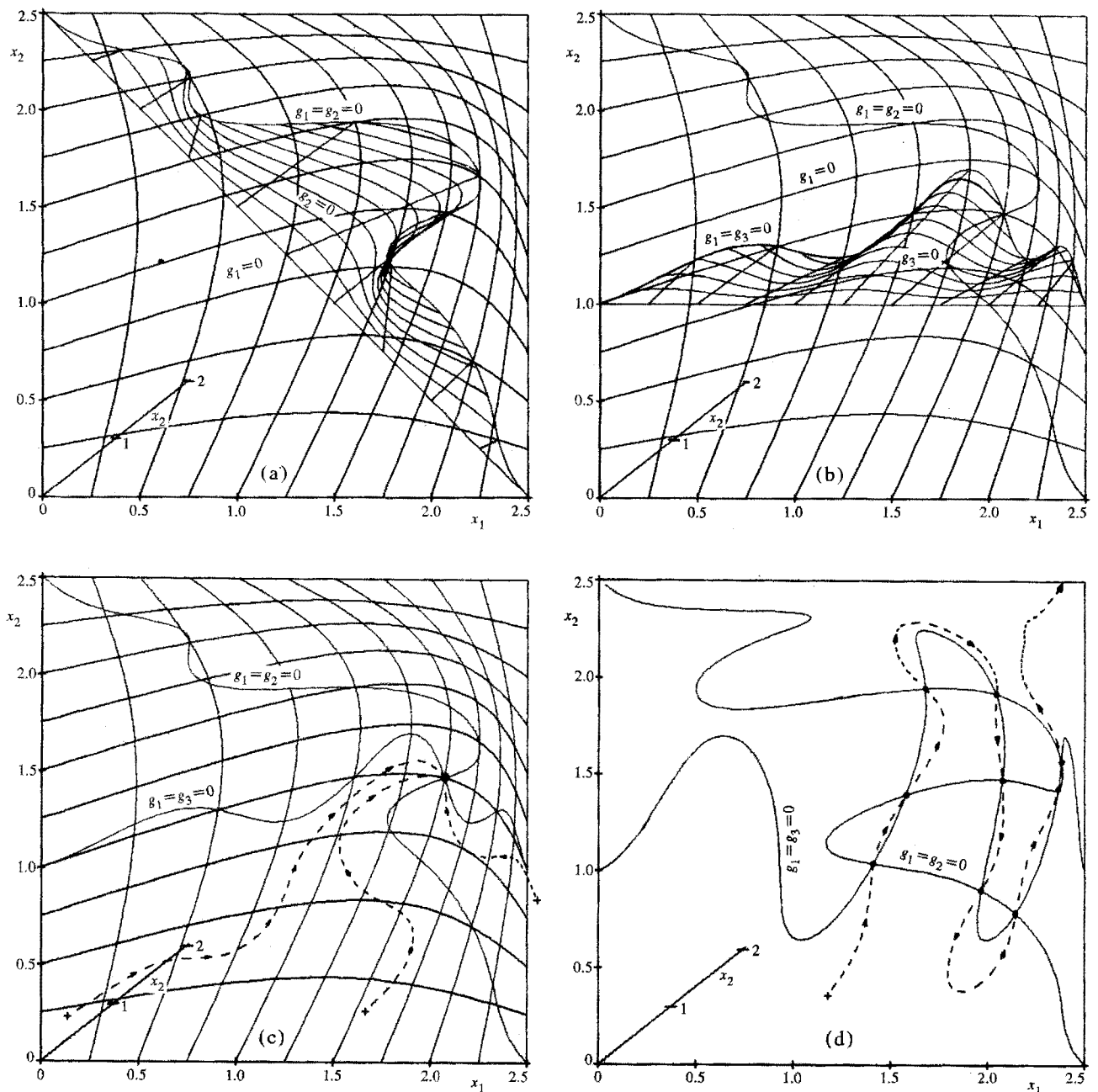
The only three-dimensional problem studied so far is defined by the equations

$$g_1(x_1, x_2, x_3) = a \sin(bx_1) \sin(bx_3) - x_2 = 0 \quad (47a)$$

$$g_2(x_1, x_2, x_3) = c - dx_3 + ex_2 \sin(fx_3) - x_1 = 0 \quad (47b)$$

$$g_3(x_1, x_2, x_3) = g + hx_2 \sin(kx_1) - x_3 = 0. \quad (47c)$$

In Fig. 7(a), the two surfaces  $g_1 = 0$  and  $g_2 = 0$ , corresponding to Eqs. (47a) and (47b), are depicted along with their intersection curve. Fig. 7(b) shows the surfaces defined by Eqs. (47a) and (47c) and their intersection



**Figure 7** Three-dimensional trigonometric problem defined by Eq. (47). (a)  $g_1 = 0$  and  $g_2 = 0$  and their intersection curve; (b)  $g_1 = 0$ ,  $g_3 = 0$ , and their intersection curve; (c) three trajectories to single solution point; (d) with values of constants  $e$  and  $h$  changed, single trajectory through nine solution points.

curve  $g_1 = g_3 = 0$ , as well as the curve  $g_1 = g_2 = 0$ . In Fig. 7(c), which shows only  $g_1 = 0$  and the two foregoing intersection curves, three trajectories converging on the lone solution point are displayed. These trajectories were calculated by using a fourth-order Runge-Kutta integration formula. In all three figures, the values of the coefficients are:  $a = 2$ ,  $b = 0.4\pi$ ,  $c = 2.5$ ,  $d = 1$ ,  $e = 0.1$ ,  $f = 2\pi$ ,  $g = 1$ ,  $h = 0.1$ , and  $k = 2\pi$ .

By increasing the amplitude constants  $e$  and  $h$  to 0.75 and 0.80, respectively, the intersection curves shown in Fig. 7(d) are generated. In this instance, there are nine solution points and there is one trajectory, again computed by fourth-order Runge-Kutta integration, passing through all of these solutions. This problem has not been studied enough to ascertain the nature of its singularities, but there has been occasional evidence of trajectories

that try to take the “long way around” what is probably a region of nonconvergence, as in Fig. 2(b). The most interesting aspect of this problem will be to establish the nature of the two-dimensional singular *regions* predicted by Eq. (23).

### Topics for future development

The most serious drawback of the present method of solving nonlinear equations is the bothersome side-effects of extraneous singular points. And yet the very existence of these unwanted singularities is an almost inevitable consequence of our attempt to deal with Jacobian singularities—at least using Eq. (14). For a singular Jacobian implies a singular adjoint—and this is a necessary (but not sufficient) condition for the existence of extraneous singular points. Moreover, if we insist on looking for multiple solutions of  $g(x) = 0$ , then the inverse function theorem forewarns us that we are bound to encounter at least one Jacobian singularity on each trajectory that joins two adjacent solution points.

Now we may be able to modify Eq. (14) if not to eliminate, then at least to ameliorate, the effects of extraneous singular points.

One possible way of doing this is to inject a smattering of the gradient method by writing

$$\frac{dx}{dt} = \mp (J^{-1} + kJ^t)g(x). \quad (48)$$

For if we define the scalar function  $f(x) = g^t g$  as the square magnitude of  $g(x)$ , then the gradient of  $f$  is

$$\frac{df}{dx} = \left(\frac{\partial g}{\partial x}\right)^t g + g^t \left(\frac{\partial g}{\partial x}\right) = 2J^t g. \quad (49)$$

The coefficient matrix of Eq. (48) is still singular when  $\det J = 0$  so that the extraneous singular points are not eliminated—although their character may change. But the solution points of  $g(x) = 0$  are still guaranteed to be degenerate nodes as long as  $k \geq 0$ . To show this, we first note that the secondary Jacobian matrix corresponding to Eq. (48) is

$$\frac{\partial \dot{x}}{\partial x} = \mp \left[ \left( \frac{\partial J^{-1}}{\partial x} + k \frac{\partial J^t}{\partial x} \right) g + I + kJ^t J \right]. \quad (50)$$

When  $g(x) = 0$ , this reduces to  $I + kJ^t J$  which is positive definite since if  $k \geq 0$ , then  $x \neq 0$  implies

$$x^t (I + kJ^t J) x = \|x\|^2 + k\|(Jx)\|^2 > 0. \quad (51)$$

Hence all the eigenvalues of this matrix are real and positive, assuring that the singularity is a degenerate node.

Now the real payoff in using Eq. (48) is the following: if we introduce Eq. (17) and modify the scalar  $k$  so that  $|\det J|$  can be factored out, Eq. (48) may be written as

$$\frac{dx}{dt} = \frac{1}{|\det J|} (\text{adj } J + kJ^t)g(x). \quad (52)$$

When  $\det J = 0$ , both  $\text{adj } J$  and  $J^t$  become singular. But the rank of the coefficient matrix is no longer unity, as it was in the case of Eqs. (17) and (22); instead it is equal to the rank of  $J^t$  (presumably  $n - 1$ ) no matter how small  $k$  may be. Consequently, instead of having an  $(n - 1)$ -dimensional subspace of  $g$  vectors satisfying the orthogonality condition of Eq. (23), we now have the condition

$$(\text{adj } J + kJ^t)g = 0. \quad (53)$$

Thus, the admissible  $g$  vectors are confined to a one-dimensional subspace so that instead of singular  $(n - 1)$ -regions, as predicted by Eq. (23), we have singular *points* again. The practical value of this approach remains to be established.

Some efforts have been directed towards developing a practical algorithm for implementing the present method. Although the results are inconclusive at present, several useful guidelines can be reported. Because of the side condition expressed in Eq. (13), the usual focus on the pros and cons of implicit vs explicit, multistep vs singlestep methods of integrating Eq. (14) is of little consequence. Indeed, any numerical technique that can be devised to generate a trajectory that will satisfy Eq. (13) is acceptable, whether or not it happens to be classified as an “integration” process. It should also be remembered that the shape of the trajectory generated during the solution of Eq. (14) is of no consequence as long as the solutions of  $g(x) = 0$  can be found.

With these principles in mind, we have found that very large step sizes can be used with Eq. (24) when coupled with the steering techniques described above. One danger, however, is that if too large a step is taken, even though Eq. (13) is satisfied, the trajectory may actually skip over several solution points of  $g(x) = 0$ . To avoid this difficulty, it is helpful to compute the size of the orthogonal component of the  $g$  vector, from Eq. (30), as a function of distance along the tangent to the trajectory. If this orthogonal component increases rapidly, then a smaller step size is requisite. Another danger is that the Jacobian matrix may change sufficiently to invalidate the steering correction. (This may happen in a very nonlinear region.) The remedy in this case is to use Broyden's method [3] for updating  $J^{-1}$  (or its  $LU$  factors) after each steering correction is applied. In this way “local” information about the Jacobian matrix at the new point can be fed into the algorithm.

Another useful technique is to keep a record of several previous points along the trajectory and to fit these points with a polynomial or spline curve. Extrapolation of this curve may be used in place of an integration formula to

predict the next point on the trajectory. The steering correction can then be applied to insure satisfaction of Eq. (13). This technique has been tried with some success but needs further refinement.

Since regions of nonconvergence appear to be an inevitable hazard of the present method, any effective algorithm must include means of recognizing a nonconvergent situation, very likely by detecting cyclic trajectories. It must then find a way out of the region of nonconvergence, perhaps by the temporary expedient of using sloppy integration. But it would be better, if possible, to avoid such regions entirely.

One possibility of doing this stems from the following basic property of Eq. (13): if any element of the initial error vector  $g[x(0)]$  is zero, the corresponding element of  $g[x(t)]$  must remain zero. In other words, if one or more of the  $n$  equations of the system  $g_i(x) = 0, i = 1, 2, \dots, n$  are initially satisfied, they will remain so as long as the side condition of Eq. (13) is satisfied.

Now if there are, say,  $m$  linear equations in the original system, we can immediately satisfy all of them by taking one full Newton step from any arbitrary starting point. If this be done, then any subsequent trajectory defined by Eq. (14) will be confined to the  $(n - m)$ -dimensional hyperplane in  $x$ -space defined by the simultaneous satisfaction of these  $m$  linear equations. Clearly, all solution points of  $g(x) = 0$  must lie on this hyperplane. (See Appendix B.)

If we can then devise a way to satisfy one additional nonlinear equation at a time until  $n - 1$  elements of the  $g$  vector are zero, we will have reached a point on the particular space curve that is defined by the satisfaction of the corresponding  $n - 1$  equations of the system. This curve not only passes through all the solution points of  $g(x) = 0$  but also is a valid trajectory defined by Eq. (14). Therefore, if this curve is simply connected, one complete traversal of it will locate all the desired solutions. If the curve consists of several branches (as in the case of  $g_1 = 0$  in Rosenbrock's problem), the traversal of any one branch will locate only those solution points, if any, that lie on it.

It is not yet clear how to devise an "inflation" method for solving one additional equation at a time and it is by no means certain that such a method will successfully avoid regions of nonconvergence. Neither is there an obvious way of deciding beforehand which  $n - 1$  equations should be solved first so as to arrive at a simply (or simplest) connected space curve. But there may be some merit in pursuing this approach.

### Summary and conclusions

1. A new method of finding multiple solutions of a system of nonlinear equations,  $g(x) = 0$ , has been developed. It is based on solving the related differen-

tial equation  $dg/dt \pm g(x) = 0$  which defines trajectories, or integration paths, in  $x$ -space. The sign of this differential equation is changed whenever a trajectory encounters a change of sign in the Jacobian determinant  $|\partial g/\partial x|$  and whenever it passes through a solution point of  $g(x) = 0$ . This procedure causes the trajectories to pass through one solution after the other.

2. The solution points of  $g(x) = 0$  are shown to be degenerate nodes of the basic differential equation and are classed as *essential* singular points. *Extraneous* singular points may also exist on the loci defined by  $|\partial g/\partial x| = 0$ ; only saddle points and vortex points have been observed. These extraneous singularities may prevent the method from being globally convergent and/or prohibit certain trajectories from passing through any or all of the solution points of  $g(x) = 0$ .
3. The solution of the basic differential equation in  $g$ -space is analytic and serves as a valuable side condition for computing accurate trajectories. This side condition is the foundation of the "steering" technique for altering the direction of the Newton vector so as to speed convergence to a solution point.
4. The "steering" technique, coupled with Euler's integration formula, leads to an effective predictor-corrector scheme for computing trajectories. But the predictor formula need not be based on numerical integration techniques; instead, it may involve a curve-fitting process.
5. The influence of extraneous singular points on the trajectories of several two-dimensional problems has been discussed in conjunction with computer-generated drawings of these trajectories. One three-dimensional problem has also been described.
6. Several suggestions and guidelines for future development of the method have been proposed.
7. It has been pointed out that the underlying cause of the difficulties encountered is the fact that the method attempts to deal directly with the related problems of multiple solutions and Jacobian singularities. If neither of these problems is encountered, then global convergence is assured and can be rapidly achieved with the aid of the steering technique.
8. The method has been extended to the problem of finding multiple extrema of a scalar function of  $n$  variables by finding multiple zeroes of the gradient of the function. Examination of the Hessian matrix of the function (i.e., the Jacobian matrix of the gradient) indicates whether the extremum is a maximum, minimum, or saddle point.
9. A discussion of the possible use of Broyden's method for updating the inverse Hessian matrix has been given. Broyden's method is not recommended for use in integrating the basic differential equation; but

it can be used in conjunction with the "steering" technique, provided that the integration process is supplanted with a curve-fitting method for predicting the next point on a trajectory.

10. Two sample problems, involving functions of two variables, have been treated. One function has an infinite number of minima, but no global minimum; the other has a finite number of maxima, and a global maximum. In both cases, it has been demonstrated that all of the stationary points can be found by this method.

#### Appendix A. Computation of the adjoint of a singular matrix

To compute the adjoint of a singular matrix of rank  $n - 1$ , we may follow a procedure similar to that described by Frazer, Duncan and Collar [15] in which one element of the singular matrix is perturbed by adding an infinitesimal increment. If  $|J| = 0$  and  $\text{rank}(J) = n - 1$ ,  $J$  can always be partitioned in the manner

$$J = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}, \quad (\text{A1})$$

where  $J_{11}$  is a nonsingular submatrix of dimension  $n - 1$ ,  $J_{12}$  is a column vector,  $J_{21}$  a row vector, and  $J_{22}$  a scalar.

Consider the perturbed matrix  $J + \epsilon E_{nn}$  where  $E_{nn}$  is an elementary matrix with 1 in the  $nn$  position and zeroes everywhere else. Assume that the adjoint of the perturbed matrix is  $A + \epsilon B$ . In accord with Eq. (20), we must have

$$(J + \epsilon E_{nn})(A + \epsilon B) = |J + \epsilon E_{nn}| \cdot I. \quad (\text{A2})$$

The determinant of  $J + E_{nn}$  can be computed as follows:

$$|J + \epsilon E_{nn}| = |J| + \begin{vmatrix} J_{11} & 0 \\ J_{21} & \epsilon \end{vmatrix} = \epsilon |J_{11}|, \quad (\text{A3})$$

where  $|J_{11}| \neq 0$  by hypothesis.

Next, assume that the perturbed matrix is subjected to the  $LU$  decomposition [16] in the form

$$\begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} + \epsilon \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ r & 1 \end{bmatrix} \begin{bmatrix} U_{11} & c \\ 0 & s \end{bmatrix}, \quad (\text{A4})$$

where  $J_{11} = L_{11}U_{11}$ , so that both  $L_{11}$  and  $U_{11}$  are nonsingular, and where  $r$  is a row vector,  $c$  is a column vector and  $s$  is a scalar. Expansion of the right hand side of Eq. (A4) shows that

$$c = L_{11}^{-1}J_{12}, \quad (\text{A5})$$

$$r = J_{21}U_{11}^{-1}, \text{ and} \quad (\text{A6})$$

$$s = J_{22} - rc + \epsilon. \quad (\text{A7})$$

But since  $|J| = 0$  by hypothesis,  $J_{22} = rc$ , and so  $s = \epsilon$ .

Substituting Eq. (A4) in Eq. (A2) and then inverting the  $L$  and  $U$  factors, we find that

$$\begin{aligned} A + \epsilon B &= \epsilon |J_{11}| \begin{bmatrix} U_{11}^{-1} & -U_{11}^{-1}c/\epsilon \\ 0 & 1/\epsilon \end{bmatrix} \begin{bmatrix} L_{11}^{-1} & 0 \\ -rL_{11}^{-1} & 1 \end{bmatrix} \\ &= |J_{11}| \begin{bmatrix} \epsilon U_{11}^{-1} & -U_{11}^{-1}c \\ 0 & 1 \end{bmatrix} \begin{bmatrix} L_{11}^{-1} & 0 \\ -rL_{11}^{-1} & 1 \end{bmatrix}. \end{aligned} \quad (\text{A8})$$

Accordingly, in the limit  $\epsilon \rightarrow 0$ , it follows that

$$A = |J_{11}| \begin{bmatrix} -U_{11}^{-1}c \\ 1 \end{bmatrix} \begin{bmatrix} -rL_{11}^{-1} & 1 \end{bmatrix}, \quad (\text{A9})$$

which is the desired dyadic form of the adjoint matrix. Finally, since the determinant of  $J_{11}$  is the product of the diagonal terms of  $U_{11}$ , we have

$$\text{adj } J = \prod_{i=1}^{n-1} U_{ii} \begin{bmatrix} -U_{11}^{-1}c \\ 1 \end{bmatrix} \begin{bmatrix} -rL_{11}^{-1} & 1 \end{bmatrix} \quad (\text{A10})$$

when  $\text{rank}(J) = n - 1$ .

#### Appendix B. The "funneling" theorem

The following theorem is central to the method described in this paper.

**Theorem:** Let  $g_i(x_1, x_2, \dots, x_n) = 0$  for  $i = 1, 2, \dots, n$  be any system of nonlinear equations and let  $x(t)$  be a trajectory defined by a solution of the related system of differential equations  $dg/dt \pm g(x) = 0$ . Finally, let  $S_k$  denote any  $(n - k)$ -dimensional hypersurface in  $x$ -space defined by the simultaneous satisfaction of any  $k$  of the equations  $g_i(x) = 0$ . Then if the starting point  $x(0)$  lies on any given hypersurface  $S_k$ , the entire trajectory  $x(t)$  must also lie on  $S_k$ . Conversely, if  $x(0)$  does not lie on  $S_k$ , the trajectory  $x(t)$  can never intersect  $S_k$  except at a common solution of all  $n$  equations of the system  $g_i(x) = 0$ .

**Proof:** The solution (in  $g$ -space) of the underlying differential equation  $dg/dt \pm g(x) = 0$  is  $g[x(t)] = g[x(0)]e^{\mp t}$ . Therefore, if  $x(0)$  lies on any particular hypersurface  $S_k$ , it follows that the corresponding  $k$  elements of the image vector  $g[x(0)]$  are zero. Since the same  $k$  elements of the image vector  $g[x(t)]$  must also remain zero for all  $t$ , the trajectory  $x(t)$  necessarily remains on  $S_k$ . On the other hand, if any one of these  $k$  elements of  $g[x(0)]$  is nonzero, corresponding to the fact that  $x(0)$  is not on  $S_k$ , the same is true of  $g[x(t)]$  until and unless  $e^{-t} = 0$  as  $t \rightarrow \infty$ . But in this case, all elements of  $g[x(\infty)]$  vanish simultaneously—and this, of course, corresponds to a solution of  $g(x) = 0$ .

A consequence of this theorem is the fact that no trajectory defined by Eq. (14) can penetrate through any

hypersurface  $S_k$  except at a common intersection of all such hypersurfaces, that is, when  $g(x) = 0$ . Hence, these hypersurfaces (of all dimensions) serve to “funnel” the trajectory into each solution point.

### Appendix C. Finding multiple extrema of a function of $n$ variables

The problem of determining the global maximum and/or minimum of a scalar function of  $n$  variables is crucial in the field of optimization. This problem may be approached (though not definitively solved) using the present method of solving nonlinear equations to find multiple zeroes of the gradient of the function. Since the gradient vector vanishes at each stationary point of the function, a corresponding number of these stationary points are thereby located. Then, by examining the Hessian matrix of the function (which is identical with the Jacobian matrix of the gradient) a determination can be made as to whether a particular stationary point is a maximum, a minimum, or a saddle point. Thus, the “global” maximum and/or minimum—at least among those stationary points actually found—can be identified.

Although the present method cannot guarantee finding all of the zeroes of the gradient, which are needed to insure identifying the true global maximum and/or minimum of the function, the method does constitute a useful advance in the techniques of nonlinear optimization.

The procedures already described are applied directly to the gradient of the function being examined. As each solution point of  $g(x) = 0$  is found, the corresponding value of  $f(x)$  is saved and the Hessian (Jacobian) matrix examined. If the Hessian matrix is positive definite, the point is a local maximum; if the Hessian is negative definite, the point is a local minimum. Finally, if the Hessian is indefinite, the point is a saddle point.

The necessary and sufficient condition for the Hessian to be positive definite is that its determinant and the determinants of all of its principal minors must be positive. These determinants may be computed by multiplying the appropriate diagonal terms of the  $U$  factor (upper triangular matrix) obtained from the  $LU$  decomposition provided that proper account has been taken of the effects of row-column interchanges on the signs of these determinants [19]. Accordingly, the “adjusted” signs of all the diagonal elements of  $U$  (adjusted with respect to row-column interchanges) must be positive if  $f(x)$  is to be a local maximum.

Correspondingly, the necessary and sufficient condition for the Hessian to be negative definite is that the adjusted signs of the diagonal terms of  $U$  be negative; this will indicate that  $f(x)$  is a local minimum. Finally, if the adjusted signs of these diagonal terms are mixed, a saddle point has been found. Thus, for negligible extra cost beyond that required for the  $LU$  decomposition of  $H$ —which is

needed anyway in Eq. (24)—the type of each stationary point can be identified.

#### • Use of quasi-Newton methods

In order to avoid the computational cost of evaluating the Hessian matrix and carrying out the  $LU$  decomposition at each iteration for Eqs. (24) and (31), it is possible to use Broyden's approximation method for updating the inverse Hessian [3]. This may be particularly advantageous when the Hessian involves very complicated functions. Using the definitions

$$H_n = \frac{\partial g(x_n)}{\partial x}, \quad (C1)$$

$$B_n = H_n^{-1}, \quad (C2)$$

$$p_n = -B_n g_n, \text{ and} \quad (C3)$$

$$y_n = g_{n+1} - g_n, \quad (C4)$$

it follows straightforwardly that Broyden's “full step” formula for updating the inverse Hessian is

$$B_{n+1} = B_n + \frac{(\pm h_n p_n - B_n y_n) p_n^t B_n}{p_n^t B_n y_n}. \quad (C5)$$

By factoring out  $B_n$ , the right-hand side of Eq. (C5) can be written in a more convenient form so that

$$B_{n+1} = (I + a_n b_n^t) B_n, \quad (C6)$$

where  $I$  is the identity matrix,

$$a_n = \pm h_n p_n - B_n y_n \text{ and} \quad (C7)$$

$$b_n = p_n / p_n^t B_n y_n. \quad (C8)$$

Now Eq. (C6) can be used to compute the determinant of  $B_{n+1}$  as follows:

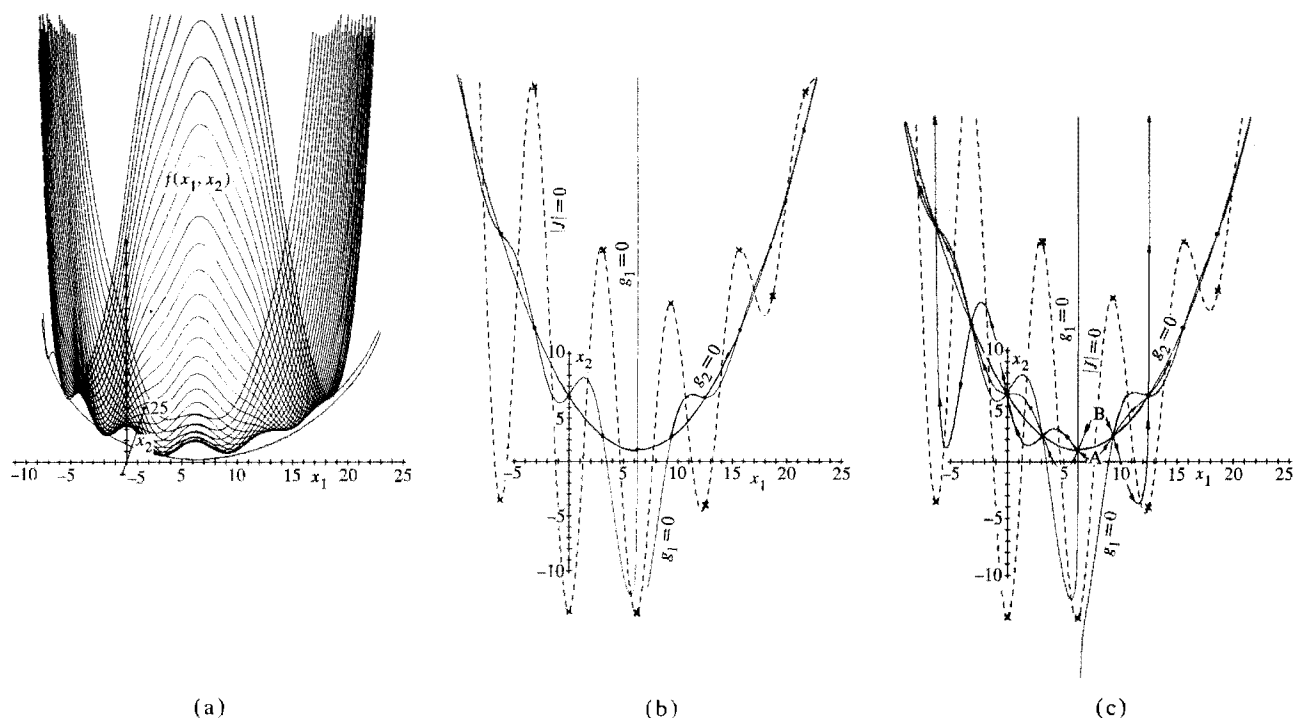
$$|B_{n+1}| = |I + a_n b_n^t| \cdot |B_n| = (1 + a_n^t b_n) \cdot |B_n|. \quad (C9)$$

If the initial Hessian,  $H_0$ , has been calculated explicitly and its determinant found, this defines the determinant of  $B_0$  and permits the determinant of  $B_n$  to be computed by the expression

$$|B_n| = |B_0| \prod_{k=1}^n (1 + a_k^t b_k). \quad (C10)$$

Thus, the changes in sign of  $|B_n|$  can be monitored simply by observing the changes in sign of the successive expressions  $(1 + a_n^t b_n)$ .

If  $B_n$  is a good enough approximation to  $H_n^{-1}$ , then the sign of  $|B_n|$  will change at the same time as that of  $|H_n|$ , thus enabling Eq. (C10) to be used for deciding when to change signs in Eq. (24). In practice, however, this procedure proves to be unreliable since  $|B_n|$  often changes sign a few steps before or after  $|H_n|$  does. It is possible to recover from this discrepancy at the cost of actually evaluating  $H_n$  whenever  $|B_n|$  changes sign and



**Figure C1** Undulating valley function. (a) Curves defined by Eq. (C12); (b) gradient equations; (c) two trajectories through several solution points.

retracing any erroneous steps. But this is not recommended.

If a polynomial or spline curve-fitting technique is used in place of Eq. (24) to predict  $x_{n+1}$ , the  $B_n$  matrix can still be used in Eq. (31) during application of the steering correction. Thus, the computational advantages of Broyden's updating algorithm can be realized. This approach has not yet been fully developed, but preliminary experiments indicate that it has considerable promise.

#### • Experimental results

Earlier we described the application of the present method to the problem of finding the minimum of Rosenbrock's function,

$$f(x_1, x_2) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2, \quad (\text{C11})$$

which is characterized by having a curving valley with steep walls. A similar function whose curving valley has an undulating floor is the following:

$$f(x_1, x_2) = a(x_2 - bx_1^2 + cx_1 - d)^2 + e(1 - f) \cos x_1 + e. \quad (\text{C12})$$

The valley follows the parabolic curve defined by the equation

$$x_2 - bx_1^2 + cx_1 - d = 0 \quad (\text{C13})$$

and the undulation of the valley floor is due to the cosine term.

A perspective drawing is shown in Fig. C1(a) for Eq. (C12) with  $a = 1$ ,  $b = 5/4\pi^2$ ,  $c = 5/\pi$ ,  $d = 6$ ,  $e = 10$ , and  $f = 1/8\pi$ . This function has an infinite number of minima, of which six can be seen. Included in the drawing is the parabola of Eq. (C13) as well as its projection on the surface  $f(x_1, x_2)$ .

The gradient equations, whose solutions determine the locations of the stationary points of this function, are

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= g_1(x_1, x_2) = 2a(x_2 - bx_1^2 + cx_1 - d)(-2bx_1 + c) \\ &\quad - e(1 - f) \sin x_1 = 0 \end{aligned}$$

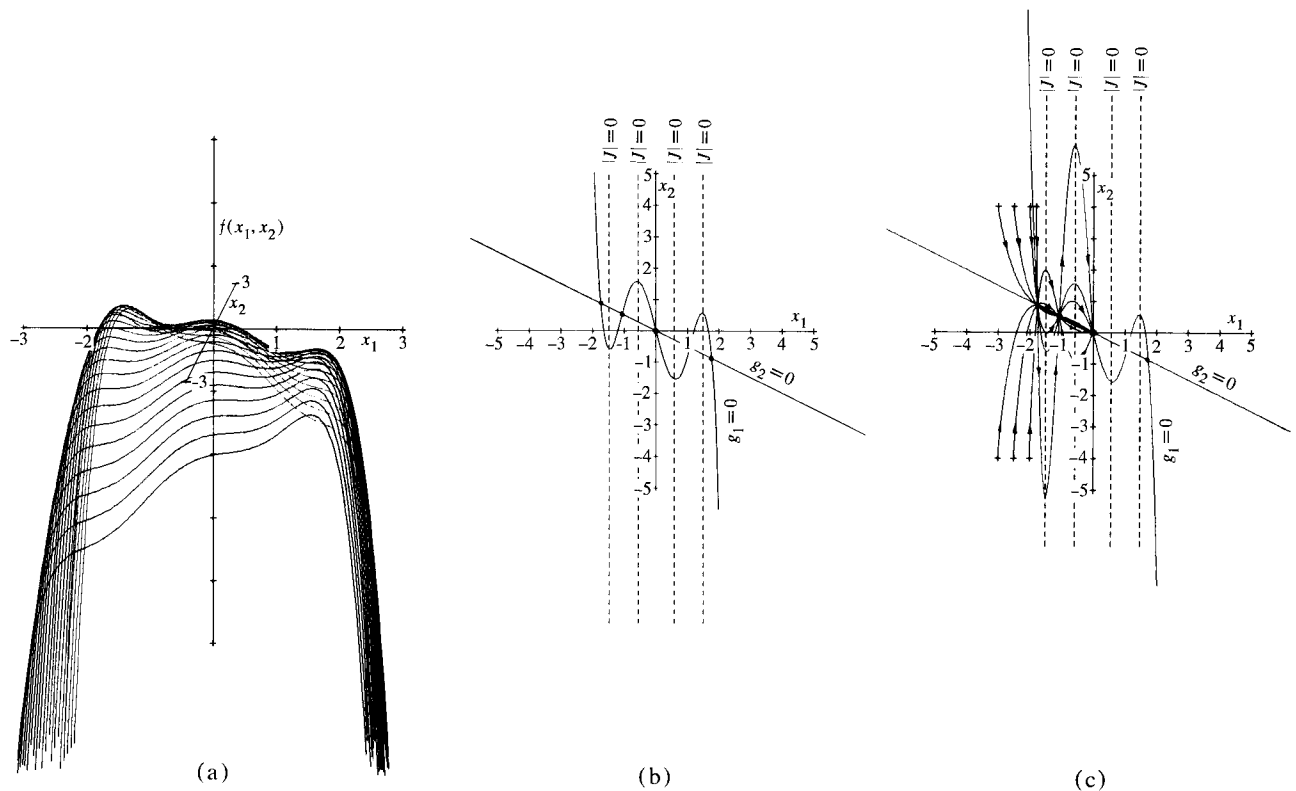
and

$$\frac{\partial f}{\partial x_2} = g_2(x_1, x_2) = 2a(x_2 - bx_1^2 + cx_1 - d) = 0. \quad (\text{C14})$$

The corresponding curves are shown in Fig. C1(b). Here the dashed line depicts the locus of points for which  $|H| = 0$ . The cross marks on this locus designate the extraneous singular points of the differential equation, Eq. (14). All of these singular points turn out to be saddle points.

In Fig. C1(c), two trajectories are shown passing through several solution points of Eq. (C14). These trajectories, like those in subsequent drawings, were computed by means of Eqs. (24)–(32). In Table 1, the coordinates of a number of the stationary points of  $f(x_1, x_2)$  are listed





**Figure C2** Camelback function with three humps. (a) Curves defined by Eq. (C15); (b) gradient equations; (c) trajectories through five solution points.

**Table 1** Stationary points for Fig. C1(a).

$x_1$	$x_2$	$f(x_1, x_2)$	Type
-9.4536	32.3651	-3.7557	minimum
-6.3149	21.1013	22.5063	saddle
-3.1769	12.3344	-1.2570	minimum
-0.0397	6.0634	20.0079	saddle
3.0962	2.2863	1.2409	minimum
6.2303	1.0003	17.5105	saddle
9.3614	2.2001	3.7373	minimum
12.4874	5.8751	15.0157	saddle
15.6034	12.0019	6.2291	minimum
18.6954	20.5124	12.5309	saddle
21.7071	31.1303	8.6915	minimum

**Table 2** Stationary points for Fig. C2(a).

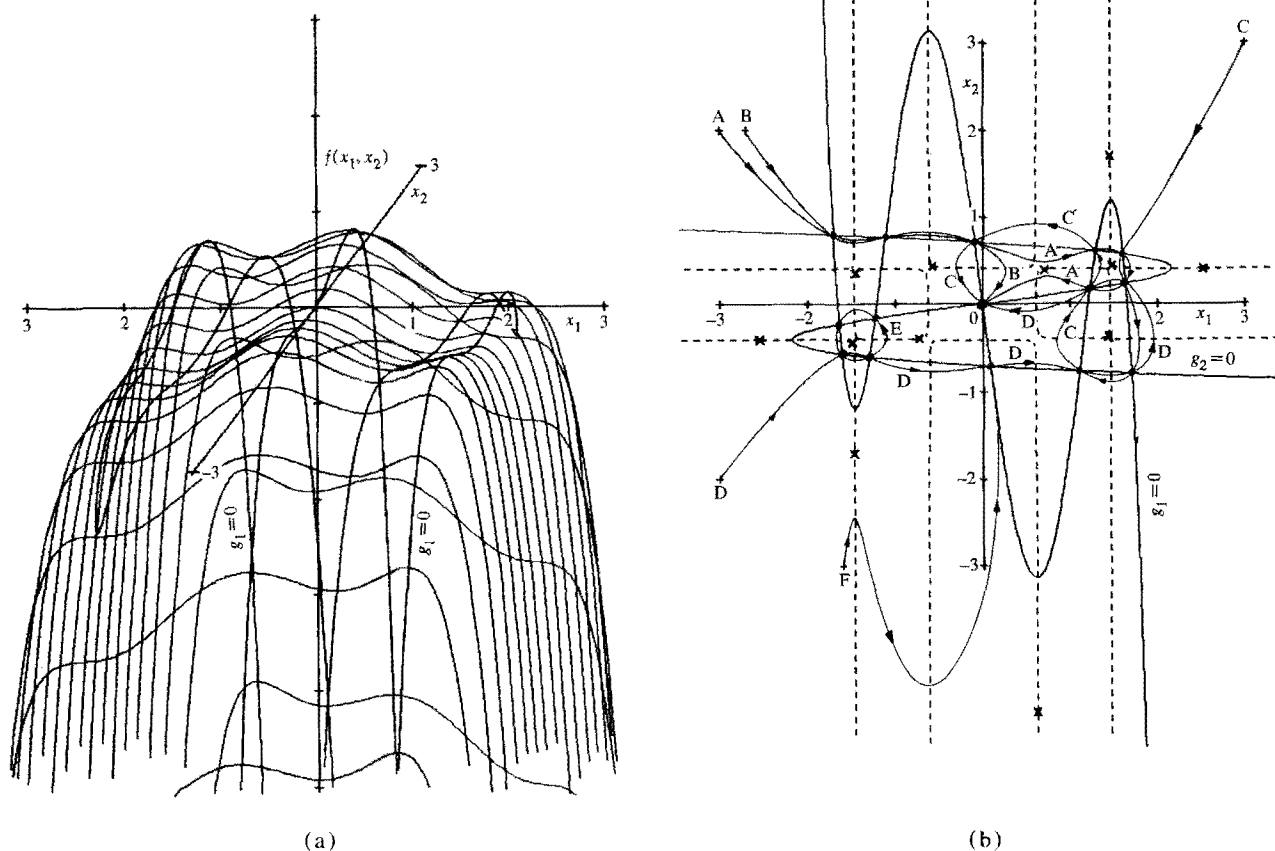
$x_1$	$x_2$	$f(x_1, x_2)$	Type
-1.7475	0.8737	-0.2986	maximum
-1.0705	0.5352	-0.8773	saddle
0.0000	0.0000	0.0000	maximum (global)
1.0705	-0.5352	-0.8773	saddle
1.7475	-0.8737	-0.2986	maximum

along with the corresponding function values and type of stationary point.

Notice that the trajectory marked A exhibits a decreasing undulation as it moves to the right. Once this trajectory passes through the solution at (9.36, 2.20), it gets captured between the curves  $g_1 = 0$  and  $g_2 = 0$ . Since it cannot escape this captivity, by virtue of the "funnelling theorem" described in Appendix B, it must certainly pass through all the solution points along the right hand branch of the curve  $g_1 = 0$ .

By contrast, as this trajectory moves to the left, it necessarily crosses to the outside of the area bounded by  $g_1 = 0$  and  $g_2 = 0$ . As a result, the amplitude of its undulation increases and eventually a point of instability is reached, beyond which the trajectory recedes to infinity. Thus, it passes through only a finite number of the solution points on the left hand branch of the curve  $g_1 = 0$ . A similar situation, but with the reverse sense of direction, occurs on trajectory B. This trajectory passes through all of the solutions on the left hand branch of  $g_1 = 0$  and through only a few of those on the right hand branch. Thus, these two trajectories are sufficient to locate all of the stationary points of Eq. (C12).

Another problem that has been solved by the present method is that of finding the global maximum of the "camelback" function,



**Figure C3** Camelback function with six humps. (a) Curves defined by Eq. (C15) with different values of constants from those used in plotting Fig. C2(a); (b) trajectories through solution points.

$$f(x_1, x_2) = ax_1^2 + bx_1^4 + cx_1^6 - x_1x_2 + dx_2^2 + ex_2^4. \quad (C15)$$

This function has three maxima when  $a = -2$ ,  $b = 1.05$ ,  $c = -1/6$ ,  $d = -1$  and  $e = 0$ , as shown in Fig. C2(a). The corresponding gradient equations,

$$\frac{\partial f}{\partial x_1} = g_1(x_1, x_2) = 2ax_1 + 4bx_1^3 + 6cx_1^5 - x_2 = 0$$

and

$$\frac{\partial f}{\partial x_2} = g_2(x_1, x_2) = -x_1 + 2dx_2 + 4ex_2^3 = 0 \quad (C16)$$

are depicted in Fig. C2(b). This instance of Eq. (C16), with  $e = 0$ , can be shown to be free of extraneous singular points since the equations are a special case of those used to describe the "tunnel diode" problem, which has no extraneous singularities. (A fifth-degree polynomial was used to represent the tunnel diode characteristic curve.)

In all the problems that have been studied using the present method, the presence of at least one extraneous

singular point has always appeared to be a necessary condition for the existence of a region of nonconvergence and for the deflection of a trajectory so that it fails to pass through all solution points. Therefore, even in the absence of a mathematical proof of its validity, it is reasonable to make the claim that for the case depicted in Fig. C2(b), the method will be globally convergent and all trajectories (except those whose  $x_1$  coordinate is identical to that of any solution point) will pass through every solution. This conclusion is borne out by the trajectories shown in Fig. C2(c). (Here, only half of each trajectory is shown since the problem is centrosymmetric.) The coordinates and function values of all five stationary points are listed in Table 2.

By using the values  $a = -4$ ,  $b = 2.1$ ,  $c = -1/3$ ,  $d = 4$ ,  $e = -4$ , Eq. (C15) may be made to exhibit six maxima, as shown in Fig. C3(a). The corresponding zero gradient curves are shown in Fig. C3(b) along with several  $|H| = 0$  loci and a number of extraneous singularities, all of which

**Table 3** Stationary points for Fig. C3(a).

$x_1$	$x_2$	$f(x_1, x_2)$	Type
-1.7035	0.7960	0.2154	maximum
-1.1092	0.7682	-0.5437	saddle
-0.0898	-0.7126	1.0316	maximum (global)
1.2960	0.6050	-2.2294	saddle
1.6071	0.5686	-2.1042	maximum
1.6380	0.2287	-2.2293	saddle
1.2302	0.1623	-2.4962	minimum
0.0000	0.0000	0.0000	saddle

(The remaining seven points are conjugate to the first seven.)

prove to be saddle points. Of the five trajectories shown, only those labelled *A*, *C*, and *D* pass through all fifteen solution points. (Again, only half of each centrosymmetric trajectory is drawn; this reduces clutter in the figure.)

Trajectory *B*, however, passes through only seven solution points; it is deflected by the saddle point at (0.72, 0.40) so that it entirely misses the right-most four solutions. These four points are threaded by a cyclic trajectory conjugate to the one labelled *E*. Thus, even though there are many extraneous singularities present, these trajectories and others that have been examined give no evidence that any regions of nonconvergence exist in this problem. Accordingly, it is an easy task to find all the stationary points in this problem and to identify the global maximum. The pertinent data are listed in Table 3.

## References

1. F. H. Branin, Jr., "Solution of Nonlinear DC Network Problems via Differential Equations," *Mem. Mexico 1971 Internat. IEEE Conference on Systems, Networks and Computers*, Oaxtepec, Mexico, pp. 93-101.
2. F. H. Branin, Jr. and S. K. Hoo, "A Method for Finding Multiple Extrema of a Function of  $N$  Variables," *Proceedings of the Conference on Numerical Methods for Nonlinear Optimization*, University of Dundee, Scotland, June 28-July 1, 1971, *Numerical Methods of Nonlinear Optimization*, Academic Press, London, (August 1972).
3. C. G. Broyden, "A Class of Methods for Solving Nonlinear Simultaneous Equations," *Math. Comp.* **19**, 577-593 (1965).
4. L. E. Kugel, private communication.
5. F. A. Ficken, "The Continuation Method for Functional Equations," *Comm. Pure Appl. Math.*, **4**, 435-456 (1951).
6. D. F. Davidenko, "On a New Method of Numerical Solution of Systems of Nonlinear Equations," *Dokl. Akad. Nauk SSSR* **88**, 601-602 (1953).
7. M. K. Gavurin, "Nonlinear Functional Equations and Continuous Analogs of Iterative Methods," *Izv. Vyssh. Uchebn. Zaved. Matematika*, **6**, No. 5, 18-31 (1958).
8. M. N. Jacovlev, "On the Solution of Systems of Nonlinear Equations by Differentiation with Respect to a Parameter," *USSR Comput. Math. and Math. Phys.* **4**, 146-149 (1964).
9. G. H. Meyer, "On Solving Nonlinear Equations with a One-parameter Operator Imbedding," *SIAM J. Numer. Anal.* **5**, 739-752 (1968).
10. P. T. Boggs, "The Solution of Nonlinear Systems of Equations by  $A$ -Stable Integration Techniques," *SIAM J. Numer. Anal.* **8**, No. 4, 767-785 (1971). Also Ph.D. Dissertation, Cornell University (1970).
11. J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York (1970), pp. 230-236, 334-341.
12. R. Brent, private communication.
13. W. J. Cunningham, *Introduction to Nonlinear Analysis*, McGraw-Hill Book Co., New York, 1958.
14. W. Hurewicz, *Lectures on Ordinary Differential Equations*, John Wiley and Sons, New York 1961.
15. R. A. Frazer, W. J. Duncan, and A. R. Collar, *Elementary Matrices*, Macmillan, New York, (1947).
16. G. E. Forsythe and C. B. Moler, *Computer Solution of Linear Algebraic Systems*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1967.
17. C. G. Broyden, "A New Method of Solving Nonlinear Simultaneous Equations," *Comp. J.* **12**, 94-99 (1969).
18. H. H. Rosenbrock, "An Automatic Method for Finding the Greatest or Least Value of a Function," *Comp. J.* **3**, 175-184 (1960).
19. J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965, pp. 237-239.

Received February 15, 1972

The author is located at the Systems Development Division Laboratory in Kingston, New York 12401.