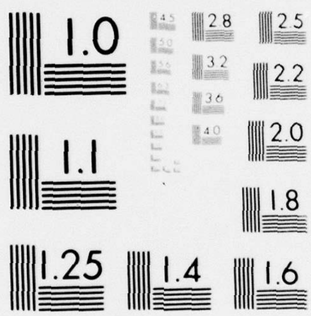
AD-A037 820    GEORGE WASHINGTON UNIV WASHINGTON D C   INST FOR MANAG--ETC   F/G 12/1
STRATEGIES FOR THE MINIMIZATION OF AN UNCONSTRAINED NONCONVEX F--ETC(U)
NOV 76   G P MCCORMICK                                    AF-AFOSR-2504-73

UNCLASSIFIED    SERIAL-T-343    AFOSR-TR-77-0171    NL

1 OF 1
ADA037820

END
DATE
FILMED
4-77




MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFOSR - TR - 77 - 017 F

(3)

ADA037820

THE GEORGE WASHINGTON UNIVERSITY

STUDENTS FACULTY STUDY R
ESEARCH DEVELOPMENT FUT
URE CAREER CREATIVITY CO
MMUNITY LEADERSHIP TECH
NOLOGY FRONTI
ENGINEERING APP
GEORGE WASHIN

DDC
APR 4 1977
C

Approved for public release;
distribution unlimited.

INSTITUTE FOR MANAGEMENT
SCIENCE AND ENGINEERING
SCHOOL OF ENGINEERING
AND APPLIED SCIENCE

THIS DOCUMENT HAS BEEN APPROVED FOR PUBLIC RELEASE AND SALE; ITS DISTRIBUTION IS UNLIMITED

AD NO.
DDC FILE COPY

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER AFOSR - TR - 77 - 0171 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle) STRATEGIES FOR THE MINIMIZATION OF AN UNCONSTRAINED NONCONVEX FUNCTION | | 5. TYPE OF REPORT & PERIOD COVERED Interim |
| | | 6. PERFORMING ORG. REPORT NUMBER T-343 |
| 7. AUTHOR(s) Garth P. McCormick | | 8. CONTRACT OR GRANT NUMBER(s) AFOSR 73-2504 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS The George Washington University Operations Research Department Washington, DC 20052 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A6 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bldg 410 Bolling AFB, Washington, DC 20332 | | 12. REPORT DATE 1 November 1976 |
| | | 13. NUMBER OF PAGES 21 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report) UNCLASSIFIED |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

406 743

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

unconstrained optimization
nonconvex optimization
Newton's method

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Several strategies are discussed which modify the classical form of Newton's method for minimizing an unconstrained function of several variables when at some point the Hessian matrix is not positive definite. A simple example is solved to explain the different methods.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# STRATEGIES FOR THE MINIMIZATION OF AN UNCONSTRAINED NONCONVEX FUNCTION

by

Garth P. McCormick

Serial T-343
1 November 1976

The George Washington University
School of Engineering and Applied Science
Institute for Management Science and Engineering

ACCESSION for
NTIS White Section
Buff Section

Research Sponsored by
Air Force Office of Scientific Research
Air Force Systems Command, USAF
under AFOSR Contract 73-2504

This document has been approved for public
sale and release; its distribution is unlimited.

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

Abstract
of
Serial T-343
1 November 1976

# STRATEGIES FOR THE MINIMIZATION OF AN UNCONSTRAINED NONCONVEX FUNCTION

by

Garth P. McCormick

Several strategies are discussed which modify the classical form of Newton's method for minimizing an unconstrained function of several variables when at some point the Hessian matrix is not positive definite. A simple example is solved to explain the different methods.

Research Sponsored by
Air Force Office of Scientific Research

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

# STRATEGIES FOR THE MINIMIZATION OF AN UNCONSTRAINED NONCONVEX FUNCTION

by

Garth P. McCormick

## 1. Introduction

Newton's method for finding a zero of a function of one variable is the prototype algorithm. There is a large body of literature dealing with Newton's method for solving n simultaneous equations in n unknowns (see [17] for material relating to this). Since finding the unconstrained minimizer of a function of a twice differentiable function of several variables involves finding a point where the first partial derivatives vanish, Newton's method is applied toward solving this problem also. In this paper Newton's method, with computationally important variations for the case when the Hessian matrix of the function is occasionally indefinite, is analyzed.

In Section 2 Newton's method is derived from a natural point of view using the "gradient path" approach. There is an interesting connection between the classical Cauchy [3] method of steepest descent and Newton's method. This point of view is helpful in Section 3, where modifications to the basic approach are discussed in order to obtain convergence to a second order point (one satisfying the requirement that the Hessian matrix be positive semi-definite, as well as one where the gradient vector vanishes). The different strategies for doing this involve directions of negative curvature. The strategies are compared in a simple example.

## 2. A Cauchy-Newton Approach to Unconstrained Minimization

A natural way to develop a method for minimizing an unconstrained function $f(x)$ is to consider a physical situation. The trajectory of a boulder down the side of a mountain (with the boulder restrained by ropes) would approximately satisfy the differential equation

$$\dot{x}(t) = -\nabla f[x(t)] \ . \tag{1}$$

In general it is not possible to obtain a solution in closed form of (1). When $f(x)$ is a quadratic form there is a general solution, since the differential equations are linear. Let $x(0) = x_0$ be the initial point, and let $E\lambda E^T$ be an eigenvector-eigenvalue decomposition of $\nabla^2 f$ , i.e., $\lambda$ is the diagonal matrix of eigenvalues, and $EE^T = I$ with $E$ the matrix of eigenvectors. Since $f$ is assumed quadratic, this decomposition is independent of $x$ .

It is well known (see [4]) that the solution of (1) is

$$x(t) = x_0 - E\gamma(t)E^T\nabla f(x_0) \ , \tag{2}$$

where $\gamma(t)$ is a diagonal matrix whose jth diagonal element $\gamma_j(t)$ is

$$\gamma_j(t) = \begin{Bmatrix} \left(1 - e^{-\lambda_j t}\right)/\lambda_j \ , & \text{if } \lambda_j \neq 0 \\ t \ , & \text{if } \lambda_j = 0 \end{Bmatrix} .$$

For $t$ small, (2) yields

$$x(t) \doteq x_0 - \nabla f(x_0)t \ .$$

This is similar to the algorithm known as Cauchy's method of steepest descent [3]. One version of this algorithm is to generate a sequence of minimizing points as

$$x_{k+1} = x_k - \nabla f(x_k)t_k \ , \quad k=0,1,\ldots \ , \tag{3}$$

where each $t_k$ is obtained from the following step size problem (with $s_k = -\nabla f(x_k))$ .

SSP I (Step Size Problem I). At a point $x_k$ , given a direction of search $s_k$ , set $x_{k+1} = x_k + s_k t_k$ , where $t_k$ is a local solution to

$$\underset{t \geq 0}{\text{minimize}} \; f(x_k + s_k t) \; .$$

An important distinction should be made between the continuous form of steepest descent given by (1) and the discrete form in (3). The former is suitable for implementation on an analogue computer. There are not many published results on this. One such experiment (Fiacco and McCormick [6], Section 7.3) demonstrated the feasibility of this approach. Implementation of this, however, requires an extraordinary amount of equipment and time. The discrete approach is easily implemented on a digital computer but is notoriously slow. Another important point is that the solution of the first order continuous differential equations seems to imply that the appropriate algorithm for implementation on a digital computer is not the discrete form of steepest descent, but rather some form of Newton's method. With this in mind, analysis of (1) and (2) is continued.

When $t$ is large the trajectory $x(t)$ depends upon the signs and magnitudes of the eigenvalues. In the case when $\nabla^2 f$ is a positive definite matrix, i.e., when $\lambda_j > 0$ for all $j$ ,

$$x(\infty) = x_0 - (\nabla^2 f)^{-1} \nabla f(x_0) \; .$$

This is the algorithm prescribed by the classical version of Newton's method which minimizes a positive definite quadratic form in one iteration.

The classical method without modifications iterates as

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k) \; , \quad \text{for } k=0,1,\ldots \; . \qquad (4)$$

Near a point $x^*$ where $\nabla f(x^*) = \underset{\sim}{0}$ and $\nabla^2 f(x^*)$ is positive definite, the rate of convergence is "at least quadratic," i.e., there is a value $M > 0$ such that

$$||x_{k+1} - x^*|| \leq M\, ||x_k - x^*||^2 .$$

3. Global Convergence Using Directions of Negative Curvature

There are many theoretical and computational objections to the use of Newton's method. The computational objections will not be taken up in this paper; here the theoretical difficulties are discussed and the basic method modified so that convergence to a second order point can be obtained.

When $x_k$ is far away from an isolated local minimizer the Hessian matrix may be indefinite, and occasionally may even be singular. In this instance the traditional move may not be a descent direction (and in the singular case is not defined). Furthermore, even if the Hessian is positive definite, the quadratic approximation may be so poor that little progress is made. Since the computation of the Newton direction is relatively expensive computationally and since it is not known how close the current point is to the isolated local minimizer, it is argued that a simpler algorithm is to be preferred. Another problem (which may be more computational than theoretical) is that the condition number of the Hessian may be so high that the numerical procedure for computing the Newton direction gives a false indication that the Hessian is not positive definite. There is no way, using the traditional Newton equation, that this possibility can be adequately handled.

These objections have led investigators recently to propose modifications to the basic algorithm. Indeed, any computer program which implements a form of Newton's method must take into account these difficulties. The problem of what form Newton's method should take when at the point $x_k$, when the Hessian matrix $\nabla^2 f(x_k)$ is not positive definite, has been investigated by several people. The strategies covered by these papers fall into five general categories.

A. When in the course of computing the inverse Hessian (usually in some implicit form) an indication occurs that the Hessian is not positive definite, force the numerical procedure to generate a positive definite matrix. The reasoning behind this strategy is that in most cases when this occurs, it is because of numerical round off errors (caused sometimes by ill-conditioning) and that this will tend to correct the round off problem. In any event, it is argued, the resulting direction will be one of descent.

This strategy will not be pursued further here, except to note that compared to those discussed below it is wasteful of information. The same numerical procedure that is used to get the inverse Hessian in the positive definite case should be able to compute information that will hasten the search for the minimizer. There is no reason why the descent direction above will be any better than, say, the steepest descent direction when the Hessian is indefinite. In other words, if the generation of a direction of descent is the only concern, there are cheaper ways to do it. For more information on these techniques the reader is referred to Matthews and Davies [13], Greenstadt [10], Levenberg [11], and Marquardt [12].

B. When it is discovered that the Hessian matrix is not positive definite, modify the numerical procedure and compute $d_k$, a descent direction of negative curvature, i.e., a vector such that

$$d_k^T \nabla^2 f(x_k) d_k \leq 0 ,$$

and

$$d_k^T \nabla f(x_k) \leq 0 .$$

Set $s_k = d_k$ and find $t_k$ the step size scalar by using SSP I.

The motivation behind this strategy is to hasten the search for a region in which the Hessian matrix of $f$ is positive definite so that the classical approach will apply and an ultimate quadratic rate of convergence be obtained. The reason this strategy should accomplish that is that the direction $d_k$ is one in which the function decreases, and also one (at least initially) in which the rate of decrease is decreasing. To see this, simply note that

$$df(x_k+d_k t)/dt = d_k^T \nabla f(x_k) \leq 0 , \quad \text{at} \quad t = 0 ,$$

and

$$d^2 f(x_k+d_k t)/dt^2 = d_k^T \nabla^2 f(x_k) d_k \leq 0 , \quad \text{at} \quad t = 0 .$$

If the problem is well posed, the function is bounded below and $\{x|f(x) \leq f(x_k)\}$ is a bounded set. Thus, ultimately the curvature of $f$ in the direction $d_k$ emanating from $x_k$ will become nonnegative.

If the directions $\{d_k\}$ are chosen with care, eventually the sequence of minimizing points will enter a region in which the Hessian is positive (semi) definite and remain there. It is not difficult to compute a direction of negative curvature; what is difficult is to compute one which has some resemblance to an eigenvector of the Hessian associated with its minimum eigenvalue. When it does, the minimum eigenvalue of the Hessian can be expected to increase each time the above strategy is used. Eventually the minimum eigenvalue is brought above zero, as hoped for.

Intuitively this strategy makes sense. Theorems can be used to prove second order convergence of the strategy if the directions $\{d_k\}$ have certain properties. If one is willing to go to the trouble to compute $e_k^{min}$, convergence (except under pathological circumstances) can be established. The problem is how to obtain a good direction of negative curvature and use no more arithmetic operations than would be required to compute the usual Newton vector. This matter has been discussed elsewhere [15]. For recent attempts to handle this problem the reader is referred to Gill and Murray [8], Fletcher and Freeman [7], Fiacco and McCormick [6], and the survey by Murray [16].

C. The most appealing strategy is based on (2). If this were done, the continuous steepest descent trajectory, or "gradient path" (based on a quadratic approximation at the point $x_k$), would be approximated. This has many desirable features. For one thing, the method is not as sensitive to numerical round off errors which may give a false indication of indefiniteness. It is easy to show that

$$\lim_{\lambda_j \to 0} \left(1 - e^{-\lambda_j t}\right)/\lambda_j = t \;, \quad \text{for all} \quad t > 0 \;.$$

There have been some experiments based on this approach (see [18], [9], [5], and in particular [2]). The major drawback is that it requires a full eigenvalue-eigenvector decompositon; that is, unless techniques for exponentiating a matrix are used.

An obvious modification to this would be to do another decomposition which resembles the eig-eig decomposition using some numerically stable procedure. There are many possibilities for this, but no published work seems to have been done in this area. There is one approach which used a quasi-Newton updating technique to approximate the inverse Hessian [19]. Some numerical experience has been reported there.

The formal statement of this approach is: let $x_k(t)$ be the solution given by (2) to the quadratic approximation problem. Set $x_{k+1} = x_k(t_k)$ where $t_k$ is a local solution to $\min_{t \geq 0} f[x_k(t)]$ .

D. The fourth strategy is to create a trajectory which is a combination of a descent trajectory and a trajectory given by a descent direction of negative curvature. The desire to simultaneously minimize the function in the directions in which the Hessian has positive eigenvalues and to move also in a direction of negative curvature can take several forms. Below is one form for which convergence to a second order point can be proved.

Two new step size procedures are necessary to implement this strategy. They are generalizations by McCormick [14] of those suggested by Armijo [1].

<u>SSP II</u> At iteration $k$ , given $x_k$ and $s_k$ a descent direction of search emanating from it, find $i(k)$ , the smallest integer from $i=0,1,\ldots$ , such that

$$f(x_k+s_k 2^{-i}) - f(x_k) \leq \alpha\, 2^{-i}\, s_k^T \nabla f(x_k) \quad ,$$

where $\alpha$ is a preassigned constant where $0 < \alpha < 1$ . Set

$$x_{k+1} = x_k + s_k 2^{-i(k)} \; .$$

<u>SSP III</u> At iteration $k$ , given $x_k$ , $s_k$ a descent direction, and $d_k$ a descent direction of negative curvature, find $i(k)$ , the smallest integer from $i=0,1,\ldots$ , such that

$$f[y_k(2^{-i})] - f(x_k) \leq \alpha\, 2^{-i}[s_k^T \nabla f(x_k) + \tfrac{1}{2}\, d_k^T \nabla^2 f(x_k) d_k] \; .$$

where $y_k(2^{-i}) \equiv x_k + s_k 2^{-i} + d_k 2^{-i/2}$ . Set

$$x_{k+1} = x_k + s_k 2^{-i(k)} + d_k 2^{-i(k)/2} \; .$$

(Here again, $\alpha$ is a preassigned constant with $0 < \alpha < 1$).

<u>Algorithm</u>. Let $x_0$ be a given point. In general, at iteration $k$ there is available a point $x_k$ . Movement to $x_{k+1}$ takes a different form depending upon which of two cases holds.

Case (i): The Hessian matrix $\nabla^2 f(x_k)$ is positive definite. Set $s_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$ and obtain $x_{k+1}$ using SSP II.

Case (ii): The Hessian matrix $\nabla^2 f(x_k)$ is not positive definite. Compute a descent direction $s_k$ and a descent direction of negative curvature $d_k$ . Use SSP III to obtain $x_{k+1}$ .

A specific realization of $s_k$ and $d_k$ in Case (ii) would be the negative gradient $-\nabla f(x_k)$ and $\pm e_k^{min}$ (with the sign chosen appropriately to make it a descent direction).

Again, the difficulty in using this strategy is in finding an efficient numerical procedure for computing $d_k$ and a natural choice for $s_k$ which allows for minimization of some "part of" $f(x)$ .

E. Another strategy, related to the fourth one, is to create an iteration which consists of several steps. Using $x_k$ as a base point, move successively optimizing along computed directions of negative curvature. The last step of the iteration is tc move so as to minimize the "positive part" of the quadratic approximation at $x_k$ .

Specifically, define the positive part of the Hessian $\nabla^2 f(x_k)$ to be

$$\mathscr{P}^k = \sum_{\lambda_j^k > 0} e_j^k \lambda_j^k (e_j^k) ,$$

where $E^k \lambda^k (E^k)^T$ is the eigenvalue-eigenvector decomposition of the Hessian matrix. Compute $P^k$ , a positive semi-definite matrix which is an approximation to $\mathscr{P}^k$ , and $d_1^k, \ldots, d_q^k$ , descent directions of negative curvature for it.

Set $y_1^k = x_k$ . In general, for the jth step of the kth iteration, set $s_j^k = d_j^k$ (where without loss of generality it is assumed that $(d_j^k)^T \nabla f(x_k) \leq 0$) and find $y_{j+1}^k$ by solving the usual step-size problem:

$$\underset{t \geq 0}{\text{minimize}} \quad f[y_j^k + s_j^k t] . \qquad (5)$$

This is to be done for $j=1,\ldots,q$ .

At the (q+1)st step the direction of search is set equal to

$$s_{q+1}^k = -(P^k)^{\dagger} \nabla f(x_k) ,$$

and the outcome of the optimal step size problem (5) yields a point which is taken to be $x_{k+1}$ .

Partial motivation for this algorithm can be given as follows. Suppose at $x_k$ there are $q$ negative eigenvalues associated with its Hessian matrix. Suppose further that in general the Hessian matrix of $f$ at any general point $x$ can be approximated by

$$\nabla^2 f(x) \doteq E\lambda(x)E^T , \tag{6}$$

where $EE^T = I$ , and $\lambda(x)$ is a diagonal matrix. This is tantamount to assuming that although the eigenvalues of the Hessian matrix of $f(x)$ may vary from it point to point, the associated eigenvectors do not change very much.

For definiteness assume that

$$\lambda_j(x_k) \leq 0 , \quad \text{for } j=1,\ldots,q ,$$

and

$$\lambda_j(x_k) > 0 , \quad \text{for } j=q+1,\ldots,n .$$

Set $q_j^k = \pm e_j$ (the sign chosen so that $(e_j)^T \nabla f(x_k) \leq 0$ ). Because the optimal step-size procedure (5) is used, it follows that

$$e_j^T \nabla f(y_{j+1}^k) = 0 , \quad \text{for } j=1,\ldots,q . \tag{7}$$

Because of the approximation above it follows that

$$\nabla f(y_{j+1}^k) = \nabla f(y_j^k) + \int_0^I E\lambda[y_j^k + (y_{j+1}^k - y_j^k)s]E^T ds(e_j t_j^k) , \quad \text{for } j=1,\ldots,q . \tag{8}$$

A simple induction argument shows that

$$e_j^T \nabla f(y_j^k) = e_j^T \nabla f(x_k) , \quad \text{for } j=1,\ldots,q . \tag{9}$$

If the problem is well-posed, i.e., if $\{x \mid f(x) \leq f(x_0)\}$ is a bounded set, then the step-size problem terminates at a finite point and the second order necessary conditions imply

$$e_j^T \nabla^2 f(y_{j+1}^k) e_j = \lambda_j(y_{j+1}^k) \geq 0 . \tag{10}$$

Thus this strategy brings the eigenvalues up to zero. Assume that strict inequality holds in (10). Assume further that the eigenvalues which were strictly positive have not changed at all (or, not very much). Then at $y^k_{q+1}$, by recomputing the new information, a usual Newton move would be made. We shall now show that no new computations are necessary, i.e., that the step (q+1) accomplishes this.

The Newton direction at $y^k_{q+1}$ is given by

$$-\nabla^2 f(y^k_{q+1})^{-1} \nabla f(y^k_{q+1}) = -\sum_{j=1}^{n} e_j \, \lambda_j(y^k_{q+1})^{-1} e_j^T \, \nabla f(y^k_{q+1}) \; . \qquad (11)$$

Now because of (7),

$$e_q^T \, \nabla f(y^k_{q+1}) = 0 \; .$$

For $1 \le j < q$, by virtue of (8) and the fact that $e_i^T e_j = 0$ for $i \neq j$, induction yields

$$e_j^T \, \nabla f(y^k_{q+1}) = e_j^T \, \nabla f(y^k_{j+1}) = 0$$

(using (7) again). Thus (11) above is equivalent to

$$-\sum_{j=q+1}^{n} e_j \, \lambda_j(y^k_{q+1})^{-1} e_j^T \, \nabla f(y^k_{q+1}) \; .$$

The fact that $e_i^T e_j = 0$, for $i \neq j$, coupled with (8), readily yields

$$e_j^T \, \nabla f(y^k_{q+1}) = e_j^T \, \nabla f(x_k) \; , \qquad \text{for } j=q+1,\ldots,n \; .$$

This, coupled with the assumption that the eigenvalues which were positive at $x_k$ do not change much (i.e., $\lambda_j(y^k_{q+1}) = \lambda_j(x_k)$ for $j=q+1,\ldots,n$), implies that

$$-\sum_{j=q+1}^{n} e_j \, \lambda_j(y^k_{q+1})^{-1} e_j^T \, \nabla f(y^k_{q+1}) = -\sum_{j=q+1}^{n} e_j \, \lambda_j(x_k)^{-1} e_j^T \, \nabla f(x_k)$$

which is just $-(\mathcal{P}^k)^T \, \nabla f(x_k)$ .

Example. The nonlinear programming problem is:

$$\underset{(x^1,\, x^2)}{\text{minimize}} \quad \sin(x^1+x^2) + (x^1-x^2)^2 - 3x^1/2 + 5x^2/2 \; .$$

The vector $(x^1, x^2)$ is further restricted so that $x^1 \geq -1.5$ , $x^2 \geq -2.5$ must hold. This is a nonconvex programming problem and has an infinite number of local minimizers. The plot of the problem is in Figure 1. Assume that the starting point of the process is $x_0 = (0., -.5)^T$ . The solution of the differential equation (1) starting from that point is given by

$$x^1(t) = [y^1(t) + y^2(t)]/2 \; , \quad x^2(t) = [y^1(t) - y^2(t)]/2 \; ,$$

where

$$y^1(t) = 2 \arctan \left[ \sqrt{3} \left[ \frac{-\exp(\sqrt{.75}t - .297007) + 1}{\exp(\sqrt{.75}t - .297007) + 1} \right] \right]$$

and

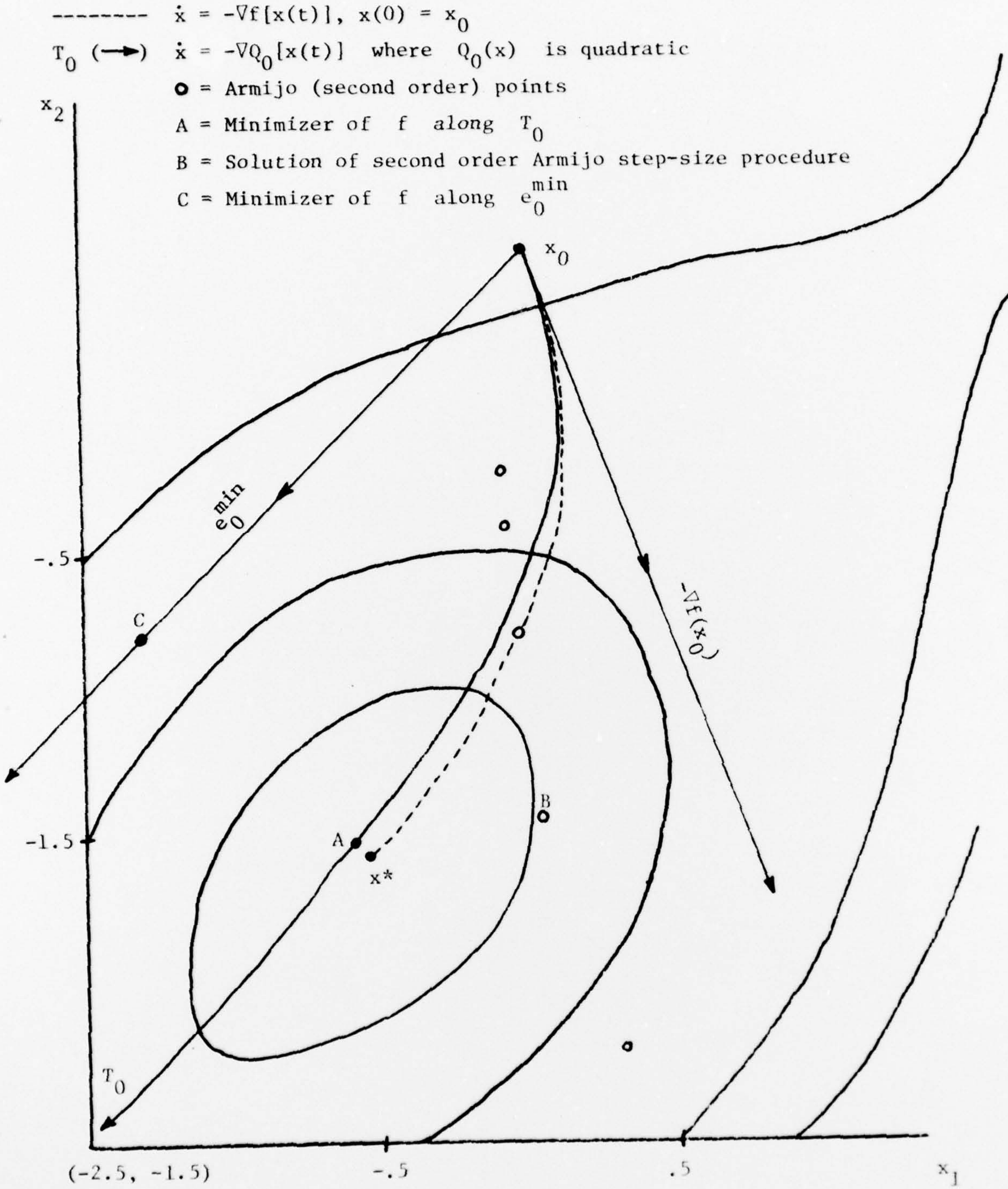
$$y^2(t) = 1.5 \exp(-2t) \; .$$

As $t$ approaches infinity this trajectory approaches the local unconstrained minimizer (-.5471975512, -1.547197551) . The trajectory is plotted (the dotted curve) in Figure 1.

At $x_0$ , $\nabla f(x_0)^T = (-1.6224, 4.3776)$ , and $\nabla^2 f(x_0) = \begin{pmatrix} 1.520574, & -2.479426 \\ -2.479426, & 1.520574 \end{pmatrix}$ . This has an eigenvalue (-.95885) in the direction $\pm(1,1)^T$ , and an eigenvalue (4.) in the direction $\pm(1,-1)^T$ . For simplicity the normalization of the eigenvectors will be incorporated into the diagonal matrix. The trajectory given by the quadratic approximation method (Strategy C) is

$$\begin{pmatrix} 0. \\ .5 \end{pmatrix} - \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} (1-\exp[-4t])/8 & 0 \\ 0 & -(1-\exp[.95885t])/1.9177 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1.6224 \\ 4.3376 \end{pmatrix} .$$

This trajectory is plotted in Figure 1. The point to notice is how closely this follows the solution of the differential equation. Initially the




$$\text{MIN} \quad \sin(x_1+x_2) + (x_1-x_2)^2 - 3x_1/2 + 5x_2/2 \quad \text{S.T.} \quad -2.5 \le x_1, \ -1.5 \le x_2$$

Figure 1.--Results of different strategies for a non-convex unconstrained minimization problem.

tangent to the trajectory is the steepest descent direction, and later it turns to point in the direction $(-1,-1)^T$. It passes very close to the unconstrained minimizer. The solution of the optimal step-size problem is at $t_0 = .67$ with $x_0(t_0) = (-.67, -1.50)^T$. The objective function value there is $f[x_0(t_0)] = -2.85$. This is given as point A in Figure 1.

Use of Strategy B on this example involves minimizing $f$ along the ray $(-1,-1)^T$ starting from $(0.,.5)^T$. The result of this is point C in Figure 1, approximately $(-1.2972, -.7972)^T$. The region of positive definiteness has been found and the regular Case (i) version of Newton's method would apply now.

The results of applying Strategy D are summarized in Table 1 and are also given in Figure 1. A value of $\alpha = 1/2$ was used in the Armijo procedure. For $i=0$ the point generated was outside the given bounds. The threshold criterion failed for $i=1$ but passed for $i=2$. The terminal point is labelled B in the figure and is approximately $(.0520, -1.43795)^T$.

Applying the classical version of Newton's method (4) at this point yields

$$x_2 = \begin{pmatrix} .0520 \\ -1.43795 \end{pmatrix} - \begin{pmatrix} 2.982964, & -1.017035 \\ -1.017035, & 2.982964 \end{pmatrix}^{-1} \begin{pmatrix} 1.6636955 \\ -.2961045 \end{pmatrix} = \begin{pmatrix} -.540799242 \\ -1.54079799 \end{pmatrix},$$

which is very close to the unconstrained minimizer.

The result using the general Strategy E is as follows. The positive portion of the Hessian matrix is

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix} 2(1,-1) = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} 4(1/\sqrt{2},\ 1/\sqrt{2}) .$$

The move from $x_0$ along the direction of negative curvature is that prescribed by Strategy B and can be used as a starting point. This is point C in Figure 1.

TABLE 1.

USE OF STRATEGY D ON EXAMPLE

| $i$ | $y_0(2^{-i})$ | $f[y_0(2^{-i})] - f(x_0)$ | $\frac{1}{2}\, 2^{-i}\left[-\lvert\lvert\nabla f(x_0)\rvert\rvert^2 + \frac{1}{2}\,(e_0^{min})^T \nabla^2 f(x_0) e_0^{min}\right]$ |
|---|---|---|---|
| 0 | (.9152, -4.5447) (Infeasible) | N.A. | -10.9632 |
| 1 | (.3112, -2.1688) | -2.41535 | -5.4816 |
| 2 | (.0520, -1.43795) | -4.4153308 | -2.7408 |

The direction vector is then

$$-(\mathcal{P}^0)^{\dagger}\ \nabla f(x_0) = -\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \frac{1}{4}\ (1/\sqrt{2},\ 1/\sqrt{2}) \begin{pmatrix} -1.6224 \\ 4.3776 \end{pmatrix} = \begin{pmatrix} .75 \\ -.75 \end{pmatrix}.$$

With a step size of one, this yields

$$\begin{pmatrix} -.5472 \\ -1.5472 \end{pmatrix} = \begin{pmatrix} -1.2972 \\ -.7972 \end{pmatrix} + \begin{pmatrix} .75 \\ -.75 \end{pmatrix}$$

In this example this is the desired unconstrained minimizer and no further computations are necessary.

## 4. Summary

In this paper several strategies have been presented for modifying the classical form of Newton's method when the Hessian matrix is not positive definite at some iterate. The emphasis was on the geometric motivation for the methods, rather than on convergence theorems which can be proved for specific algebraic implementations of the methods. In the papers referenced there is a general concensus that although the natural way to look at the methods is from the point of view of the eigenvector-eigenvalue decomposition of the Hessian matrix, it is also agreed that computationally this is too expensive. Most of the authors quoted intend to work on imitative algorithms which do not require this (and in some cases do not require explicit computation of the second derivatives).

The computational problems associated with these methods have not been elaborated upon here, nor have the difficulties in getting good test problems. Since the modifications to be made occur infrequently compared to the usual Newton or quasi-Newton steps, care in generating test situations to compare these different strategies must be taken.

REFERENCES

[1] ARMIJO, L. (1966). Minimization of functions having Lipschitz continuous first partial derivatives. Pacific J. Math. 16 (1) 1-3.

[2] BOTSARIS, C. A. and D. H. JACOBSEN (1976). A Newton-type curvilinear method for optimization. J. Math. Anal. Appl. 54 217-229.

[3] CAUCHY, A. (1971). Méthode générale pour la resolution des systèmes d'équations simultanées. C. R. Acad. Sci. Paris Sér. A-B 25 536-538.

[4] CODDINGTON, E. A. and N. LEVINSON (1955). Theory of Ordinary Differential Equations. McGraw-Hill, New York.

[5] DIXON, L. C. W., J. GOMULKA and S. E. HERSOM (1975). Reflections on the global optimization problem. The Hatfield Polytechnic Technical Report Number 64, Numerical Optimization Center.

[6] FIACCO, A. V. and G. P. McCORMICK (1968). Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Wiley, New York.

[7] FLETCHER, R. and T. L. FREEMAN (1975). A modified Newton method for minimization. Report Number 7, Department of Mathematics, University of Dundee, Dundee, Scotland.

[8] GILL, P. E. and W. MURRAY (1974). Newton-type methods for unconstrained and linearly constrained optimization. Math. Programming 7 311-350.

[9] GREENBERG, H. J. and J. E. KALAN (1972). Methods of feasible paths in nonlinear programming. Technical Report CP 72004, Computer/Operations Research Center, Southern Methodist University, Dallas, Texas.

[10] GREENSTADT, J. (1967). On the relative efficiencies of gradient methods. Math. Comp. **21** 360-367.

[11] LEVENBERG, U. (1944). A method for the solution of certain non-linear problems in least squares. Quart. Appl. Math. **2** 164-168.

[12] MARQUARDT, D. W. (1963). An algorithm for least squares estimation of nonlinear parameters. SIAM J. Appl. Math. **11** 431-441.

[13] MATTHEWS, A. and D. DAVIES (1969). A comparison of modified Newton methods for unconstrained optimisation. Report CMS/69/18, Category A, Mathematics and Statistics Group, Imperial Chemical Industries Limited.

[14] McCORMICK, G. P. (1976). Second order convergence using a modified Armijo step-size rule for function minimization. Technical Paper Serial T-328, Institute for Management Science and Engineering, The George Washington University.

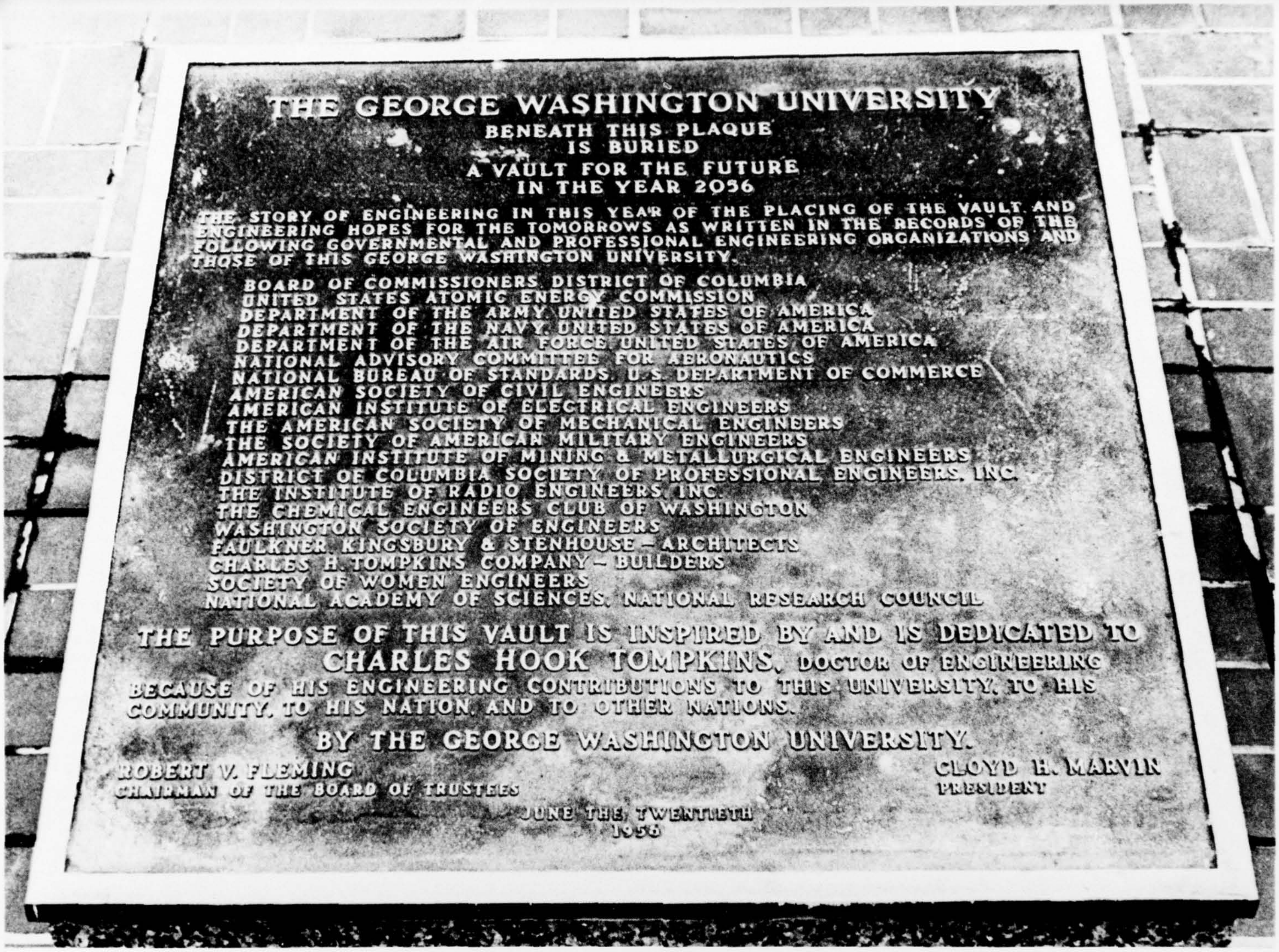
[15] McCORMICK, G. P. (1976). On computing the positive part and a direction of negative curvature for a symmetric matrix given in dyadic form. Technical Paper Serial T-337, Institute for Management Science and Engineering, The George Washington University.

[16] MURRAY, W. (1972). Second derivative methods. Numerical Methods for Unconstrained Optimization (W. Murray, ed.) 107-122. Academic Press, New York.

[17] ORTEGA, J. M. and W. C. RHEINBOLDT (1970). Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York.

[18] RAMSAY, J. O. (1970). A family of gradient methods for optimization. Comput. J. **13** (4) 413-417.

[19] VIAL, JEAN-PHILIPPE and I. ZANG (1975). Unconstrained optimization by approximation of the gradient path. CORE Discussion Paper No. 7513, Center for Operations Research and Econometrics, de Croylaan, 54, 3030 Heverlee, Belgium. Revised, August 1976.




To cope with the expanding technology, our society must be assured of a continuing supply of rigorously trained and educated engineers. The School of Engineering and Applied Science is completely committed to this objective.