

**Computer Science and Applied Mathematics**  
A SERIES OF MONOGRAPHS AND TEXTBOOKS

Editor  
**Werner Rheinboldt**  
*University of Maryland*

Hans P. Küinzi, H. G. Tzschach, and C. A. Zehnder  
NUMERICAL METHODS OF MATHEMATICAL OPTIMIZATION: WITH ALGOL  
AND FORTRAN PROGRAMS, 1968

Azriel Rosenfeld  
PICTURE PROCESSING BY COMPUTER, 1969

James Ortega and Werner Rheinboldt  
ITERATIVE SOLUTION OF NONLINEAR EQUATIONS IN  
SEVERAL VARIABLES, 1970

INSTITUTO SUPERIOR DE CIÊNCIAS  
CENTRO DE CÁLCULO CIENTÍFICO

3949

518.6 : 517.9

**ITERATIVE SOLUTION  
OF NONLINEAR EQUATIONS  
IN SEVERAL VARIABLES**

*J. M. Ortega and W. C. Rheinboldt*

UNIVERSITY OF MARYLAND  
COLLEGE PARK, MARYLAND



ACADEMIC PRESS New York and London

INSTITUTO SUPERIOR TECNICO



inverse defined on all of  $Y$ . More generally, the method has been considered on topological linear spaces (Hirasawa [1954]). Newton's method has been studied extensively in Banach spaces by many authors, most notably L. Kantorovich and his colleagues (see Kantorovich and Akilov [1959]), L. Collatz and his students (see Collatz [1964]), and R. Bellman and his associates under the name *quasi-linearization* (see Bellman and Kalaba [1965]).

**NR 7.1-4.** Crockett and Chernoff [1955] were apparently the first to consider the iteration (21) in a formal way; see also Gleyzal [1959]. The iteration (22) dates back at least to Levenberg [1944]. Convergence results for these two iterations are given in 10.2, 11.2, and 14.4.

**NR 7.1-5.** Traub [1964] has studied the iteration (24) and shown that it possesses the property of "cubic convergence" for  $m = 2$  and even higher-order convergence for  $m > 2$ . We give these results in 10.2. Šamanski [1967a] has considered the same procedure with  $F'(x)$  replaced by the approximation (16) with  $h_{ij} = h, i, j = 1, \dots, n$ .

**NR 7.1-6.** Several authors (see, e.g., Altman [1961a]; Jankó [1962b]; Lika [1965]; Mertveova [1953]; Nečepurenko [1954]; Šafiev [1964]; and Traub [1964]) have considered extensions to  $n$  dimensions, and also infinite dimensions, of other one-dimensional methods which possess a high rate of convergence. Such a typical process is the *method of tangent hyperbolas*, which may be formulated as

$$x^{k+1} = x^k - \left\{ f - \frac{1}{3} F'(x^k)^{-1} F''(x^k) F'(x^k)^{-1} F x^k \right\}^{-1} F'(x^k)^{-1} F x^k, \quad (25)$$

and can be shown to exhibit cubic convergence. However, methods of this type, which require second- and higher-order derivatives, are rather cumbersome from the computational viewpoint. Note that, while computation of  $F'(x)$  involves only the  $n^2$  first partial derivatives  $\partial_j f_i$ , computation of  $F''$  requires the  $n^3$  second derivatives  $\partial_j \partial_s f_i$ ; in general, an exorbitant amount of work. Indeed, much recent research has been devoted to finding methods needing fewer derivative computations than Newton's method. One, not very satisfactory approach to reducing the derivative requirements in methods such as (26) is to consider difference analogs similar to the secant analogs of Newton's method; for work along these lines, see Ufm [1963b, 1965b].

## EXERCISES

**E 7.1-1.** Assume that  $F: R^n \rightarrow R^n$  is  $G$ -differentiable at  $x$  and that  $F'(x)$  is invertible. Set  $y = F'(x)^{-1} Fx$  and show that there exist  $\lambda > 0$  so that

$$\|F(x - \lambda y)\| \leq \|Fy\|, \quad (27)$$

Show, moreover, that if

$$\|F'(x - ty) - F'(x)\| \leq Kt\|y\|, \quad \forall t \in [0, 1],$$

then (27) holds for all  $\lambda \in [0, \|Fy\| / (K\|y\|^2)]$ .

**E 7.1-2.** For  $A, B \in L(R^n)$ , the eigenvalue problem  $Ax = \lambda Bx, x^T x = 1$ , is equivalent with the equation  $Fx = 0$ , where

$$F: R^n \times R^1 \rightarrow R^{n+1}, \quad F \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} Ax - \lambda Bx \\ x^T x - 1 \end{pmatrix}.$$

Write down Newton's method for this mapping  $F$ .

## 7.2. SECANT METHODS

The discretized Newton methods discussed in the last section constitute  $n$ -dimensional generalizations of the one-dimensional discretized Newton methods:

$$x^{k+1} = x^k - \left[ \frac{f(x^k + h^k) - f(x^k)}{h^k} \right]^{-1} f(x^k), \quad k = 0, 1, \dots, \quad (1)$$

Two important special cases of (1) are the *regula falsi* iteration

$$x^{k+1} = x^k - \left[ \frac{f(\bar{x}) - f(x^k)}{\bar{x} - x^k} \right]^{-1} f(x^k), \quad k = 0, 1, \dots, \quad (2)$$

in which  $h^k = \bar{x} - x^k$  for some fixed  $\bar{x}$ , and the *secant method*

$$x^{k+1} = x^k - \left[ \frac{f(x^{k-1}) - f(x^k)}{x^{k-1} - x^k} \right]^{-1} f(x^k), \quad k = 0, 1, \dots, \quad (3)$$

where  $h^k = x^{k-1} - x^k$ .

It is possible to use similar choices of  $h^k$  in the  $n$ -dimensional discretized Newton methods. However, in order to discuss the resulting methods in suitable generality, it is desirable to begin with a somewhat different approach to the one-dimensional methods (1). As drawn in Fig. 7.3, the next iterate  $x^{k+1}$  of (1) is the solution of the linearized equation

$$l(y) = \left[ \frac{f(x^k + h^k) - f(x^k)}{h^k} \right] (x - x^k) + f(x^k) = 0.$$

The important point now is that  $l$  can be viewed in two different ways; either it is regarded as an approximation of the tangent line

$l_1(x) = f'(x^k)(x - x^k) + f(x^k)$ , or as a linear interpolation of  $f$  between the points  $x^k$  and  $x^k + h^k$ . In generalizing (1) to  $n$  dimensions, we arrive at different methods depending on which view we take. In the case of the discretized Newton methods, the first interpretation was used, and we replaced the derivative  $f'(x^k)$  by a matrix  $J(x^k, h^k)$  of difference quotients approximating it.

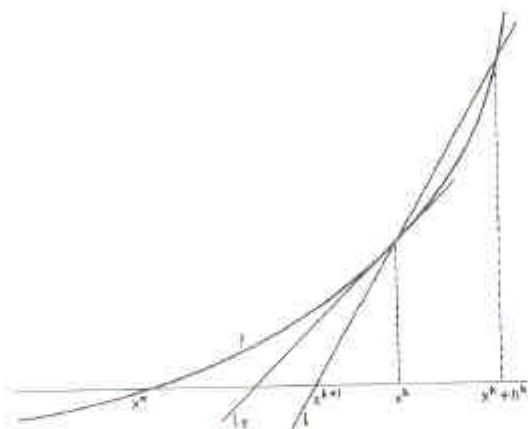


FIGURE 7.3

In order to extend the second point of view to  $n$  dimensions, we replace each "component surface"  $f_i$ ,  $i = 1, \dots, n$ , in  $R^{n+1}$  by a hyperplane which interpolates  $f_i$  at  $n + 1$  given points  $x^{k,j}$ ,  $j = 0, \dots, n$ , in a neighborhood of  $x^k$ . That is, a vector  $a^i$  and a scalar  $\alpha_i$  are to be found such that the affine mapping  $L_i x = \alpha_i + x^T a^i$  satisfies

$$L_i x^{k,j} = f_i(x^{k,j}), \quad j = 0, 1, \dots, n.$$

The next iterate  $x^{k+1}$  is then obtained as the intersection of these  $n$  hyperplanes in  $R^{n+1}$  with the hyperplane  $x = 0$ ; that is,  $x^{k+1}$  is the solution of the linear system  $L_i x = 0$ ,  $i = 1, \dots, n$ . This describes the *general secant method* in  $n$  dimensions. Depending on the choice of the interpolation points  $x^{k,j}$ ,  $j = 0, \dots, n$ , there are numerous possible different specific methods, but, before giving any of these, we develop some results on  $n$ -dimensional linear interpolation in order to see how the next iterate can actually be calculated.

**7.2.1. Definition.** Any  $n + 1$  points  $x^0, \dots, x^n$  in  $R^n$  are in *general position* if the vectors  $x^j - x^i$ ,  $j = 1, \dots, n$  are linearly independent. ■

This definition appears to depend upon the order of enumeration of the  $x^i$ , but this is not the case. Indeed, we have the following equivalent conditions for points to be in general position.

**7.2.2.** Let  $x^0, \dots, x^n$  be any  $n + 1$  points in  $R^n$ . Then the following statements are equivalent:

- $x^0, \dots, x^n$  are in general position.
- For any  $j$ ,  $0 \leq j \leq n$ , the vectors  $x^i - x^j$ ,  $i = 0, \dots, n$ ,  $i \neq j$ , are linearly independent.
- The  $(n + 1) \times (n + 1)$  matrix  $(e, X^T)$ , where  $e^T = (1, \dots, 1)$  and  $X = (x^0, \dots, x^n)$ , is nonsingular.
- For any  $y \in R^n$ , there exist scalars  $\alpha_0, \dots, \alpha_n$  with  $\sum_{i=0}^n \alpha_i = 1$  such that  $y = \sum_{i=0}^n \alpha_i x^i$ .

*Proof.* By the matrix identity

$$\begin{aligned} & \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ x^0 & x^0 - x^j & \dots & x^{j-1} - x^j & x^{j+1} - x^j & \dots & x^n - x^j \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ x^j & x^0 & \dots & x^{j-1} & x^{j+1} & \dots & x^n \end{pmatrix} \begin{pmatrix} 1 & -1 & \dots & -1 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 \end{pmatrix} \end{aligned}$$

we have

$$\begin{aligned} & \det(x^k - x^j, \dots, x^{j-1} - x^j, x^{j+1} - x^j, \dots, x^n - x^j) \\ &= \det \begin{pmatrix} 1 & 1 & \dots & 1 & 1 & \dots & 1 \\ x^j & x^0 & \dots & x^{j-1} & x^{j+1} & \dots & x^n \end{pmatrix} = (-1)^j \det(e, X^T), \end{aligned}$$

for any  $j = 0, \dots, n$ , which shows the equivalence of (a), (b) and (c).

Finally, (d) is equivalent to the statement that the linear system

$$\begin{pmatrix} e^T \\ X \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} 1 \\ y \end{pmatrix} \quad (4)$$

has a solution for any  $y$ , so that, clearly, (c) implies (d). Conversely, by solving (4) for  $y$  successively set equal to  $0, e^1, \dots, e^n$ , we see that  $(e, X^T)$  is nonsingular. ■

The geometrical interpretation of general position is that the points  $x^0, \dots, x^n$  do not lie in an affine subspace of dimension less than  $n$ . Thus, for  $n = 2$ , the points  $x^0, x^1, x^2$  are in general position if they are not colinear, that is, if they do not lie on a line in  $R^2$ . Note, however, that the vectors  $x^0, \dots, x^n$  may span  $R^n$  even if they are not in general position (see E 7.2-6).

The following result now gives a complete answer to the linear interpolation problem in  $R^n$ .

**7.2.3.** Let  $x^0, \dots, x^n$  and  $y^0, \dots, y^n$  be given points in  $R^n$ . Then there exists a unique affine function  $Lx = a + Ax$ , where  $a \in R^n$  and  $A \in L(R^n)$ , such that  $Lx^j = y^j, j = 0, \dots, n$ , if and only if  $x^0, x^1, \dots, x^n$  are in general position. Moreover,  $A$  is nonsingular if and only if  $y^0, \dots, y^n$  are in general position.

*Proof.* The conditions  $Lx^j = y^j, j = 0, 1, \dots, n$ , can be written in matrix form as

$$(a, X^T) \begin{pmatrix} a^T \\ A^T \end{pmatrix} = (y^0, \dots, y^n)^T, \quad (5)$$

where again  $e^T = (1, 1, \dots, 1)$  and  $X = (x^0, \dots, x^n)$ . Hence, the first part is an immediate consequence of 7.2.2. Now,  $Lx^j = y^j, j = 0, \dots, n$ , implies that

$$A(x^j - x^0) = y^j - y^0, \quad j = 1, \dots, n. \quad (6)$$

and, since  $x^j - x^0, j = 1, \dots, n$ , are linearly independent, it follows that  $A$  is nonsingular if and only if the vectors  $y^j - y^0, j = 1, \dots, n$  are linearly independent, and, hence, if and only if  $y^0, \dots, y^n$  are in general position. ■

In line with these results, one step of any general secant method can now be phrased as follows.

**7.2.4. Definition.** Let  $F: D \subset R^n \rightarrow R^n$  and assume that the two sets of points  $x^0, \dots, x^n \in D$  and  $Fx^0, \dots, Fx^n$  are in general position. Then the point

$$x^* = -A^{-1}a, \quad (7)$$

where  $a$  and  $A$  satisfy

$$a + Ax^j = Fx^j, \quad j = 0, \dots, n, \quad (8)$$

is a *basic secant approximation* with respect to  $x^0, \dots, x^n$ . ■

Note that 7.2.3 ensures that  $x^*$  is well defined. Note also that in one dimension the conditions reduce to  $x^0 \neq x^1$  and  $f(x^0) \neq f(x^1)$ , which are just the conditions which ensure that the unique secant line intersects the  $x$ -axis.

The computation of a basic secant approximation may be carried out by finding  $a$  and  $A$  to satisfy (8)—which, in turn, requires solving the linear system (5) with  $y^j = Fx^j$ —and then by solving  $a + Ax = 0$ . However, it turns out that there is no need to compute the interpolating function  $a + Ax$  explicitly. We consider next two alternative formulations, both of which show that  $x^*$  can be obtained by solving only one linear system.

**7.2.5. Wolfe Secant Formulation.** Let  $x^0, \dots, x^n$ , as well as  $Fx^0, \dots, Fx^n$ , be in general position. Then the basic secant approximation satisfies

$$x^* = Xz = \sum_{i=0}^n z_i x^i, \quad (9)$$

where  $z = (z_0, \dots, z_n)^T$  is the unique solution of the  $(n+1) \times (n+1)$  linear system

$$\begin{pmatrix} 1 & & \\ Fx^0 & \dots & Fx^n \end{pmatrix} z = (1, 0, \dots, 0)^T, \quad (10)$$

*Proof.* Since the  $Fx^j$  are in general position, 7.2.2(b) ensures that (10) has a unique solution which satisfies  $\sum_{i=0}^n z_i = 1$  and  $\sum_{i=0}^n z_i Fx^i = 0$ . Hence, by (8),

$$0 = \sum_{i=0}^n z_i Fx^i = \sum_{i=0}^n z_i (a + Ax^i) = a + A \left( \sum_{i=0}^n z_i x^i \right),$$

and, since  $x^*$  is the unique solution of  $Ax + a = 0$ , (9) holds. ■

Note that (9) and (10) uniquely determine a vector  $x^*$  provided only that  $Fx^0, \dots, Fx^n$  are in general position; therefore, the Wolfe formulation can be carried out even if  $x^0, \dots, x^n$  are not in general position. In that case, however,  $x^*$  will lie in the lower-dimensional affine subspace  $\{x \mid x = \sum_{i=0}^n c_i x^i, \sum_{i=0}^n c_i = 1\}$  and no linear interpolator  $a + Ax$  can exist such that  $a + Ax^j = Fx^j, j = 0, \dots, n$  (see E 7.2-7).

For the next formulation, it is convenient to introduce the operator

$$J: D_H \subset R^n \times L(R^n) \rightarrow L(R^n)$$

defined by

$$J(x, H) = (F(x + He^1) - Fx, \dots, F(x + He^n) - Fx) H^{-1} \quad (11)$$

where, if  $D$  is the domain of definition of  $F$ ,

$$D_x = \{(x, H) \mid x + He^i \in D, \quad i = 1, \dots, n; \quad H \text{ nonsingular}\}.$$

**7.2.6. Newton Formulation.** Assume that  $x^0, \dots, x^n$  and  $Fx^0, \dots, Fx^n$  are in general position, and set

$$H = (x^1 - x^0, \dots, x^n - x^0). \quad (12)$$

Then  $J(x^0, H)$  is nonsingular and the basic secant approximation  $x^1$  is given by

$$x^1 = x^0 - J(x^0, H)^{-1}Fx^0. \quad (13)$$

*Proof.* Since  $Fx^1 = F(x^0 + He^1)$ , it follows from (6) with  $y^j = Fx^j$  that

$$AH = (F(x^0 + He^1) - Fx^0, \dots, F(x^0 + He^n) - Fx^0)$$

so that, since  $H$  is nonsingular,  $A = J(x^0, H)$ . Thus 7.2.3 ensures that  $J(x^0, H)$  is nonsingular, and, from  $x^1 = -A^{-1}a$  and  $a = Fx^0 - Ax^0$ , we obtain

$$x^1 = -A^{-1}(-Ax^0 + Fx^0) = x^0 - J(x^0, H)^{-1}Fx^0. \quad \blacksquare$$

Note that if we set  $T = (Fx^1 - Fx^0, \dots, Fx^n - Fx^0)$ , then (13) may be written in the form

$$x^1 = x^0 - HT^{-1}Fx^0. \quad (14)$$

Therefore, as with the Wolfe formulation, the Newton formulation may be carried out provided only that  $Fx^0, \dots, Fx^n$  are in general position. Again, however,  $x^1$  will then lie in the affine subspace

$$\left\{ x \mid x = \sum_{i=0}^n \epsilon_i x^i, \sum_{i=0}^n \epsilon_i = 1 \right\}$$

(see E 7.2-7).

Observe also that a basic secant approximation, by either the Wolfe or Newton formulation, requires indeed only the solution of one linear system of equations—namely, (10) in the first case, and  $Tx = Fx^0$  in the second—followed by the calculation of a linear combination of the vectors  $x^0, \dots, x^n$  by means of (9) or (14), respectively.

It is of interest to note that the following representation of  $x^1$  is also valid:

$$x^1 = x^0 - [(Fx^1 - Fx^0, Fx^1 - Fx^0, \dots, Fx^1 - Fx^0)(x^1 - x^0, \dots, x^n - x^{n-1})^{-1}]^{-1}Fx^0. \quad (15)$$

This is an immediate consequence of the following lemma, which will be of use in Chapter 11, since (15) follows from (16) if  $x = x^0$  and  $H$  is defined by (12).

**7.2.7.** Let  $J(x, H)$  be defined by (11) with  $H = (h^1, \dots, h^n)$ . Then

$$J(x, H) = (F(x + h^1) - Fx, F(x + h^2) - F(x + h^1), \dots, F(x + h^n) - F(x + h^{n-1}))\hat{H}^{-1} \quad (16)$$

where

$$\hat{H} = (h^1, h^2 - h^1, \dots, h^n - h^{n-1}).$$

*Proof.* We note that, for any  $n + 1$  vectors  $r^0, r^1, \dots, r^n \in R^n$ , we have

$$(r^1 - r^0, r^2 - r^1, \dots, r^n - r^{n-1})P = (r^1 - r^0, r^2 - r^1, \dots, r^n - r^{n-1}),$$

where  $P \in L(R^n)$  is defined by

$$P = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & -1 \\ \circ & & & & 1 \end{bmatrix}.$$

Clearly,  $P$  is nonsingular, and (16) then follows from

$$J(x, H) = (F(x + h^1) - Fx, \dots, F(x + h^n) - Fx)P(HP)^{-1}. \quad \blacksquare$$

The Newton formulation allows the general secant method described at the beginning of this section to be expressed in the compact form

$$\begin{cases} x^{k+1} = x^k - J(x^k, H_k)^{-1}Fx^k, & k = 0, 1, \dots, \\ H_k = (x^{k+1} - x^k, \dots, x^{k+n} - x^k), \end{cases} \quad (17)$$

where we have set  $x^{k,0} = x^k$ .

We consider next several possible ways of choosing the auxiliary points  $x^{k,1}, \dots, x^{k,n}$ . A first simple choice is given by

$$x^{k,j} = x^k + (x_j^{k-1} - x_j^k)e^j, \quad j = 1, \dots, n. \quad (18)$$

In this case,  $H_k$  is the diagonal matrix

$$H_k = \text{diag}(x_1^{k-1} - x_1^k, \dots, x_n^{k-1} - x_n^k),$$

and, if we set  $h_j^k = x_j^{k-1} - x_j^k$ ,  $j = 1, \dots, n$ , then

$$J(x^k, H_k) = ((1/h_1^k)[F(x^k + h_1^k e^1) - Fx^k], \dots, (1/h_n^k)[F(x^k + h_n^k e^n) - Fx^k]).$$

Substituting this into the Newton formulation (17), we see that the resulting method is exactly the same as the discretized Newton method (7.1.19) using the difference approximation (7.1.16) with  $h_{ij} = h_i$ ,  $i = 1, \dots, n$ . For future use, it is convenient to redefine  $J$  in this case as a mapping

$$\left\{ \begin{array}{l} J: D_J \times D_h \subset R^n \times R^n \rightarrow L(R^n) \\ D_J \times D_h = \{(x, h) \in R^n \times R^n \mid x + h_i e^i \in D, \quad h_i \neq 0, \quad i = 1, \dots, n\} \\ J(x, h) = (h_1^{-1}[F(x + h_1 e^1) - Fx], \dots, h_n^{-1}[F(x + h_n e^n) - Fx]), \end{array} \right. \quad (19)$$

and to write the method as

$$x^{k+1} = x^k - J(x^k, x^{k-1} - x^k)^{-1} Fx^k, \quad k = 0, 1, \dots \quad (20)$$

If we choose instead of (18) the points

$$x^{k+1} = x^k + \sum_{j=1}^n (x_j^{k-1} - x_j^k) e^j, \quad j = 1, \dots, n, \quad (21)$$

then a simple computation shows that the iteration (17) is exactly the same as the discretized Newton method using the difference approximation (7.1.15) with  $h_{ij} = x_j^{k-1} - x_j^k$ . In this case, we may define the iteration by (20), where  $J$  is now given by

$$J(x, h) = \left( h_1^{-1} [F(x + h_1 e^1) - Fx], \dots, h_n^{-1} \left[ F \left( x + \sum_{j=1}^n h_j e^j \right) - F \left( x + \sum_{j=1}^{n-1} h_j e^j \right) \right] \right). \quad (22)$$

More generally, we can consider the choice of auxiliary points

$$x^{k,j} = x^k + P_{j,k}(x^{k-1} - x^k), \quad j = 1, \dots, n, \quad (23)$$

where  $P_{j,k} \in L(R^n)$  are given linear operators. Clearly, (18) is the special case of (23) in which  $P_{j,k} = (0, \dots, 0, e^j, 0, \dots, 0)$ ,  $j = 1, \dots, n$ ,  $k = 0, 1, \dots$ , while (23) reduces to (21) if  $P_{j,k} = (e^1, \dots, e^j, 0, \dots, 0)$ ,  $j = 1, \dots, n$ ,  $k = 0, 1, \dots$ .

The choice (23) of the auxiliary points depends only on  $x^k$  and  $x^{k-1}$ . In general, if the auxiliary points  $x^{k,j}$  depend on precisely  $p$  of the previous iterates  $x^k, \dots, x^0$ , we say that the iteration (17) is a  $p$ -point secant method, while if the  $x^{k,j}$  depend upon  $x^k, \dots, x^{k-p+1}$ , it is called a sequential  $p$ -point secant method.

The iterations (17) with  $x^{k,j}$  given by (18) or (21) are examples of sequential two-point methods, while the iteration

$$x^{k+1} = x^k - J(x^k, H_k)^{-1} Fx^k, \quad H_k = (x^{k-1} - x^0, \dots, x^{k-n} - x^k), \quad k = 0, 1, \dots, \quad (24)$$

is a sequential  $(n+1)$ -point method. As an example of a nonsequential  $(n+1)$ -point method, the auxiliary points may be chosen from the set of previous iterates by the criterion that  $x^{k-1}, \dots, x^{k-n}$  are those  $n$  vectors among  $x^0, x^1, \dots, x^{k-1}$  for which the  $\|Fx^i\|$  are smallest.

General  $(p+1)$ -point methods may be generated in a variety of ways. For example, in analogy with (23), we may choose

$$x^{k,j} = x^k + \sum_{i=1}^p P_{i,j,k}(x^{k-i} - x^k), \quad j = 1, \dots, n, \quad (25)$$

where again the  $P_{i,j,k}$  are given linear operators.

In general, the secant method requires  $n+1$  evaluations of the function  $F$  at each stage—namely at the points  $x^k, x^{k-1}, \dots, x^{k-n}$ . In particular, this is true for the two-point method defined by (18). This amount of computation is comparable to that of Newton's method if the evaluation of  $f_j(x)$  takes about as much work as that of  $\partial_j f_j(x)$ .

In certain cases, however, the particular choice of the  $x^{k,j}$  permits fewer function evaluations. For example, if (21) is used, then  $x^{k,n} = x^{k-1}$ , and, since  $Fx^{k-1}$  is available from the previous stage, only  $n$  new function evaluations are required. The most spectacular saving, however, is available through the  $(n+1)$ -point method (24). Here, since  $Fx^{k-1}, \dots, Fx^{k-n}$  are already available (except at the first stage, when  $Fx^0, \dots, Fx^n$  must all be calculated), only one new function evaluation is required—namely,  $Fx^k$ . Moreover, another possible computational savings is available in the solution of the linear system of (24). To see this, recall from 7.2.7 that (24) can be written in the alternative form

$$x^{k+1} = x^k - H_k \Gamma_k^{-1} Fx^k, \quad (26)$$

where now

$$\begin{aligned} H_k &= (x^k - x^{k-1}, x^{k-2} - x^{k-3}, \dots, x^{k-n+1} - x^{k-n}), \\ \Gamma_k &= (Fx^k - Fx^{k-1}, \dots, Fx^{k-n+1} - Fx^{k-n}). \end{aligned} \quad (27)$$

Then we find:

**7.2.8.** Assume that the matrices  $\Gamma_p$  and  $\Gamma_{p+1}$  defined by (27) for

$k = p, p + 1$ , are both nonsingular, and denote the rows of  $\Gamma_p^{-1}$  by  $v^1, \dots, v^n$ . Then

$$\Gamma_{p+1}^{-1} = B - \frac{B(q^p - q^{p-n})v^n}{1 + v^n(q^p - q^{p-n})}, \quad (28)$$

where  $q^t = Fx^{t+1} - Fx^t$ , and  $B$  is the matrix with the rows  $v^n, v^1, \dots, v^{n-1}$ .

*Proof.* Let  $P$  be a permutation matrix such that

$$\Gamma_n P^t = (q^{p-n}, q^{p-1}, \dots, q^{p-n+1});$$

then

$$\Gamma_{p+1} = (q^p, \dots, q^{p-n+1}) - \Gamma_n P + (q^p - q^{p-n})(e^1)^T.$$

As a permutation matrix,  $P$  is nonsingular, and, hence, the same is true for  $\Gamma_n P$ . It then follows, by the Sherman-Morrison formula (2.3.14), that

$$\Gamma_{p+1}^{-1} = P^{-1} \Gamma_n^{-1} - (1/\alpha) P^{-1} \Gamma_n^{-1} (q^p - q^{p-n})(e^1)^T P^{-1} \Gamma_n^{-1}, \quad (29)$$

where  $\alpha = 1 + (e^1)^T P^{-1} \Gamma_n^{-1} (q^p - q^{p-n}) \neq 0$ , since, by hypothesis,  $\Gamma_{p+1}^{-1}$  exists. But the effect of multiplication on the left by  $P^{-1}$  is to place the last row in the position of the first row and move all other rows down one place. ■

Although the  $(n+1)$ -point sequential secant method requires the least amount of computation per step, it will be shown in Chapter 11 that the method is prone to unstable behavior and that no satisfactory convergence results can be given. In contrast, the two-point methods defined by (18) and (21) will be shown to retain the essential properties of Newton's method and, in particular, satisfactory local convergence theorems will be obtained for them in Section 11.2.

We end this section by describing a closely related class of iterative processes known as Steffensen methods.

Consider again the basic one-dimensional secant method (1). If we set  $h^k = f(x^k)$ , we obtain Steffensen's method in one dimension:

$$x^{k+1} = x^k - \frac{f(x^k)}{f(x^k + f(x^k)) - f(x^k)} f(x^k), \quad k = 0, 1, \dots \quad (30)$$

This iteration is of interest, since—under suitable conditions—it exhibits the same quadratic convergence as Newton's method while not requiring any derivatives of  $f$ . (See Chapter 11).

The concepts already developed for the secant method now permit immediate and natural extensions of (30) to  $n$  dimensions. In corre-

spondence with the two-point secant methods defined by the choice of points (23), we can define the analogous Steffensen method by (17) with the choice of auxiliary points

$$x^{k,j} = x^k + P_{j,k} Fx^k, \quad j = 1, \dots, n.$$

For example, if  $P_{j,k} = (0, \dots, 0, e^j, 0, \dots, 0)$ , we obtain the particular Steffensen method

$$x^{k+1} = x^k - J(x^k, Fx^k)^{-1} Fx^k, \quad (31)$$

where  $J$  is defined by (19); this is the direct analog of the two-point secant method (19)–(20). Similarly, corresponding to the secant method (20), (22), the choice  $P_{j,k} = (e^1, \dots, e^j, 0, \dots, 0)$  gives a Steffensen method of the form (31) in which  $J$  is defined by (22).

More generally, corresponding to (25), we can choose

$$x^{k,j} = x^k + \sum_{i=1}^k P_{i,j,k} Fx^{k-i+1}, \quad j = 1, \dots, n. \quad (32)$$

As a special case of (32), let

$$x^{k,j} = x^k + Fx^{k-j+1}, \quad j = 1, \dots, n,$$

which leads to the method

$$x^{k+1} = x^k - J(x^k, H_k)^{-1} Fx^k, \quad H_k = (Fx^k, \dots, Fx^{k-n+1}), \quad (33)$$

corresponding to the  $(n+1)$ -point secant method (24). Here, of course,  $J(x, H)$  is defined by (11).

Note that in the Steffensen method (31), (19) it is necessary to evaluate  $F$  at the points  $x^k + f_j(x^k) e^j$ ,  $j = 1, \dots, n$ , as well as at  $x^k$ , so that precisely the same number of function evaluations are needed as for the corresponding secant method defined by (19)–(20). On the other hand, in the case of (33), it is necessary to obtain  $F(x^k + Fx^{k-j})$ ,  $j = 1, \dots, n$ , so that, again,  $n+1$  function evaluations are required. Thus, the advantage of (24), in which only one new evaluation of  $F$  is needed at each stage, does not carry over to (33).

Another form of Steffensen's method arises in connection with the fixed-point equation  $x = Gx$ . Here, the auxiliary points  $x^{k,j}$  may be taken as the iterates  $x^{k,i} = G^i x^k$ ,  $i = 1, \dots, n$ , generated by the operator  $G$ . This then leads to the iteration

$$x^{k+1} = x^k - J(x^k, H_k)^{-1} [x^k - Gx^k], \quad H_k = (Gx^k - x^k, \dots, G^n x^k - x^k) \quad (34)$$

by setting  $Fx = x - Gx$  in (17). Note that the evaluation of  $J(x^k, H_k)$  involves the computation of the  $n$  vectors

$$F(x^i + H_k e^i) - Fx^i = F(G^i x^i) - Fx^i = G^i x^i - G^{i+1} x^i = x^i + Gx^i, \\ i = 1, \dots, n,$$

so that  $n+1$  evaluations of  $G$  are required. The iteration (34) is, of course, not restricted to equations in fixed-point form, since the conversion  $Gx = x - Fx$  may always be made. Note also that, in contrast to the multistep method (33), (34) is a one-step method.

### NOTES AND REMARKS

**NR 7.2-1.** For a thorough discussion of the one-dimensional secant method and of related higher-order methods, such as that of Muller, the reader is referred to Ostrowski [1966] and Traub [1964].

**NR 7.2-2.** The idea of replacing  $F$  by a linear interpolating function in order to extend the secant method to higher dimensions dates back to Gauss in the two-dimensional case (see Ostrowski [1966, Appendix D]). Its modern revival and generalization to  $n$  dimensions is apparently due to H. Heinrich in unpublished lectures (circa 1955), and its first rigorous analysis due to Bittner [1959]. However, Wolfe [1959] independently suggested the  $(n+1)$ -point method, described in the text, in which the vector with largest function value is dropped. More recent works on the  $(n+1)$ -point or related methods include Tornheim [1964], Anderson [1965] (see NR 7.2-10), and Barnes [1965] (see NR 7.3-2).

**NR 7.2-3.** The Steffensen iteration (34) was first considered by Ludwig [1952], and, more recently, by Henrici [1964], but from an entirely different point of view. One way of deriving the Steffensen iteration in one dimension is by means of the Aitken  $\mathcal{A}^2$ -process, which may be extended to  $n$  dimensions as follows: Given  $n+2$  points  $y^0, \dots, y^{n+1}$ , define the matrices

$$H = (y^1 - y^0, y^2 - y^1, \dots, y^n - y^{n-1})$$

and

$$S = (y^2 - 2y^1 + y^0, \dots, y^{n+1} - 2y^n + y^{n-1}),$$

and then introduce an "extrapolated" vector  $\hat{x}$  by  $\hat{x} = y^0 - HS^{-1}(H^* - y^0)$ . For the fixed-point equation  $x = Gx$ , set  $y^0 = x^k$ ,  $y^i = G^i x^k$ ,  $i = 1, \dots, n+1$ , and  $x^{k+1} = \hat{x}$ . Although slightly different in form, it is easy to see, by 7.2.7, that the  $x^{k+1}$  thus produced is precisely that of (34).

**NR 7.2-4.** The two-point secant and related Steffensen methods have been considered by a variety of authors. Koganoff [1961] handles the particular method in which the auxiliary points  $x^{k,i}$  are given by (18), while Maerigoiz [1967] and Wegge [1966] treat the corresponding Steffensen method (31), (19). Other authors have considered the method (22), (31) in the context of divided differences (NR 7.2-6).

**NR 7.2-5.** There is a fundamental difficulty in extending the interpolation approach to infinite dimensions, since, presumably, we would require that the linear interpolator  $L$  agree with  $F$  at infinitely many points. One possibility is to assume that  $\{x^i\}$  span the space in some sense and that  $Lx^i = Fx^i$ ,  $i = 0, 1, \dots$ . Then the basic secant approximation is defined as the solution of  $Lx = 0$ . This type of extension to infinite dimensions does not seem very promising, and has not been explored in the literature.

However, extensions of the two-point secant methods to infinite dimensions have been made by two related approaches. For one such approach, see NR 7.2-6; the other might be called "pointwise extension," and has been treated, for example, by Collatz [1964]. Consider the two-point boundary problem

$$u''(t) = f(u(t)), \quad u(0) = u(1) = 0, \quad t \in [0, 1].$$

"Newton's method" applied directly to this equation gives the sequence of linear boundary value problems

$$u_{k+1}'' = f(u_k) + (u_{k+1} - u_k) f'(u_k), \quad u_{k+1}(0) = u_{k+1}(1) = 0.$$

Now let  $h_k$  be a function on  $[0, 1]$  and define the corresponding discrete Newton method by

$$u_{k+1}'' = f(u_k) + (u_{k+1} - u_k) \frac{f(u_k + h_k) - f(u_k)}{h_k}.$$

Hence, for  $h_k = u_{k+1} - u_k$ , we have the two-point secant method:

$$u_{k+1}'' = f(u_k) + (u_{k+1} - u_k) \frac{f(u_k) - f(u_{k-1})}{u_k - u_{k-1}}.$$

Related Steffensen-type procedures, treated as extrapolation formulas, have been given by Bellman, Kagiwađa, and Kalaba [1965], and Noble [1964, p. 274].

**NR 7.2-6.** A more axiomatic approach to the extension of the secant method was taken by Schmidt [1961, 1963a] by means of the concept of a divided-difference operator. Briefly, Schmidt defines a *first divided difference* of  $F$  on a Banach space  $X$  as a mapping  $J: D \times D \subset X \times X \rightarrow L(X)$  which satisfies

$$J(x, h)h = F(x+h) - Fx, \quad \forall x, x+h \in D, \quad (35)$$