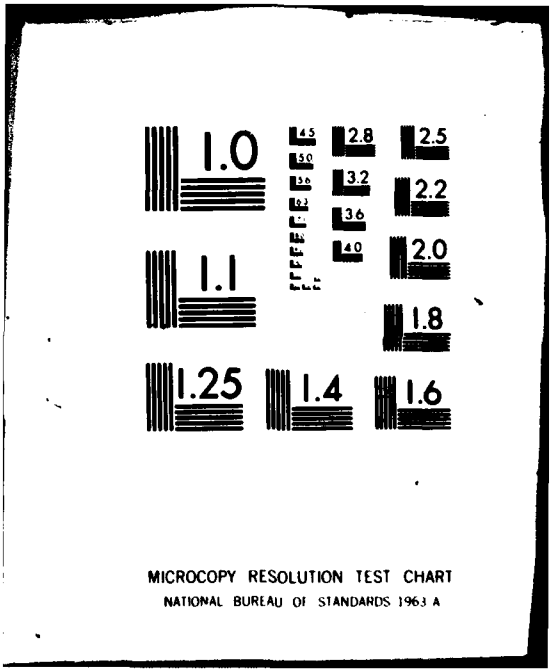
AD-A092 254 GEORGE WASHINGTON UNIV WASHINGTON DC INST FOR MANAGE--ETC F/G 12/1
CONVERGENCE THEORY FOR UNCONSTRAINED MINIMIZATION,(U)
SEP 80 G P MCCORMICK DAAG29-79-C-0062
UNCLASSIFIED SERIAL-T-431 NL

END
DATE FILMED
1 8
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

AD A092254

LEVEL II

12 B.S

THE GEORGE WASHINGTON UNIVERSITY

# STUDENTS FACULTY STUDY R ESEARCH DEVELOPMENT FUT URE CAREER CREATIVITY CC MMUNITY LEADERSHIP TECH NOLOGY FRONTIE SIGN ENGINEERING APP ENC GEORGE WASHIN NIV

DEC 1 1980
C

DDC FILE COPY

SCHOOL OF ENGINEERING AND APPLIED SCIENCE

8

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

(12)

# CONVERGENCE THEORY FOR UNCONSTRAINED MINIMIZATION

by

Garth P. McCormick

Serial T-431
30 September 1980

DTIC
DEC 1 1980

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

Contract DAAG-79-C-0062
Army Research Office
and
Contract N00014-75-C-0729
Project NR 347 020
Office of Naval Research

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

NONE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>14 SERIAL T-431 | 2. GOVT ACCESSION NO.<br>AD-A092 | 3. RECIPIENT'S CATALOG NUMBER<br>254 |
| 4. TITLE (and Subtitle)<br>6 CONVERGENCE THEORY FOR UNCONSTRAINED MINIMIZATION | | 5. TYPE OF REPORT & PERIOD COVERED<br>SCIENTIFIC |
| | | 6. PERFORMING ORG. REPORT NUMBER<br>29 T-431 |
| 7. AUTHOR(s)<br>10 GARTH P. McCORMICK | | 8. CONTRACT OR GRANT NUMBER(s)<br>15 DAAG-79-C-0062<br>N00014-75-C-0729 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>THE GEORGE WASHINGTON UNIVERSITY<br>INSTITUTE FOR MANAGEMENT SCIENCE & ENGINEERING<br>WASHINGTON, DC 20052 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>12 20 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>U.S. ARMY RESEARCH OFFICE<br>BOX 12211, RESEARCH<br>TRIANGLE PARK, NC 27709 | OFFICE OF NAVAL RESEARCH<br>CODE 434<br>ARLINGTON, VA 22217 | 12. REPORT DATE<br>30 September 1980<br>13. NUMBER OF PAGES<br>18 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>NONE<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC SALE AND RELEASE; DISTRIBUTION UNLIMITED.

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

18. SUPPLEMENTARY NOTES

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

NONLINEAR PROGRAMMING
CONVERGENCE THEORY
UNCONSTRAINED OPTIMIZATION

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

Convergence theory for unconstrained minimization centers on proving which characteristics hold at point of accumulation of "minimizing sequences" generated by unconstrained minimization algorithms. To a great extent, convergence proofs for particular algorithms have common elements. In this paper some results have been synthesized and put in a general context. Some applications of the theory are included.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

NONE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

406745

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

Abstract
of
Serial T-431
30 September 1980

# CONVERGENCE THEORY FOR UNCONSTRAINED MINIMIZATION

by

Garth P. McCormick

Convergence theory for unconstrained minimization centers on proving which characteristics hold at points of accumulation of "minimizing sequences" generated by unconstrained minimization algorithms. To a great extent convergence proofs for particular algorithms have common elements. In this paper some results have been synthesized and put in a general context. Some applications of the theory are included.

Research Supported by

Contract DAAG-79-C-0062
Army Research Office
*and*
Contract N00014-75-C-0729
Office of Naval Research

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist Avail and/or Special
A

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

# CONVERGENCE THEORY FOR UNCONSTRAINED MINIMIZATION

by

Garth P. McCormick

## 1. Introduction

Convergence theory for unconstrained minimization centers around proving which characteristics hold at points of accumulation of "minimizing sequences" generated by unconstrained minimization algorithms. First order convergence refers to proofs that accumulation points are stationary points. Second order convergence is concerned with the additional property that the Hessian matrix at a point of accumulation satisfies the second order necessary condition, namely, that the Hessian matrix there be positive semidefinite. To a great extent convergence proofs for particular algorithms have many common elements. In this paper these results have been synthesized and put in a general context. Applications are given in Section 5.

## 2. Step-size Procedures

For the unconstrained minimization problem

$$\min \quad f(x)$$

$$\text{s.t.} \quad x \in H \subseteq E^n \quad \text{(here } H \text{ is an open set)},$$

a general algorithm usually takes the following form.

At iteration $k$, given $x_k \in H$, construct a directed curve $y_k(t)$ parameterized by the single variable $t$. The curve should have the two properties that $y_k(0) = x_k$, and for $t$ positive and small $f[y_k(t)] \leq f(x_k)$. Use some suitable step size procedure to obtain a value $t_k$ and set

$$x_{k+1} = y_k(t_k) .$$

Most of the time, $y_k(t) = x_k + s_k t$, where $s_k$ is an $n \times 1$ direction of search.

Below are the five step size procedures which will be used at different times throughout this paper. First, definitions are required.

A vector $s_k$ is a *nonascent direction* at $x_k$ if

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

A vector $d_k$ is a *direction of nonpositive curvature* at $x_k$ if

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

A vector $s_k$ is a *descent direction* at $x_k$ if

$$s_k^T \nabla f(x_k) < 0 .$$

A vector $d_k$ is a *direction of negative curvature* at $x_k$ if

$$d_k^T \nabla^2 f(x_k) d_k < 0 .$$

The first three step size procedures are called *optimal* step size procedures and refer to the following problem: given a point $x_k \in H$, a given open set, and given a direction $s_k$, solve

$$\begin{aligned} &\text{minimize} \quad f[x_k + s_k t] \\ &t > 0 \end{aligned} \tag{1}$$

subject to the restriction that

$$t \in \{t > 0 \mid x_k + s_k t \in H\} .$$

FIRST LOCAL MINIMIZER (SSP I):

Set $t_k$ to be the *first* local minimizer for (1).

GLOBAL MINIMIZER (SSP II):

Set $t_k$ to be a *global* minimizer for (1).

LOCAL MINIMIZER WITH SMALLER FUNCTION VALUE (SSP III):

Set $t_k$ to be any local minimizer for (1) with the additional property that

$$f[x_k + s_k t_k] \leq f(x_k) .$$

There are several difficulties associated with the use of SSP I - SSP III. In the first place, a solution may not exist; i.e., any infimum may be taken on at a point $t_k$ where $x_k + s_k t_k$ is on the boundary of $H$. Any algorithm which uses these step size procedures is required to ascertain that a solution exists. (If it does, it is obviously unconstrained.) Second, even if solutions to SSP I and SSP II are known to exist, there is no guarantee usually that the desired $t_k$ can be found when the function $f(x)$ is a general not-necessarily-convex function. A method for solving SSP I and SSP II is described in [McCormick, 1979].

When, for fixed $x_k, s_k$ the function $f[x_k + s_k t]$ is convex in $t$, all three of the first step size procedures reduce to the same problem.

Most of the published convergence theorems for algorithms solving the unconstrained optimization problem rely on either SSP I or SSP II. The guaranteed ability to solve SSP III has not proved a strong enough tool for proving these theorems.

A third difficulty which has concerned algorithmists in recent years is that these step size procedures are idealized in that it is

usually impossible to find a local minimizer *exactly*. Convergence proofs and rate of convergence proofs relying on exact minimization are thus suspect. Some effort has been put forward in stating weaker requirements on the three optimal step size procedures. This amounts essentially to deciding how far from exact minimization one can come and still prove the desired theorems. The reader is referred to [Polak, 1971], [Cohen, 1972], and [McCormick and Ritter, 1974] for more on this subject.

When $f(x)$ is strictly convex, it is easy to show that the $t_k$ specified by the optimal step size procedures is the same and that an algorithm can be specified to guarantee finding the value (if it exists). For methods of conjugate directions, the optimal step size procedure is necessary in accelerating the rate at which the algorithm converges to the global minimizer. Many unconstrained minimization algorithms (e.g., Newton's method, and some quasi-Newton methods) do not depend upon optimal step size procedures for their rate of convergence properties. For these it is possible to use step size procedures of a type introduced by [Armijo, 1966]. Some generalizations of his approach are described below. As before, assume that $x_k \in H$, a given open set. Let $s_k, d_k$ be directions, where $s_k$ is one of nonascent and $d_k$ is one of nonascent and also nonpositive curvature. Let $0 < \alpha < 1$ be a preassigned constant.

FIRST ORDER ARMIJO (SSP IV):

Set $t_k = 2^{-i(k)}$, where $i(k)$ is the smallest integer from $i=0,1,\ldots$ such that

$$x_k + s_k 2^{-i} \in H ,$$

and

$$f[x_k + s_k 2^{-i}] - f(x_k) \leq \alpha 2^{-i} s_k^T \nabla f(x_k) .$$

In order for a finite value of $i$ satisfying the inequality to exist, it is sufficient that $s_k$ be a direction of descent. When

certain conditions on the sequence $\{s_k\}$ are met, it is possible to prove that points of accumulation of the sequence $\{x_k\}$ are stationary points. When variations of Newton's method are used which involve a direction of nonpositive curvature, it is possible to prove convergence (see [McCormick 1977] and Theorem 4) to a second order point when the following step size procedure is used.

SECOND ORDER ARMIJO (SSP V):

Set $t_k = 2^{-i(k)}$, where $i(k)$ is the smallest integer from $i=0,1,\ldots$ such that

$$y_k(2^{-i}) = x_k + s_k 2^{-i} + d_k 2^{-i/2} \in H ,$$

and

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

In order for a finite value $i(k)$ to exist, it is sufficient that $s_k, d_k$ be nonascent directions and in addition, that $s_k \nabla f(x_k) < 0$ whenever $\nabla f(x_k) \neq 0$, and

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

whenever $\nabla f(x_k) = 0$.

## 3. First Order Convergence

*Theorem 1.* Consider any algorithm for minimizing the continuously differentiable function $f(x)$ in the open set $H$ which has the following properties: the algorithm is a nonascent algorithm, i.e., $f(x_{k+1}) \leq f(x_k)$ for all $k$; consecutive points are of the form $x_{k+1} = x_k + s_k t_k$ where $s_k^T \nabla f(x_k) \leq 0$; and $t_k$ is found by solving the step-size problem SSP I. Let $\bar{x} \in H$ be a point of accumulation of $\{x_k\}$ and $K_1$ a set of indices such that $\lim_{k \in K_1} x_k = \bar{x}$. Assume that $\|s_k\| < M$ for all $k \in K_1$. Let $\bar{s}$ be any point of accumulation of $\{s_k\}$ for

$k \in K_1$ . Then

$$\bar{s}^T \nabla f(\bar{x}) = 0 \ .$$

*Proof:* Let $K_2 \subseteq K_1$ be a set of indices such that $\bar{s} = \lim_{k \in K_2} s_k$ . If $\bar{s} = 0$ , the theorem is obviously true. Assume otherwise.

Case (i). There exists a set of indices $K_3 \subseteq K_2$ such that $\lim_{k \in K_3} t_k = 0$ . Because the optimal step size problem generates $t_k$ ,

$$0 = \nabla f(x_k + s_k t_k)^T s_k \ .$$

Taking the limit (by assumption each $\|s_k\|$ is uniformly bounded above and $t_k \to 0$) ,

$$0 = \nabla f(\bar{x})^T \bar{s} \ .$$

Case (ii). Here, $\liminf_{k \in K_2} t_k = \bar{t} > 0$ . Let $K_4 \subseteq K_2$ be a set of indices such that $t_k \geq \bar{t}/2$ for all $k \in K_4$ . Assume the contrary of the theorem conclusion, i.e., $\nabla f(\bar{x})^T \bar{s} < -\delta < 0$ . Then there is a neighborhood $N(\bar{x})$ about $\bar{x}$ and a set of indices $K_5 \subseteq K_4$ such that for $x \in N(\bar{x})$ and $k \in K_5$ , $\nabla f(x)^T s_k \leq -\delta/2 < 0$ .

Let $\hat{t} > 0$ be a scalar small enough such that for all $0 < t < \hat{t}$ , all $k \in K_5$ ,

$$x_k + s_k t \in N(\bar{x}) \ .$$

Pick $t^* = \min(\bar{t}/2 , \hat{t})$ . Then

$$f(\bar{x}) - f(x_0) = \sum_{k=0}^{\infty} [f(x_{k+1}) - f(x_k)] \leq \sum_{k\in K_5} [f(x_{k+1}) - f(x_k)] \quad \text{(the nonascent property)}$$

$$\leq \sum_{k\in K_5} [f(x_k + s_k t^*) - f(x_k)] \quad \text{(because of the step-size procedure)} \quad (2)$$

$$= \sum_{k\in K_5} \nabla f(x_k + s_k \tau)^T s_k t^* \quad \text{(Taylor's Theorem with } 0 \leq \tau \leq t^*)$$

$$\leq \sum_{k\in K_5} -(\delta/2)t^* = -\infty \ .$$

This contradiction shows the truth of the theorem for Case (ii).

*Theorem 2* [Convergence of SSP II]: Consider the same hypotheses and statements as in Theorem 1 except that the step size procedure SSP II is used. Then the same conclusions hold.

*Proof:* Nothing in the proof of Theorem 1 changes except it is noted that Inequality (2) follows when SSP II is used because $t_k$ is chosen as a global minimizer in $H$ of $f(x)$ along the vector $s_k$ emanating from $x_k$.

*Comment:* Unfortunately, the same proof cannot be used for SSP I''. It is theoretically possible that if $t_k$ is chosen to be any local minimum [albeit with value less than $f(x_k)$ ], Inequality (2) may not be valid.

Under similar assumptions, convergence to a stationary point can be shown if the Armijo step size procedure SSP IV is used.

*Theorem 3:* Suppose SSP IV is used as the step size procedure. Let $\bar{x} \in H$ be a point of accumulation of $\{x_k\}$ and $K_1$ a set of indices such that $x_k \to \bar{x}$ for $k \in K_1$. Assume that $f \in C^1$, that there exists a value $\beta > 0$ such that

$$\|s_k\| \geq \beta \|f'(x_k)\| \tag{3}$$

for all $k \in K_1$, that $f'(x_k)s_k \leq 0$ for each $k$, that the sequence $\{s_k\}$ is uniformly bounded ($k \in K_1$), and that

$$\text{limsup}_{k \in K_1} \; f'(x_k)s_k \; / \; \|f'(x_k)\| \; \|s_k\| \; = \; -\delta < 0 \; . \tag{4}$$

Then

$$f'(\bar{x}) = 0 \; .$$

*Proof:* There are two cases to consider.

Case (i). The set of indices $\{i(k)\}$ for $k \in K_1$ is uniformly bounded above (by a number I). Because the sequence $\{f(x_k)\}$ is monotone decreasing and each $f'(x_k)s_k < 0$, it follows by summing appropriately that

$$f(\bar{x}) - f(x_0) \; < \; \sum_{k=0}^{\infty} \alpha 2^{-i(k)} \; f'(x_k)s_k \; < \; \sum_{k \in K_1} \alpha 2^{-i(k)} \; f'(x_k)s_k \; .$$

Further, it follows from (3), (4), and the assumed uniform bound on $i(k)$ that

$$2^{-i(k)} \; f'(x_k)s_k \; \leq \; -\alpha\delta\beta 2^{-I} \; \|f'(x_k)\|^2 \; ,$$

for each $k \in K_1$. It follows directly, then, that $f'(x_k) \to 0$ for $k \in K_1$, and by the continuity of $f'(x)$ that $f'(\bar{x}) = 0$.

Case (ii). There is a subset of indices $K_2 \subseteq K_1$ such that $\lim_{k \in K_2} i(k) = \infty$. If the cause for termination of iteration $k$ were that $x_k + s_k 2^{-i(k)+1} \notin H$ infinitely often, then since $i(k) \to \infty$ for $k \in K_2$, and because the $\{s_k\}$ are uniformly bounded, it follows that $\bar{x}$ is on the boundary of $H$. Since $H$ is an open set, $\bar{x} \notin H$, a contradiction to the theorem assumption. Thus, without loss of generality, assume for all $k \in K_2$,

$$f'(x_k)s_k 2^{-i(k)+1} + o\left(\|s_k\| 2^{-i(k)+1}\right) =$$

$$= f\left(x_k + s_k 2^{-i(k)+1}\right) - f(x_k) > \alpha 2^{-i(k)+1} f'(x_k)s_k \ .$$

Transposing, using (4) and dividing by $\|s_k\| 2^{-i(k)+1}$ yields

$$o\left(\|s_k\| 2^{-i(k)+1}\right) / 2^{-i(k)+1}\|s_k\| > (\alpha-1)f'(x_k)s_k / \|s_k\|$$

$$\geq (1-\alpha)\|f'(x_k)\| \ .$$

Because the $\{s_k\}$ are uniformly bounded, taking the limit as $k \to \infty$ for $k \in K_2$ yields the desired result.

## 4. Second Order Convergence

Some algorithms are concerned with producing accumulation points which have in addition to the stationarity property the property that their Hessian matrices are positive semidefinite. So far only algorithms which compute explicit second derivative information have been modified to produce this kind of convergence. It is theoretically possible that by using a finite difference approximation technique, similar convergence results could be obtained.

Algorithms which produce second order convergence must check the Hessian matrix at each iteration. If it is indefinite, a nonascent direction of nonpositive curvature must be computed as well as a nonascent direction. A convenient step size rule to use, then, is SSP V.

If the vector $s_k$ forms a sufficiently small angle with the negative gradient vector, and if the direction of nonpositive curvature acts sufficiently like an eigenvector associated with the minimum eigenvalue, then an interesting convergence theorem can be proved.

*Theorem 4* (Second Order Convergence of SSP V [McCormick 1979]): Assume that $f \in C^2$ in $H$. Suppose that in minimizing $f$ in $H$ an algorithm with the descent property uses SSP V an infinite number of times. [The *descent property* is simply that $f(x_{k+1}) < f(x_k)$ for all $k$.] Let $K_1$ be the infinite set of indices for which SSP V is used, and let $\bar{x}$ denote a point of accumulation in $H$ of $\{x_k\}$ for $k \in K_1$. Let $K_2 \subseteq K_1$ be a set of indices such that $\lim_{k \in K_2} x_k = \bar{x}$. Some regularity properties on the sequences $\{s_k\}$ and $\{d_k\}$ are required to hold. There exists a value $\beta > 0$ such that

$$\|s_k\| \geq \beta \|f'(x_k)\| \quad , \quad \text{for all} \quad k \in K_2 \; . \tag{5}$$

There is a value $\delta > 0$ such that

$$\text{limsup}_{k \in K_2} \frac{f'(x_k)}{\|f'(x_k)\|} \frac{s_k}{\|s_k\|} = -\delta < 0 \; . \tag{6}$$

There is a value $\gamma > 0$ such that

$$d_k^T f''(x_k) d_k \leq \left(e_k^{\min}\right)^T f''(x_k) e_k^{\min} \gamma \quad , \quad \text{for all} \quad k \in K_2 \; , \tag{7}$$

where $e_k^{\min}$ is an eigenvector of $f''(x_k)$ associated with its minimum eigenvalue. The sequences $\{s_k\}$ and $\{d_k\}$ are uniformly bounded. Then: $\bar{x}$ is a stationary point, i.e.,

$$f'(\bar{x}) = 0 \; ,$$

and

$$f''(\bar{x})$$

is a positive semidefinite matrix with at least one eigenvalue equal to zero.

*Proof:* There are two cases to consider.

Case (i). The integers $\{i(k)\}$ for $k$ in $K_2$ are uniformly bounded above by some value $I$. Because of the descent property, it

follows that all points of accumulation have the same function value, and

$$f(\bar{x}) - f(x_0) = \sum_{k=0}^{\infty} |f(x_{k+1}) - f(x_k)| \geq \sum_{k \in K_2} f(x_{k+1}) - f(x_k)$$

$$\geq \alpha \sum_{k \in K_2} 2^{-i(k)} \left[ f'(x_k)s_k + \frac{1}{2} d_k^T f''(x_k) d_k \right]$$

$$\geq \alpha \sum_{k \in K_2} 2^{-i} \left[ -\beta \|f'(x_k)\|^2 + \frac{1}{2} \left(e_k^{min}\right)^T f''(x_k) e_k^{min} \right] .$$

Since $f(\bar{x})$ is finite, and since each term in brackets is less than or equal to zero for each $k \in K_2$, it follows that $f'(\bar{x}) = 0$, and that $\bar{e}_{min}^T f''(\bar{x}) \bar{e}_{min} = 0$, where $\bar{e}_{min}$ is some accumulation point of $\{e_k^{min}\}$ for $k \in K_2$.

Case (ii). There is a subset $K_3 \subseteq K_2$ such that $\lim_{k \in K_3} i(k) = \infty$. Because of the definition of $i(k)$, then either

$$y_k(2^{-i(k)+1}) \notin H ,$$

or

$$f[y_k(2^{-i(k)+1})] - f(x_k) > \alpha 2^{-i(k)+1} [f'(x_k)s_k + \frac{1}{2} d_k^T f''(x_k) d_k] . \quad (8)$$

If the former condition held infinitely often, then because $y_k(2^{-i(k)+1}) \to \bar{x}$, also $(k \in K_3)$ it follows that $\bar{x}$ is on the boundary of $H$. Since $H$ is an open set, $\bar{x} \notin H$, a contradiction to the theorem hypothesis. Therefore, without loss of generality, (8) can be considered to hold for all $k \in K_3$.

Because $f \in C^2$ and because the sequences $\{s_k\}$ and $\{d_k\}$ are assumed to be uniformly bounded, the left-hand side of the inequality (8) can be written

$$f'(x_k)s_k 2^{-i(k)+1} + f'(x_k)d_k 2^{-\lfloor i(k)-1\rfloor/2}$$

$$+ \frac{1}{2}\left[s_k 2^{-i(k)+1} + d_k 2^{-\lfloor i(k)-1\rfloor/2}\right]^T f''(x_k)\left[s_k 2^{-i(k)+1} + d_k 2^{-\lfloor i(k)-1\rfloor/2}\right]$$

$$+ o(2^{-i(k)+1}) .$$

Combining like terms and incorporating where appropriate into $o(2^{-i(k)+1})$ yields [using the fact that $f'(x_k)d_k \leq 0$ ],

$$o(2^{-i(k)+1}) > (\alpha-1)\left[f'(x_k)s_k + \frac{1}{2} d_k^T f''(x_k)d_k\right]2^{-i(k)+1}$$

$$> (\alpha-1)\left[-\delta\beta\|f'(x_k)\|^2 + \frac{1}{2}\left(e_k^{\min}\right)^T f''(x_k)e_k^{\min}\gamma\right]2^{-i(k)+1} ,$$

[using (5), (6), and (7)]. Dividing by $2^{-i(k)+1}$ , taking the limit as $k \to \infty$ (for $k \in K_3$) yields, by the argument in Case (i), the desired result.

A different strategy for minimizing a function whose Hessian is not always positive definite is to compute a direction of nonpositive or negative curvature and optimize in that direction. The theorem below is useful for proving convergence of algorithms using this strategy.

*Theorem 5:* Assume as in Theorem 1, and in addition assume that $f(x)$ is twice continuously differentiable in $H$ . Then in addition to the conclusion of Theorem 1,

$$\bar{s}^T \nabla^2 f(\bar{x})\bar{s} \geq 0 .$$

*Proof:* Let $K_2 \subseteq K_1$ be a set of indices such that $\bar{s} = \lim_{k\in K_2} s_k$ . If $\bar{s} = 0$ the theorem is obviously true. Assume otherwise.

Case (i). There exists a set of indices $K_3 \subseteq K_2$ such that $\lim_{k\in K_3} t_k = 0$ . Because the optimal step size procedure generates $t_k$ ,

$$s_k^T \nabla^2 f(x_k + s_k t_k) s_k > 0 .$$

Taking the limit yields the desired result for Case (i).

Case (ii). Here, $\liminf_{k \in K_2} t_k = \bar{t} > 0$. Let $K_4 \subseteq K_2$ be a set of indices such that $t_k > \bar{t}/2$ for all $k \in K_4$. Assume the contrary of the theorem conclusion, i.e., that $\bar{s}^T \nabla^2 f(\bar{x}) \bar{s} = -\delta < 0$. Then there is a neighborhood $N(\bar{x})$ about $\bar{x}$ and a set of indices $K_5 \subseteq K_4$ such that for $x \in N(\bar{x})$, and $k \in K_5$, $s_k^T \nabla^2 f(x) s_k < -\delta/2 < 0$

Let $\hat{t} > 0$ be a scalar small enough such that for all $0 < t < \hat{t}$, all $k \in K_5$,

$$x_k + s_k t \in N(\bar{x}) .$$

Pick $t^* = \min(\bar{t}/2, \hat{t})$. Then

$$f(\bar{x}) - f(x_0) = \sum_{k=0}^{\infty} |f(x_{k+1}) - f(x_k)| < \sum_{k \in K_5} |f(x_{k+1}) - f(x_k)| \quad \text{(the non-ascent property)}$$

$$< \sum_{k \in K_5} |f(x_k + s_k t^*) - f(x_k)| \quad \text{(because of the step size procedure)}$$

$$< \sum_{k \in K_5} \left[ \nabla f(x_k)^T s_k t^* + s_k^T \nabla^2 f(x_k + s_k t) s_k (t^*)^2/2 \right] \quad \text{(Taylor's theorem with } 0 < t < t^*)$$

$$< \sum_{k \in K_5} s_k^T \nabla^2 f(x_k + s_k t) s_k (t^*)^2/2 \quad \text{(fact that } s_k^T \nabla f(x_k) < 0)$$

$$< \sum_{k \in K_5} -(\delta/2)(t^*)^2/2 = -\infty .$$

This contradiction proves the theorem for Case (ii).

## 5. Applications

The well-known method of steepest descent chooses as the direction of search each iteration the negative gradient vector, i.e., $s_k = -\nabla f(x_k)$. That points of accumulation generated by this method are stationary points follows directly from Theorem 1. The conclusion is that $\bar{s}^T \nabla f(\bar{x}) = 0$, which means that $\|\nabla f(\bar{x})\|^2 = 0$.

It is easy to cite other applications of these theorems to prove convergence of algorithms. They essentially shift the burden of proof to that of showing that the hypotheses of the theorems are as satisfied. In many cases this is difficult. The application to be pursued here is the modified Newton method. This type of algorithm (see [McCormick 1976] for a survey of these methods) is one which modifies the classical Newton procedure when a point is encountered where the Hessian matrix is not positive definite. The modification considered here is one which uses $n$ optimal steps at each iteration. It is not recommended since it is computationally prohibitive but is considered to illustrate the application of the general theorems.

Let $E_k(\Lambda_k)E_k^T = \nabla^2 f(x_k)$ be an eigenvalue-eigenvector reduction of the Hessian matrix at $x_k$, i.e., $E_k(E_k)^T = I$, and $\Lambda_k$ is a diagonal matrix. Let $e_j^k$ be the $j$*th* column of $E_k$. Set $y_1^k = x_k$. In general, for the $j$*th* step of the $k$*th* iteration, find $y_{j+1}^k$ by solving the step size problem (either SSP I or SSP II)

$$\underset{t \geq 0}{\text{minimize}} \quad f(y_j^k + e_j^k t)$$

$$\text{subject to} \quad t \in [t \mid y_j^k + e_j^k t \in \Omega] .$$

The sign is chosen so that $+ e_j^k$ is a nonascent direction at $y_j^k$.

This is to be done for $j=1,\ldots,n$ . The last point generated is taken to be $x_{k+1}$ .

*Theorem 6:* Assume that $f(x)$ is twice continuously differentiable in the open set $H$ . Suppose that the algorithm just described is applied to the unconstrained minimization of $f$ in $H$ . Assume that there is a single point of accumulation (call it $\bar{x}$ ) generated by the algorithm. Then $f'(\bar{x}) = 0$ , and $\nabla^2 f(\bar{x})$ is a positive semidefinite matrix.

*Proof:* Let $\bar{E}$ be a matrix of accumulation of $E_k$ with columns $\{\bar{e}_j\}$ . It follows from Theorem 1 (2) that $f'(\bar{x})\bar{e}_1 = 0$ . Since by assumption $\{y_2^k\}$ has the single point of accumulation $\bar{x}$ , from Theorem 1 (2), $f'(\bar{x})\bar{e}_2 = 0$ also. Inductively it is trivial to show that $f'(\bar{x})\bar{e}_j = 0$ , for $j=1,\ldots,n$ . Since the $\bar{e}_j$'s are linearly independent, it must be the case that $f'(\bar{x}) = 0$ .

Similar reasoning shows that $\bar{e}_j^T \nabla^2 f(\bar{x})\bar{e}_j > 0$ , for $j=1,\ldots,n$ , and thus that $\nabla^2 f(\bar{x})$ is a positive semidefinite matrix. Q.E.D.

Practical modified Newton algorithms for minimizing unconstrained functions differ in their strategies when faced with an indefinite Hessian, and in their computation of the estimates of the "positive part" of the Hessian and directions of nonpositive curvature. The theorems presented herein should be a help in proving convergence of such algorithms by isolating the components of the proof which are independent of the linear algebra used to generate the necessary quantities.

## REFERENCES

ARMIJO, L. (1966). Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16, (1), 1-3.

COHEN, A. (1972). Rate of convergence of several conjugate gradient algorithms. *SIAM J. Numerical Anal.*, 9, 248-259.

McCORMICK, G. P. (1976). Strategies for the minimization of an unconstrained nonconvex function. Technical Paper Serial T-343, Institute for Management Science and Engineering, The George Washington University (November).

McCORMICK, G. P. (1977). A modification of Armijo's step-size rule for negative curvature. *Math. Programming*, 13, (1), 111-115.

McCORMICK, G. P. (1979). Finding the global minimum of a function of one variable using the method of constant signed higher order derivatives. Technical Paper Serial T-411, Institute for Management Science and Engineering, The George Washington University (November).

McCORMICK, G. P. and K. RITTER (1974). Alternate proofs of the convergence properties of the conjugate gradient method. *J. Optimization Theory and Appl.*, 13, (5), 497-518.

POLAK, E. (1971). *Computational Methods in Optimization: A Unified Approach*. Academic Press, New York.