JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 23, No. 3, NOVEMBER 1977

A Modified Newton Method for Minimization[1]

R. FLETCHER[2] AND T. L. FREEMAN[3]

Communicated by A. V. Fiacco

Abstract. Some promising ideas for minimizing a nonlinear function, whose first and second derivatives are given, by a modified Newton method, were introduced by Fiacco and McCormick (Ref. 1). Unfortunately, in developing a method around these ideas, Fiacco and McCormick used a potentially unstable, or even impossible, matrix factorization. Using some recently developed techniques for factorizing an indefinite symmetric matrix, we are able to produce a method which is similar to Fiacco and McCormick's original method, but avoids the difficulties of the original method.

Key Words. Modified Newton methods, negative curvature directions, unconstrained minimization.

1. Introduction

In this paper, we consider the problem of minimizing a nonlinear function

$$f(x), \qquad f \in C^2, \qquad x \in E^n,$$

given that the first and second derivatives of the function (gradient vector and Hessian matrix) are explicitly available to the algorithm. The classical method for solving such a problem is Newton's method; this method is considered in Section 2. However, the method has its difficulties, particularly when the Hessian matrix is indefinite, remote from the minimum. A promising way of avoiding most of these difficulties is due to Fiacco and McCormick (Ref. 1), and the method has a particularly nice interpretation as one which generates directions of search which have negative curvature.

[1] Both authors gratefully acknowledge the award of a research fellowship from the British Science Research Council.

[2] Senior Research Fellow, Department of Mathematics, The University, Dundee, Scotland.
[3] Research Fellow, Department of Mathematics, The University, Dundee, Scotland.

However this method, which is described in Section 3, involves a potentially unstable LDL^T factorization when the Hessian matrix is indefinite. More recently, a stable and efficient factorization of an indefinite matrix has been suggested, based on the idea of a pivot which can be either a 1×1 or a 2×2 block diagonal element. This work is summarized in Section 4. This idea, together with that of generating a negative-curvature search direction, as in Fiacco and McCormick's method, is the basis of our new modification to Newton's method. This is described in Section 5. Section 6 describes the approach which we have used to solve the line search subproblem. There are some interesting considerations which relate to making this search efficient, especially when searching in a negative-curvature direction, when a good initial estimate of the step is not available. The paper concludes with some numerical results, and a discussion of the merits of our modification as against others which have been proposed. It is difficult to compare Newton methods on the basis of numerical experience, since most of them perform well, and the modifications only operate a small proportion of the time. Nonetheless, our modification is simple and does permit a nice interpretation in terms of generating a negative-curvature search direction.

The notation in use is to denote the gradient vector by

$$g(x) \equiv [\partial f(x)/\partial x_i]$$

and the Hessian matrix by

$$G(x) \equiv [\partial^2 f(x)/(\partial x_i \partial x_j)].$$

A local minimizer of $f(x)$ is denoted by x^*, and the ith approximation to x^* in the iterative method is denoted by $x^{(i)}$. $f^{(i)}$, $g^{(i)}$, and $G^{(i)}$ denote the function value, gradient vector, and Hessian matrix evaluated at $x^{(i)}$. The correction to $x^{(i)}$ is written as

$$\delta^{(i)} = x^{(i+1)} - x^{(i)},$$

and the direction of search as $p^{(i)}$.

2. Newton's Method

The classical method for minimizing a function when second-derivative information is available is Newton's method. It is an iterative method based on solving the set of equations

$$g(x) = 0,$$

which are the first-order necessary conditions for a function $f(x)$ to have a minimum. If the Taylor series

$$g(x^{(i)}+h)=g^{(i)}+G^{(i)}h+\cdots,$$

for $g(x)$ about any iterate $x^{(i)}$, is truncated after the linear term, then Newton's method consists of choosing a correction

$$h=\delta^{(i)}$$

to try to make

$$g^{(i+1)}=0.$$

This gives the iteration

$$\delta^{(i)}=-G^{(i)-1}g^{(i)}, \tag{1}$$

$$x^{(i+1)}=x^{(i)}+\delta^{(i)}. \tag{2}$$

From this derivation, it is clear that this method will converge to the stationary point of a quadratic function in one iteration. For general functions, Newton's method usually exhibits a quadratic rate of convergence near the solution: the precise details of what can be proved are well known and are stated, for example, in Ortega and Rheinboldt (Ref. 2). Unfortunately, this simple method may converge to a stationary point which is not a local minimum, since the condition

$$g(x)=0$$

is by no means sufficient. Even more important is the fact that the method is not reliable, remote from the minimum, even though no other stationary points exist. For instance, it is possible to find that

$$f^{(i+1)}>f^{(i)},$$

even though $G^{(i)}$ is positive definite, i.e., the quadratic approximation to f does have a minimum. In fact, when $G^{(i)}$ is positive definite, a direction

$$p^{(i)}=-G^{(i)-1}g^{(i)} \tag{3}$$

can be defined, where $p^{(i)}$ is a descent direction, namely, a direction such that

$$g^{(i)T}p^{(i)}<0;$$

hence, there is a positive step α in the direction $p^{(i)}$ such that f is reduced. It is, therefore, possible in an algorithm to choose a correction

$$\delta^{(i)}=\alpha^{(i)}p^{(i)}$$

which will ensure that the sequence $\{f^{(i)}\}$ is nonincreasing and which, therefore, improves the reliability of the method. $p^{(i)}$ is said to be a *direction* *of search.* It would seem attractive to choose $\alpha^{(i)}$ to minimize

$$f(x^{(i)}+\alpha p^{(i)})$$

(full line search), but this cannot be implemented computationally, except to some finite tolerance. It is also inefficient, and a usual compromise is to terminate the line search when

$$f^{(i+1)}<f^{(i)}.$$

However, this test alone by no means guarantees convergence. For instance, see the example

$$F(x)=(11/546)x^6-(38/364)x^4+(1/2)x^2,$$

from $x_0=1.01$ [Fletcher, Ref. 3, with a misprint corrected]. To eliminate these examples of nonconvergence, it is usual to impose some other test which is readily satisfied, for instance

$$|g^{(i+1)T}p^{(i)}|<-\beta g^{(i)T}p^{(i)}, \qquad 0<\beta<1. \tag{4}$$

However, there is as yet no general agreement on which tests are most appropriate. The resulting method may be termed Newton's method with a partial line search.

Even with the partial line search, Newton's method still has some drawbacks:

(i) Convergence to a saddle point is possible; such a point can be recognized by the occurrence of an indefinite Hessian matrix, but further progress is impossible because the gradient is zero.

(ii) When $G^{(i)}$ is singular, $p^{(i)}$ is not defined.

(iii) If $G^{(i)}$ is not positive definite, $p^{(i)}$ may not be a descent direction; for instance, $g^{(i)}$ and $p^{(i)}$ may be orthogonal, making further progress impossible.

These difficulties are due to the Hessian matrix not being positive definite, and various ideas exist for overcoming this problem. Several methods are based on the observation that

$$p^{(i)}=-A^{-1}g^{(i)}$$

is a descent direction when A is positive definite. The methods differ in the way in which they generate the positive definite matrix A, either as an arbitrary choice or by modifying the possibly indefinite matrix $G^{(i)}$. These methods will be considered again in Section 7. A different idea, which is also attractive, is due to Fiacco and McCormick (Ref. 1); we consider this method in some detail in the next section.

3. Fiacco and McCormick's Method

Fiacco and McCormick's method is a modification of Newton's method with full line search, in situations when $G^{(i)}$ has negative eigenvalues. It is based on the observation that, in this case, a direction of search $p^{(i)}$ exists which has negative curvature and along which $f(x)$ can be reduced by a line search. In view of the negative curvature along $p^{(i)}$, it is likely that this is a good way of reducing $f(x)$. Furthermore, since, if $f(x)$ is bounded below, the curvature along $p^{(i)}$ must change to positive for a sufficiently large step α, hopefully this will lead to a matrix $G^{(i+1)}$ which has one fewer negative eigenvalue than $G^{(i)}$ [although, of course, this cannot be guaranteed].

The modification is based on the use of a factorization

$$G^{(i)} = LDL^T, \tag{5}$$

where L is a unit lower triangular matrix ($l_{ii} = 1$) and D is a diagonal matrix. If $G^{(i)}$ (and, hence, D) is positive definite, then the factorization is used to generate a direction of search $p^{(i)}$ as defined by (3). If, however, there are any negative d_{ii}, then t is defined as

$$L^T t = a \tag{6}$$

where

$$a_i = \begin{cases} 1, & d_{ii} \leq 0, \\ 0, & d_{ii} > 0. \end{cases}$$

Then, it is easily verified that the direction

$$p^{(i)} = \begin{cases} t, & g^{(i)T} t \leq 0, \\ -t, & g^{(i)T} t > 0, \end{cases} \tag{7}$$

is a direction which satisfies

$$p^{(i)T} G^{(i)} p^{(i)} < 0, \qquad g^{(i)T} p^{(i)} \leq 0. \tag{8}$$

Hence, f can be decreased by searching along $+\alpha p^{(i)}$. Fiacco and McCormick recommend a full line search, although this step is better modified to use a partial line search. Their choice in regard to what to do when any

$$d_{ii} = 0$$

is not described here, since it is not entirely clear, and since it does not contribute to understanding the idea. In fact, there are one or two problems concerned with implementing this method, which are taken up in the context of the new method in Section 5.

The main disadvantage of the idea is that the factorization (5) is potentially unstable and may not even exist. However, recent work in linear algebra has devised a new method for factorizing a symmetric indefinite matrix, which does not suffer in this way and which can be used as the basis of an algorithm. This is described in the next section.

4. Stable Factorization of Symmetric Indefinite Matrices

It is well known that the LDL^T factorization of a symmetric positive-definite matrix A is both stable and efficient, but that it is potentially unstable when A becomes indefinite. In fact, for some A, for example,

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

the factorization does not exist.

The problem of factorizing symmetric indefinite matrices is considered by Bunch and Parlett (Ref. 4). They describe a factorization in which either one *or two* rows and columns of the partially factorized matrix may be eliminated. For example, if we let the $n \times n$ matrix $A^{(0)}$ be

$$A^{(0)} = \begin{bmatrix} a_{11} & \underset{\sim}{a}_{21}^T \\ \underset{\sim}{a}_{21} & A_{22} \end{bmatrix},$$

where A_{22} is $(n-1)\times(n-1)$, then elimination of one row and column would lead to the reduced matrix $A^{(1)}$:

$$A^{(1)} = A^{(0)} - d_1 \underset{\sim}{l}_1 \underset{\sim}{l}_1^T,$$

where

$$d_1 = a_{11}, \qquad \underset{\sim}{l}_1 = (1/d_1)\begin{bmatrix} a_{11} \\ \underset{\sim}{a}_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ \underset{\sim}{a}_{21}/a_{11} \end{bmatrix},$$

and $A^{(1)}$ has zero first row and column. If, however, we let $A^{(0)}$ be

$$A^{(0)} = \begin{bmatrix} A_{11} & A_{21}^T \\ A_{21} & A_{22} \end{bmatrix},$$

where A_{11} is 2×2, then elimination of two rows and columns would lead to the reduced matrix $A^{(2)}$:

$$A^{(2)} = A^{(0)} - L_1 D_1 L_1^T,$$

where

$$D_1 = A_{11}, \qquad L_1 = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} D_1^{-1} = \begin{bmatrix} I \\ A_{21}A_{11}^{-1} \end{bmatrix},$$

and $A^{(2)}$ has zero first and second rows and columns.

The algorithm chooses between these two alternative types of pivot, not only on the first iteration, but also on subsequent iterations. It is possible to resolve this choice in such a way that roundoff errors are controlled as well as possible. This must be done in conjunction with symmetric row and column interchanges, in which case a stable factorization is possible. The factorization requires

$$n^3/6+O(n^2)$$

multiplications to compute, analogous to (5), and is equally efficient in its storage requirement. The resulting factorization (of P^TAP, where P is the permutation matrix) has the structure that D is a symmetric block diagonal matrix, with 1×1 or 2×2 blocks, and L is a unit lower triangular matrix. Furthermore, the elements of L which correspond to nonzero off-diagonal elements of D are also zero. An example is given by the matrices

$$D=\begin{bmatrix} x & & & & & \\ & x & x & & & \\ & x & x & & & \\ & & & x & x & \\ & & & x & x & \\ & & & & & x \end{bmatrix},\qquad L=\begin{bmatrix} 1 & & & & & \\ x & 1 & & & & \\ x & 0 & 1 & & & \\ x & x & x & 1 & & \\ x & x & x & 0 & 1 & \\ x & x & x & x & x & 1 \end{bmatrix}$$

Bunch and Parlett (Ref. 4) describe a pivot strategy which requires $O(n^3)$ comparisons, but a more efficient $O(n^2)$ strategy, based on the concept of an error growth factor, is described by Fletcher (Ref. 5). This pivotal strategy also guarantees that each 2×2 block of D has one positive and one negative eigenvalue. Our modification of Newton's method is based directly on this type of factorization.

5. New Modified Newton's Method

As in Fiacco and McCormick's method, the modification that we suggest operates when $G^{(i)}$ is indefinite. $G^{(i)}$ is factorized as

$$G^{(i)}=LDL^T, \tag{9}$$

where L and D are as described in Section 4, and where we neglect any interchanges for simplicity of exposition. If $G^{(i)}$ is positive definite, then the factorization is used to solve (3) in accordance with Newton's method. If, however, $G^{(i)}$ has negative eigenvalues, then again we look for a negative-curvature direction $p^{(i)}$ which satisfies conditions (8), so that $f(x)$ can be reduced by searching along $+\alpha p^{(i)}$. This is given by defining

$$L^Tt=a, \tag{10}$$

where, for i corresponding to a 1×1 pivot,

$$a_i = \begin{cases} 1, & d_{ii} \leq 0, \\ 0, & d_{ii} > 0; \end{cases}$$

and, for i, $i+1$ corresponding to a 2×2 pivot,

$$\begin{bmatrix} a_i \\ a_{i+1} \end{bmatrix}$$

is the eigenvector corresponding to the negative eigenvalue of

$$\begin{bmatrix} d_{i,i} & d_{i+1,i} \\ d_{i+1,i} & d_{i+1,i+1} \end{bmatrix},$$

normalized so that

$$a_i^2 + a_{i+1}^2 = 1.$$

Then, we again let

$$p^{(i)} = \begin{cases} t, & g^{(i)T} t \leq 0, \\ -t, & g^{(i)T} t > 0. \end{cases} \tag{11}$$

It can easily be verified that $p^{(i)}$ satisfies the required conditions (8). The first condition holds, since

$$p^{(i)T} G^{(i)} p^{(i)} = p^{(i)T} LDL^T p^{(i)} = a^T D a = \sum_{i:\lambda_i<0} \lambda_i < 0,$$

where λ_i are the eigenvalues of D. The second condition follows directly from (11).

The only difficult decisions concerned with implementing this method concern what to do when $G^{(i)}$ has some zero eigenvalues, and also how to carry out the line search, especially when $G^{(i)}$ is indefinite. When $G^{(i)}$ has at least one zero eigenvalue, then a direction $p^{(i)}$ which satisfies either

$$G^{(i)} p^{(i)} = 0 \quad \text{and} \quad g^{(i)T} p^{(i)} < 0, \tag{12}$$

or

$$p^{(i)} = -L^{-T} \tilde{D}^{+} L^{-1} g^{(i)}, \tag{13}$$

where $\tilde{D}$ is D with any negative eigenvalues made zero. The expressions (12) represent an underdetermined system of equations, and there is some freedom in the way $p^{(i)}$ is defined. It should be noted that (12) generates a descent direction of zero curvature, while (13) generates a Newton direction restricted to the subspace of directions of positive curvature. If $p^{(i)}$ is generated by (12) on one iteration, then, on the next iteration, if there is a zero eigenvalue of $G^{(i+1)}$, $p^{(i+1)}$ is generated by (13) and vice versa. Were

(12) or (13) used continually, then there is a danger that all the directions of search would lie in the same subspace of E^n. In fact, there is a similar danger of nonconvergence if searches are made continually in directions of negative curvature. In the implementation described in Section 7, the direction of search is chosen alternately to be firstly a direction of negative curvature [see (11)] and then a Newton direction restricted to the subspace of positive curvature directions [see (13)], while the Hessian matrix has negative eigenvalues. This is not the only possible way of resolving this difficulty, but was found to be the best of those which were tried.

The other main feature of the new method, which has not so far been described in detail, is that of how the line search is implemented. This is the subject of the next section.

6. Line Search

At the point $x^{(i)}$, the method described in the previous section provides a direction of search $p^{(i)}$; the problem considered in this section is to find a value $\alpha^{(i)}$, such that, at

$$x^{(i+1)}=x^{(i)}+\alpha^{(i)}p^{(i)},$$

the conditions

$$f^{(i+1)}<f^{(i)}, \qquad |g^{(i+1)T}p^{(i)}|<-\beta g^{(i)T}p^{(i)}, \qquad 0<\beta<1, \tag{14}$$

are satisfied. We note that $f^{(i)}$, $g^{(i)}$, and $G^{(i)}$ are all known, and that the calculation of $\alpha^{(i)}$ is a one-dimensional minimization problem in terms of a single variable α (see Fig. 1).

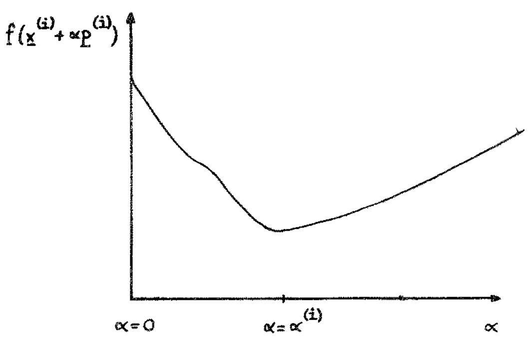

Fig. 1. One-dimensional minimization problem.

An initial approximation α_1 to $\alpha^{(i)}$ is required; it is possible to fit a quadratic function to the available information f_0, g_0, and G_0, and take the position of the minimum of this function as α_1. Here, f_0, g_0, and G_0 are defined as

$$f^{(i)}, \quad g^{(i)T}p^{(i)}, \qquad \text{and} \quad p^{(i)T}G^{(i)}p^{(i)},$$

respectively. This approach is appropriate when G_0 is positive, that is, the curvature in the direction $p^{(i)}$ is positive. However, when G_0 is negative, the quadratic function has a maximum, but no minimum; and, when G_0 is zero, the quadratic function reduces to a linear function, which has no stationary point. To overcome these difficulties, some extra information is required so that a higher-order polynomial which does have a minimum may be fitted.

A similar problem occurs with quasi-Newton methods, particularly on early iterations when the only reliable information is f_0 and g_0; this problem has been avoided by using the concept of a *target function reduction* $T^{(i)}$. This is the function reduction which one expects to achieve on the ith iteration. This allows a different quadratic function to be fitted, which gives a different prediction (α_1^*, say) of the line minimizer (see Fig. 2).

Since too large a correction $\delta^{(i)}$ must be avoided when the second-derivative information is unreliable, and since the Newton correction is required when reliable information is built up, so α_1, the initial approximation to $\alpha^{(i)}$, is taken to be

$$\min\{\alpha_1^*, 1\}.$$

The initial target $T^{(0)}$ is user-supplied and $T^{(i)}$, $i \geq 1$, is updated as the function reduction achieved on the $(i-1)$th iteration. This approach has proved to be effective, ensuring that α_1 is of the correct magnitude. It also has the advantage that, close to the minimum, where f is well approximated

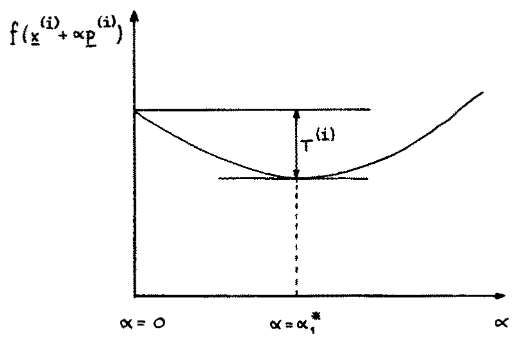

Fig. 2. Use of the target function reduction.

by a quadratic function, and where the direction generated by a quasi-Newton method is a good approximation to the true Newton direction, the value

$$\alpha_1 = 1$$

is always chosen.

The same approach can be used here to provide the extra information required when searching along a direction of negative or zero curvature. In the case of negative curvature, it has been found desirable, first of all, to predict the position of the maximum in the direction $-p^{(i)}$ by quadratic extrapolation. A quartic function, which has a maximum at this predicted point and which achieves a function reduction $T^{(i)}$, is then fitted to the available information. This quartic function is used to give an initial approximation α_1 to $\alpha^{(i)}$. It should be noted that Ineq. (14-2) is not the most appropriate test when $p^{(i)}$ is a direction of negative curvature. A more appropriate test is

$$|g^{(i+1)T}p^{(i)}| < -\beta g_{\max},$$

where $g_{\max}$ is the component of g in the direction $p^{(i)}$ at the point of inflection; the quartic function is also used to estimate $g_{\max}$. In the case of zero curvature, a cubic function is fitted to the available data, and extrapolation is performed to give α_1. In the cases of negative and zero curvature, it is not appropriate to let the initial approximation to $\alpha^{(i)}$ be

$$\min\{\alpha_1, 1\},$$

since a unit step does not have the same significance that it has in the case when $p^{(i)}$ is a Newton direction; hence, the initial approximation to $\alpha^{(i)}$ is simply taken to be α_1. It has been found best to update $T^{(i)}$ as the function reduction obtained on the most recent Newton step, since the function reduction achieved on Newton and non-Newton steps has been found to differ substantially in magnitude.

Having generated an initial approximation α_1 to $\alpha^{(i)}$, further approximations to $\alpha^{(i)}$ can be generated easily by interpolation or extrapolation until the convergence criterion (14) is satisfied. Upper and lower limits are imposed on the interpolants to prevent nonconvergence (e.g., convergence to one of the endpoints of the range of interpolation). Extrapolation is simply performed by increasing the step by a fixed multiplier, although it is possible to use polynomial extrapolation.

7. Numerical Results and Discussion

The method described in the previous sections has been implemented in FORTRAN on an Elliot 4130 computer at Dundee University. The algorithm was terminated when the absolute value of each component of the direction of search $|(p^{(i)})_j|$ was less than a specified tolerance TOL. For all but one test function, TOL was taken to be 10^{-6}. For Powell's singular function, TOL was set to 10^{-12} in order to test the program's ability to deal with a singular Hessian matrix; three searches in directions of zero curvature were performed; the program converged and output a warning that the Hessian matrix was singular at the solution.

For all the test problems, the results of Table 1 include the number of iterations and function evaluations required to achieve convergence and the final function value. The number of non-Newton steps taken by the method is included, since this is an indication of the difficulty which the test problem poses.

Also tabulated are the average number of eigenvalues which change sign from negative to positive on each search in a negative-curvature direction; it should be noted that, with the exception of the trigonometric functions with $N = 4$ and $N = 8$, on average at least one negative eigenvalue is eliminated by each search in a negative-curvature direction. This numerical evidence supports our belief that following a direction of negative curvature leads to a Hessian matrix with less negative eigenvalues. It should be noted that the number of negative eigenvalues of $G^{(i)}$ is simply equal to

Table 1. Numerical results.

Function	Number of variables	Number of iterations	Number of function evaluations	Number of Non-Newton steps	Average number of eigenvalues which become positive in each negative-curvature search	Final function value
Wood's	4	25	67	2	1	1.14×10^{-19}
Trigonometric	2	2	5	0	—	2.15×10^{-9}
Trigonometric	4	8	19	2	$\frac{1}{2}$	8.17×10^{-16}
Trigonometric	6	8	15	1	1	3.90×10^{-15}
Trigonometric	8	9	17	1	0	2.26×10^{-12}
Trigonometric	10	12	28	1	1	1.23×10^{-9}
Trigonometric	20	13	31	2	$\frac{3}{2}$	3.41×10^{-12}
Trigonometric	30	16	40	4	1	1.63×10^{-12}
Powell's	4	37	72	3	—	7.04×10^{-26}

the number of negative 1×1 blocks of D plus the number of 2×2 blocks of D. The results for the three test problems tabulated are certainly comparable with the results given by the method's competitors. However, comparing modified Newton methods on numerical evidence alone is not very reliable. It is possible for a good method to converge to a region in which the Hessian matrix is indefinite and spend time moving away from this region, while a worse method may completely avoid the difficult region, more by good fortune than anything else. For example, with Wood's function, taking a slightly different path for the first few iterations can cause a substantial difference in the number of iterations and function evaluations required for convergence. Because of the unreliability of comparisons based solely on numerical results, we present in the next paragraphs a more qualitative comparison of the new method with its competitors.

Of the currently available modified Newton methods, several can be eliminated from the comparison because they suffer from major drawbacks. The methods of Goldstein and Price (Ref. 6) and Dixon and Biggs (Ref. 7) fail to use useful information in the indefinite Hessian matrix by allowing the direction of search to be the steepest-descent direction, $-g^{(i)}$, when $G^{(i)}$ is indefinite or singular; the method proposed by Mathews and Davies (Ref. 8) involves a potentially unstable factorization, while Greenstadt's (Ref. 9) method requires an eigenvector analysis of the Hessian matrix on each iteration, and this is very time-consuming. This leaves a few alternatives to our new method, and these are now considered.

Gill and Murray (Ref. 10) propose a method based on the LDL^T factorization of $G^{(i)}$. During the factorization process, the elements of D are modified so that the factorization is stable and so that each element of D is positive. The result is the LDL^T factorization of a positive-definite matrix A,

$$A = G^{(i)} + E^{(i)},$$

where $E^{(i)}$ is a diagonal matrix which is zero when $G^{(i)}$ is sufficiently positive definite; the direction of search $p^{(i)}$ is then defined as $-A^{-1}g^{(i)}$. While this approach will always generate a downhill search direction, it is our belief that our new method, by isolating a negative-curvature direction, is more likely to leave a difficult region, in which the Hessian matrix is indefinite, in fewer iterations. In fact, Gill and Murray find it necessary to use the idea of a negative-curvature search direction in order to move away from a saddle point.

Another family of modified Newton methods is based on the concept of restricting the step $\delta^{(i)}$ taken on the ith iteration. These methods involve minimizing $Q^{(i)}(x)$, subject to

$$\|\delta^{(i)}\| \leq h,$$

where $Q^{(i)}(x)$ is a quadratic approximation to $f(x)$ at $x^{(i)}$, and h is such that $Q^{(i)}(x)$ is a good approximation to $f(x)$ for all x such that

$$\|x^{(i)}-x\|\leq h.$$

It is, of course, desirable to make h as large as possible, so as not to overrestrict $\delta^{(i)}$, while ensuring that $Q^{(i)}(x)$ is a good approximation to $f(x)$. The method of hypercubes, described by Fletcher (Ref. 3), chooses the norm to be the ∞-norm and involves solving a simple quadratic programming subproblem on each iteration. The method of Goldfeld, Quandt, and Trotter (Ref. 11) chooses the norm to be the 2-norm. This method requires the solution of a set of linear equations

$$(G^{(i)}+\mu I)\delta^{(i)}=-g^{(i)},$$

for possibly more than one value of μ on each iteration, where μ is the Lagrange multiplier of constraint

$$\|\delta^{(i)}\|_2\leq h.$$

Goldfeld, Quandt, and Trotter base this calculation on an eigenvector analysis of $G^{(i)}$; however, Hebden (Ref. 12) has suggested a method based on the same ideas, but which does not require the eigenvector analysis. Again, the advantage that our new method has over these methods is that it is more likely to leave a difficult region in fewer iterations, especially as the restricted step methods are likely to be restricted to small steps in such regions.

8. Appendix: Test Functions

(i) Wood's function:

$$f=100(x_2-x_1^2)^2+(1-x_1)^2+90(x_4-x_3^2)^2+(1-x_3)^2$$
$$+10.1((x_2-1)^2+(x_4-1)^2)+19.8(x_2-1)(x_4-1).$$

This function has a minimum $f=0$ at

$$x^*=(1,1,1,1)^T.$$

The starting point $x^{(1)}$ is taken to be

$$(-3,-1,-3,-1)^T.$$

(ii) Trigonometric functions:

$$f=\sum_{i=1}^{N}\left\{E_i-\sum_{j=1}^{N}(A_{ij}\sin x_j+B_{ij}\cos x_j)\right\}^2$$

where A_{ij}, B_{ij} are random integers in the range $[-100, 100]$ and

$$E_i = \sum_{j=1}^{N} \{A_{ij} \sin x_j^* + B_{ij} \cos x_j^*\}, \qquad i = 1, 2, \ldots, N,$$

where x_j^* are random numbers in the range $[-\pi, \pi]$. This function has a minimum $f = 0$ at

$$x^* = (x_1^*, x_2^*, \ldots, x_N^*)^T.$$

The starting point $x^{(1)}$ is taken to be

$$x^* + 0.1\delta,$$

where δ is a vector whose components are random numbers in the range $[-\pi, \pi]$.

(iii) Powell's singular function:

$$f = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4.$$

This function has a minimum $f = 0$ at

$$x^* = (0, 0, 0, 0)^T.$$

The starting point $x^{(1)}$ is taken to be

$$(3, -1, 0, 1)^T.$$

References

1. FIACCO, A. V., and MCCORMICK, G. P., *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley & Sons, New York, New York, 1968.
2. ORTEGA, J. M., and RHEINBOLDT, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, New York, 1970.
3. FLETCHER, R., *An algorithm for Solving Linearly Constrained Optimization Problems*, Mathematical Programming, Vol. 2, pp. 133–165, 1972.
4. BUNCH, J. R., and PARLETT, B. N., *Direct Methods for Solving Symmetric Indefinite Systems of Linear Equations*, SIAM Journal on Numerical Analysis, Vol. 8, pp. 639–655, 1971.
5. FLETCHER, R., *Factorizing Symmetric Indefinite Matrices*, Linear Algebra and its Applications, Vol. 14, pp. 257–262, 1976.
6. GOLDSTEIN, A. A., and PRICE, J. F., *An Effective Algorithm for Minimization*, Numerische Mathematik, Vol. 10, pp. 184–189, 1967.
7. DIXON, L. C. W., and BIGGS, M. C., *Meander, a Newton Based Procedure for N-Dimensional Function Minimization*, Hatfield Polytechnic, Numerical Optimization Center, Technical Report No. 9, 1970.

8. MATHEWS, A., and DAVIES, D., *A Comparison of Modified Newton Methods for Unconstrained Optimization*, Computer Journal, Vol. 14, pp. 293–294, 1971.
9. GREENSTADT, J., *On the Relative Efficiencies of Gradient Methods*, Mathematics of Computation, Vol. 21, pp. 360–367, 1967.
10. GILL, P. E., and MURRAY, W., *Two Methods for the Solution of Linearly Constrained and Unconstrained Optimization Problems*, National Physical Laboratory, Teddington, England, Report No. NAC 25, 1972.
11. GOLDFELD, S. M., QUANDT, R. E., and TROTTER, H. F., *Maximization by Quadratic Hill-Climbing*, Econometrica, Vol. 34, pp. 541–551, 1966.
12. HEBDEN, M. D., *An Algorithm for Minimization Using Exact Second Derivatives*, Atomic Energy Research Establishment, Harwell, England, Report No. TP 515, 1973.