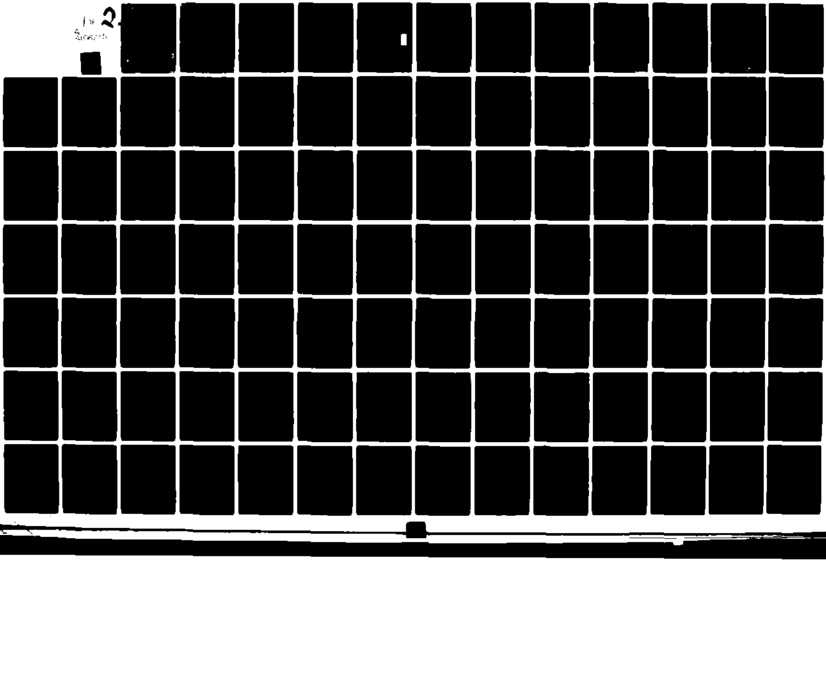
AD-A106 715 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH F/G 12/1
CONVERGENCE RATE ANALYSIS FOR ITERATIVE MINIMIZATION SCHEMES WI--ETC(U)
1981 G C HUGHES
UNCLASSIFIED AFIT-CI-81-17D NL



UNCLASS

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFIT-CI-81-17D
2. GOVT ACCESSION NO.: AD-A106 715
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Convergence Rate Analysis for Iterative Minimization Schemes with Quadratic Subproblems
5. TYPE OF REPORT & PERIOD COVERED: /THESIS/DISSERTATION
6. PERFORMING ORG. REPORT NUMBER: Doctoral thesis
7. AUTHOR(s): Capt George Crittenden Hughes
8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: AFIT STUDENT AT: North Carolina State Univ
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: AFIT/NR WPAFB OH 45433
12. REPORT DATE: 1981
13. NUMBER OF PAGES: 83
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
15. SECURITY CLASS. (of this report): UNCLASS
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)

20 OCT 1981

18. SUPPLEMENTARY NOTES: APPROVED FOR PUBLIC RELEASE: IAW AFR 190-17

FREDRIC C. LYNCH, Major, USAF
Director of Public Affairs
Air Force Institute of Technology (ATC)
Wright-Patterson AFB, OH 45433

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

NOV 5 1981

D

A

012200

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASS

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

81 10 26 141

AD A106715

DTIC FILE COPY

ABSTRACT

HUGHES, GEORGE CRITTENDEN. Convergence Rate Analysis for Iterative Minimization Schemes with Quadratic Subproblems. (Under the direction of JOSEPH C. DUNN.)

A large class of descent algorithms is analyzed for the problem $\min_\Omega f$, with Ω a convex subset of a Banach space X, and $F:X \to \mathbb{R}^1$ a differentiable functional. At each iteration a feasible direction $\hat{x}_n - x_n$ is determined, where $\hat{x}_n$ is a solution to the subproblem $\min_{y\in\Omega} \{\langle f'(x_n), y - x_n\rangle + \frac{1}{2}\langle M_n(y - x_n), y - x_n\rangle\}$ with $\{M_n\}$ a sequence of nonnegative linear operators with a uniform upper bound, and step lengths are obtained from Goldstein's rule. If f' is Lipschitz continuous and Ω is bounded, then limit points of sequences generated by this general scheme are extremals. A "worst case" convergence rate estimate of $r_n = f(x_n) - \inf_\Omega f = O(n^{-1/3})$ for convex f is shown to improve to $O(n^{-1})$ when either the condition numbers of the operators in the sequence $\{M_n\}$ are bounded away from zero or $0 \le \langle M_n u, u\rangle \le \langle f''(x)u, u\rangle$, $\forall x \in \Omega$, $\forall u \in X$, $\forall n \ge 0$; under these conditions a hierarchy or rate estimates exists ranging from finite termination of the process to $r_n = O(n^{-1})$ depending on how fast f grows near a unique minimizer ξ, i.e., depending on the value of ν in either of the conditions $\langle f'(\xi), x - \xi\rangle \ge \gamma\|x - \xi\|^\nu$ or $f(x) - f(\xi) \ge \gamma\|x - \xi\|^\nu$, $\forall x \in \Omega$, some $\gamma > 0$ and $\nu \in [1, \infty)$. A similar hierarchy of rate estimates is established for Newton's method ($M_n = f''(x_n)$) also depending on the growth of the convex functional f near ξ.

For twice differentiable, possibly nonconvex functionals f local conditions on the growth of the quadratic approximation to f at ξ in

81 10 26 141

directions leading into Ω are given as sufficient to insure linear or superlinear convergence of the sequence $\{\|x_n - \xi\|\}$ when the iterates pass sufficiently near ξ and the operators M_n are either uniformly positive definite or satisfy certain standard quasi-Newton conditions.

These results have potential applications to problems in optimal control theory.

81-17D

AFIT RESEARCH ASSESSMENT

The purpose of this questionnaire is to ascertain the value and/or contribution of research accomplished by students or faculty of the Air Force Institute of Technology (ATC). It would be greatly appreciated if you would complete the following questionnaire and return it to:

AFIT/NR
Wright-Patterson AFB OH 45433

RESEARCH TITLE: Convergence Rate Analysis of Iterative Minimization Schemes with Quadratic Subproblems

AUTHOR: Capt George C. Hughes

RESEARCH ASSESSMENT QUESTIONS:

1. Did this research contribute to a current Air Force project?

( ) a. YES ( ) b. NO

2. Do you believe this research topic is significant enough that it would have been researched (or contracted) by your organization or another agency if AFIT had not?

( ) a. YES ( ) b. NO

3. The benefits of AFIT research can often be expressed by the equivalent value that your agency achieved/received by virtue of AFIT performing the research. Can you estimate what this research would have cost if it had been accomplished under contract or if it had been done in-house in terms of manpower and/or dollars?

( ) a. MAN-YEARS ________ ( ) b. $________

4. Often it is not possible to attach equivalent dollar values to research, although the results of the research may, in fact, be important. Whether or not you were able to establish an equivalent value for this research (3. above), what is your estimate of its significance?

( ) a. HIGHLY SIGNIFICANT ( ) b. SIGNIFICANT ( ) c. SLIGHTLY SIGNIFICANT ( ) d. OF NO SIGNIFICANCE

5. AFIT welcomes any further comments you may have on the above questions, or any additional details concerning the current application, future potential, or other value of this research. Please use the bottom part of this questionnaire for your statement(s).

NAME GRADE POSITION

ORGANIZATION LOCATION

STATEMENT(s):

FOLD DOWN ON OUTSIDE - SEAL WITH TAPE

AFIT/NR
WRIGHT-PATTERSON AFB OH 45433

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE. $300

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL
FIRST CLASS PERMIT NO. 73236 WASHINGTON D.C.

POSTAGE WILL BE PAID BY ADDRESSEE

AFIT/ DAA
Wright-Patterson AFB OH 45433

FOLD IN

GEORGE C. HUGHES, Capt., USAF

Convergence Rate Analysis for Iterative Minimization Schemes with Quadratic Subproblems

1981

83 pages

PhD, Applied Mathematics

North Carolina State University

ABSTRACT

HUGHES, GEORGE CRITTENDEN. Convergence Rate Analysis for Iterative Minimization Schemes with Quadratic Subproblems. (Under the direction of JOSEPH C. DUNN.)

A large class of descent algorithms is analyzed for the problem $\min_\Omega f$, with Ω a convex subset of a Banach space X, and $F:X \to \mathbb{R}^1$ a differentiable functional. At each iteration a feasible direction $\hat{x}_n - x_n$ is determined, where $\hat{x}_n$ is a solution to the subproblem $\min_{y\in\Omega} \{\langle f'(x_n), y - x_n\rangle + \frac{1}{2}\langle M_n(y - x_n), y - x_n\rangle\}$ with $\{M_n\}$ a sequence of nonnegative linear operators with a uniform upper bound, and step lengths are obtained from Goldstein's rule. If f' is Lipschitz continuous and Ω is bounded, then limit points of sequences generated by this general scheme are extremals. A "worst case" convergence rate estimate of $r_n = f(x_n) - \inf_\Omega f = O(n^{-1/3})$ for convex f is shown to improve to $O(n^{-1})$ when either the condition numbers of the operators in the sequence $\{M_n\}$ are bounded away from zero or $0 \leq \langle M_n u, u\rangle \leq \langle f''(x)u, u\rangle$, $\forall x \in \Omega$, $\forall u \in X$, $\forall n \geq 0$; under these conditions a hierarchy or rate estimates exists ranging from finite termination of the process to $r_n = O(n^{-1})$ depending on how fast f grows near a unique minimizer ξ, i.e., depending on the value of ν in either of the conditions $\langle f'(\xi), x - \xi\rangle \geq \gamma\|x - \xi\|^\nu$ or $f(x) - f(\xi) \geq \gamma\|x - \xi\|^\nu$, $\forall x \in \Omega$, some $\gamma > 0$ and $\nu \in [1, \infty)$. A similar hierarchy of rate estimates is established for Newton's method $(M_n = f''(x_n))$ also depending on the growth of the convex functional f near ξ.

For twice differentiable, possibly nonconvex functionals f local conditions on the growth of the quadratic approximation to f at ξ in

directions leading into Ω are given as sufficient to insure linear or superlinear convergence of the sequence $\{\|x_n - \xi\|\}$ when the iterates pass sufficiently near ξ and the operators M_n are either uniformly positive definite or satisfy certain standard quasi-Newton conditions.

These results have potential applications to problems in optimal control theory.

North Carolina State University
School of Physical and Mathematical Sciences

Department of Mathematics
(919) 737-2382

Box 5548, Raleigh 27650

July 20, 1981

TO: AFIT/~~CI~~NR, Wright-Patterson AFB, OH, 45933

FROM: E. E. Burniston, Head, Mathematics Department

Capt. George C. Hughes has completed all requirements for the Degree of Ph.D. in Applied Mathematics. The Degree will be awarded on December 22, 1981.

☐ INITIAL ED PLAN ☐ REVISED ED PLAN *(Complete from 1st Term)* ☒ FINAL ED PLAN *(Note certificate on reverse side.)*

NAME OF EDUCATIONAL INSTITUTION: N. C. State University	SIGNATURE OF ADVISOR *(Not required on Final Ed Plan)*
MAJOR: Applied Mathematics / INPUT AFSC	TITLE OF ADVISOR
DEGREE SOUGHT: B BA BS M MA MS MBA D PHD MD DVM DDS DO LLB/JD OTHER ____	STUDENT SIGNATURE: George C. Hughes / DATE: 20 JUL 81
SSAN: 100-40-1906 (Circle one) FR FV / AERONAUTICAL RATING / CODE	FOR USE BY PROGRAM MANAGER CI( ) ONLY
NAME *(Rank, Last, First, MI, Present Mailing Address and Zip Code.)* PLEASE NOTE CHANGE OF: ( ) RANK ( ) ADDRESS ( ) PHONE NO. Capt. George C. Hughes 306 Rustic Ridge Rd. Cary, N. C., 27511	AFIT COMPLETION/TERMINATION DATE / MAF
TELEPHONE AREA CODE: 919 / TELEPHONE NUMBER: 469-0935	PROGRAM MANAGER APPROVAL / DATE

* For courses which satisfy degree requirements: Use a (1) to designate required courses *(no substitutions permitted)* and use a (2) to designate elective courses. For courses which do NOT specifically satisfy degree requirements: Use a (3) to designate pre-degree courses *(non-credit prerequisites)* and use a (4) to indicate extra or spare courses.

☐ CONTINUATION PAGE   TERMS LISTED ARE: ☒ SEMESTER ☐ QUARTER ☐ TRIMESTER ☐ ACADEMIC YEAR

1st TERM (8th TERM) DATES *(day, mo, year)* 28 Aug. 78 TO 20 Dec. 78

#	Dept & Course No.	Complete Course Title	Credit Hours	Grade
1	MA 515	Lin. Functional Anal. I	3	A
1	MA 600	Adv. Ordinary Diff. Eqns. I	3	A
1	MA 602	Partial Diff. Eqns. I	3	A
1	MA 611	Analytic Function Theory I	3	A
		Term Hrs	12	GPA

2nd TERM (9th TERM) DATES 8 Jan. 79 TO 9 May 79

#	Dept & Course No.	Complete Course Title	Credit Hours	Grade
1	MA 516	Lin. Functional Anal. II	3	A
1	MA 601	Adv. O.D.E. II	3	A
1	MA 613	Techniques of Complex Anal.	3	B
1	MA 623	Theory of Matrices and Applications	3	B
		Term Hrs	12	Cum GPA

3rd TERM (10th TERM) DATES 23 May 79 TO 28 Jun 79

#	Dept & Course No.	Complete Course Title	Credit Hours	Grade
1	MA 689	Special Topics in Applied Math.	3	A
		5 Jul 79 To: 10 Aug 79		
1	MA 683	Special Topics in Algebra	3	SAT
		Part A Written Qualifying Exam 25 Aug 79		PASSED
		Term Hrs	6	Cum GPA

AFIT FORM SEP 74 0-18   FC 5620   PREVIOUS EDITION OF THIS FORM MAY BE USED UNTIL EXHAUSTED.   **EDUCATION PLAN**

4th TERM (7th TERM) DATES (day, mo, year) 27 Aug 79 TO 19 Dec 79

*	Dept & Course No.	Complete Course Title	Credit Hours	Grade
2	MA 629	Vector Space Methods in Sys. Optimization	3	A
2	OR 531	Dynamical Systems and Multivariable Control	3	A
2	MA 689	Special Topics in Applied Math	6	A

Term Hrs 12 Cum GPA

5th TERM (12th TERM) DATES 9 Jan 80 TO 7 May 80

*	Dept & Course No.	Complete Course Title	Credit Hours	Grade
2	MA 689	Special Topics in Applied Math	3	A
2	MA 583	Special Topics in Numerical Solns. of O.D.E's	3	A
2	OR 506	Algorithmic Methods in Nonlinear Programming	3	A
1	MA 699	Research (Define Topic)	3	S
		Part B Written Preliminary Exam 19/21 Jan 80		PASSED

Term Hrs 12 Cum GPA

6th TERM (13th TERM) DATES 21 May 80 TO 25 Jun 80

*	Dept & Course No.	Complete Course Title	Credit Hours	Grade
1	MA 699	Research (Define Topic)	3	S
		30 Jun 80 To: 6 Aug 80		
1	FLF 401	French for Graduate Students	3	Au
		Language Exams: French 31 Jul 80		PASSED
		German 4 Aug 80		PASSED

Term Hrs 6 Cum GPA

7th TERM (14th TERM) DATES Aug 80 TO Dec. 80

*	Dept & Course No.	Complete Course Title	Credit Hours	Grade
1	MA 699	Research (Define Topic)	3	S
		Oral Preliminary Exam 16 Dec 80		PASSED

Term Hrs 3 Cum GPA

DISSERTATION/THESIS/MAJOR REPORT TITLE (Circle one)

Convergence Rate Analysis for Iterative Minimization Schemes

NUMBER OF CREDITS LISTED ON THIS EDUCATION PLAN 66

TRANSFER CREDITS ALLOWED TOWARD DEGREE RQMTS ONLY 21

TOTAL CREDITS REQUIRED BY INSTITUTION FOR DEGREE SOUGHT: No specific requirement; 87 is easily enough

CERTIFICATION REQUIRED IN FINAL EDUCATION PLAN

I hereby certify that I expended $300 for books and supplies and $280 for thesis, dissertation, or major report preparation. (Circle one)

[X] I completed degree requirements & did/will officially receive the degree of Ph.D (Applied Math) on 22 Dec 81 (date)

[X] An official transcript [(X) with degree posted] has been ordered and should arrive by (date) 1 Jan 82

[ ] I plan to complete degree requirements in absentia by (date) ______ . George C. Hughes
Student Signature

(USE ADDITIONAL 0-18 IF REQUIRED FOR CONTINUATIO[illegible]ERMS.) AFLC-WPAFB-APR 75 24M

☐ INITIAL ED PLAN ☐ REVISED ED PLAN *(Complete from 1st Term)* ☒ FINAL ED PLAN *(Note certificate on reverse side.)*

NAME OF EDUCATIONAL INSTITUTION N. C. State University	SIGNATURE OF ADVISOR *(Not required on Final Ed Plan)*
MAJOR / INPUT AFSC	TITLE OF ADVISOR
DEGREE SOUGHT B BA BS M MA MS MBA D PHD MD DVM DDS DO LLB/JD OTHER	STUDENT SIGNATURE George C. Hughes / DATE 20 JUL 81
SSAN *(Circle one)* FR FV / AERONAUTICAL RATING CODE	FOR USE BY PROGRAM MANAGER CI( ) ONLY
NAME *(Rank, Last, First, MI, Present Mailing Address and Zip Code.)* PLEASE NOTE CHANGE OF: ( ) RANK ( ) ADDRESS ( ) PHONE NO. Capt. Hughes, George C.	AFIT COMPLETION/TERMINATION DATE / MAF
TELEPHONE AREA CODE / TELEPHONE NUMBER	PROGRAM MANAGER APPROVAL / DATE

* For courses which satisfy degree requirements: Use a (1) to designate required courses *(no substitutions permitted)* and use a (2) to designate elective courses. For courses which do NOT specifically satisfy degree requirements: Use a (3) to designate pre-degree courses *(non-credit prerequisites)* and use a (4) to indicate extra or spare courses.

☒ CONTINUATION PAGE TERMS LISTED ARE: ☒ SEMESTER ☐ QUARTER ☐ TRIMESTER ☐ ACADEMIC YEAR

~~1st TERM~~ (8th TERM) DATES *(day, mo, year)* Jan 81 TO May 81

*	Dept & Course No.	Complete Course Title	Credit Hours	Grade
1	MA 699	Research (Prepare draft)	3	S

Term Hrs 3 GPA

~~2d TERM~~ (9th TERM) DATES May 81 TO 15 Jul 81

*	Dept & Course No.	Complete Course Title	Credit Hours	Grade
		Complete Dissertation		
		Final Oral Exam 14 Jul 81		PASSED

Term Hrs 0 Cum GPA 3.87

3rd TERM (10th TERM) DATES TO

*	Dept & Course No.	Complete Course Title	Credit Hours	Grade
		Selected for PI MU EPSILON		
		Honorary Mathematical Society		

Term Hrs Cum GPA

AFIT FORM SEP 74 0-18 FC 5620 PREVIOUS EDITION OF THIS FORM MAY BE USED UNTIL EXHAUSTED. **EDUCATION PLAN**

4th TERM (11th TERM) DATES *(day, mo, year)* ______ TO ______			
Dept & Course No.	*Complete Course Title*	*Credit Hours*	*Grade*
		Term Hrs	*Cum GPA*

5th TERM (12th TERM) DATES ______ TO ______			
		Term Hrs	*Cum GPA*

6th TERM (13th TERM) DATES ______ TO ______			
		Term Hrs	*Cum GPA*

7th TERM (14th TERM) DATES ______ TO ______			
		Term Hrs	*Cum GPA*

DISSERTATION/THESIS/MAJOR REPORT TITLE *(Circle one)*

NUMBER OF CREDITS LISTED ON THIS EDUCATION PLAN ______

TRANSFER CREDITS ALLOWED TOWARD DEGREE RQMTS ONLY ______

TOTAL CREDITS REQUIRED BY INSTITUTION FOR DEGREE SOUGHT ______

CERTIFICATION REQUIRED IN FINAL EDUCATION PLAN

I hereby certify that I expended $______ for books and supplies and $______ for thesis, dissertation, or major report preparation.
(Circle one)

☐ I completed degree requirements & did/will officially receive the degree of ______ on ______ *(date)*

☐ An official transcript [( ) with degree posted] has been ordered and should arrive by *(date)* ______

☐ I plan to complete degree requirements in absentia by *(date)* ______. ______
Student Signature

(USE ADDITIONAL 0-18 IF REQUIRED FOR CONTINUATION OF TERMS.) AFLC-WPAFB-APR 75 24M

CONVERGENCE RATE ANALYSIS FOR

ITERATIVE MINIMIZATION SCHEMES WITH QUADRATIC SUBPROBLEMS

by

GEORGE CRITTENDEN HUGHES

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

DEPARTMENT OF MATHEMATICS

RALEIGH

1 9 8 1

APPROVED BY:

Joe A. Marlin

R A Struble

Chairman of Advisory Committee

BIOGRAPHY

George Crittenden Hughes was born at West Point, New York, on December 31, 1947. As a member of a military family, he was raised on or near bases in the United States and Europe and graduated in 1965 from Madrid High School, in Madrid, Spain. He attended the U. S. Air Force Academy where he majored in Mathematics, and in June of 1970 he received a Bachelor of Science degree and a commission in the Air Force.

The author's first assignment was to North Carolina State University where he earned the degree of Master of Applied Mathematics, and in March of 1971 he entered pilot training in Valdosta, Georgia. From 1972 through 1976 he served tours in Germany, Greece, and Florida as a pilot of the C-130 before being assigned in 1976 to the faculty of the U. S. Air Force Academy in the Department of Mathematical Sciences. In 1978 he returned to North Carolina State University under Air Force sponsorship and will return to his teaching post at the Air Force Academy in July of 1981.

The author is married to the former Andrea M. Oosterlinck of Overmere, Belgium. The Hughes' have two children, a daughter, Eve, and a son, Kristiaan.

ACKNOWLEDGMENTS

The author wishes to express his deepest gratitude to two persons in particular: first, to his advisor, Professor Joseph C. Dunn, for his motivation, guidance, and encouragement in the research and preparation of this thesis, and to the author's wife, Andrea, for her understanding and continual moral support during the course of this study.

The author would also like to thank Professor Raimond A. Struble for his counsel and instruction over the last three years and Professors J. Luh and J. A. Marlin for their help as committee members.

The thesis was typed by Ms. Nancy Vincent, whose efficient and rapid work is greatly appreciated.

This research was supported by the U. S. Air Force through the Air Force Institute of Technology and by National Science Foundation grant ECS-8005958.

TABLE OF CONTENTS

1. Introduction

Let f be a real functional on a real Banach space X, i.e., $f: X \to \mathbb{R}^1$, and consider the following constrained optimization problem:

$$\text{(P)} \qquad \min_{x \in \Omega} f(x)$$

where Ω is a closed convex nonempty subset of X. A number of methods for solving (P) generate sequences of approximations to the solution via the following general process:

$$\text{(1.1a)} \qquad x_{n+1} = x_n + \omega_n(\hat{x}_n - x_n), \qquad \omega_n \in [0, 1],$$

where

$$\text{(1.1b)} \qquad \hat{x}_n \in \arg\min_{y \in \Omega} Q_n(y).$$

$Q_n(y)$ is a functional which approximates $f(y)$ near the vector x_n, and ω_n is a steplength parameter. Three examples of methods from this general class are:

A. The conditional gradient method corresponding to

$$Q_n = \langle f'(x_n), y - x_n \rangle;$$

here $f'(x_n)$ is the Fréchet derivative of f at x_n, and brackets, $\langle u, v \rangle$, denote the value of an element $u \in X^*$, the dual of X, operating on an element $v \in X$.

B. The "relaxed" form of the method of gradient projection corresponding to

$$Q_n(y) = \langle f'(x_n), y - x_n \rangle + \frac{1}{2\alpha_n} \|y - x_n\|^2, \ \alpha_n \geq \alpha > 0 \text{ for } n \geq 0.$$

For this class of methods, X is understood to be a Hilbert space.

C. The relaxed Newton's method corresponding to

$$Q_n(y) = \langle f'(x_n), y - x_n \rangle + \frac{1}{2} \langle f''(x_n)(y - x_n), y - x_n \rangle.$$

When the functionals $Q_n(y)$ are properly chosen, the vector $\hat{x}_n - x_n$ will be a feasible direction, i.e., for sufficiently small $\bar{\omega} \in (0, 1]$,

$$f(x_n + \omega(\hat{x}_n - x_n)) < f(x_n) \text{ for } \omega \in (0, \bar{\omega}),$$

provided x_n is not an extremal (see (2.1)). There are many methods for choosing suitable stepsize parameters ω_n which will insure that $f(x_{n+1}) < f(x_n)$ when $\hat{x}_n - x_n$ is a feasible direction. Most attempt to approximate the classical line minimization scheme in which one chooses the smallest ω_n satisfying

$$(1.2) \qquad \min_{\omega \in [0,1]} f(x_n + \omega(\hat{x}_n - x_n)).$$

For most nonlinear problems, however, (1.2) cannot be solved exactly, and methods which approximate (1.2) are necessary.

In addition to determining feasible directions, the functionals $Q_n(y)$ must have the property that subproblem (1.1b) is easy to solve relative to (P). For methods such as the conditional gradient method or the method of gradient projection on certain simple sets such as those often found in optimal control theory, (1.1b) is trivial. For more complicated functionals $Q_n(y)$ and constraint sets Ω, the utility of such methods becomes questionable. A number of authors have devised variations of the basic scheme to make the subproblem (1.1b) more feasible. Han [1] and Garcia Palomares and Mangasarian [2] minimize

$$Q_n(y) = \langle f'(x_n), y - x_n \rangle + \frac{1}{2} \langle M_n(y - x_n), y - x_n \rangle$$

over an approximation to Ω defined by linear inequalities in $\mathbb{R}^n$. In their method $\{M_n\}$ is a sequence of operators which approaches the second derivative operator of the Lagrangian of f and the constraints defining Ω. Bertsekas [3] uses a hybrid Newton method similar to gradient projection on simple sets such as orthants and cubes in $\mathbb{R}^n$. Such modifications can have considerable practical importance in special cases; however, the convergence behavior of the basic method (1.1) itself is still only partially understood.

The purpose of this thesis is to establish the convergence properties of the class of algorithms (1.1) in which $Q_n(y)$ is of the form

$$(1.3) \qquad Q_n(y) = \langle f'(x_n), y - x_n \rangle + \frac{1}{2} \langle M_n(y - x_n), y - x_n \rangle$$

where each M_n is a nonnegative bounded linear operator, i.e., $M_n \in BL(X, X^*)$ and

$$(1.4) \qquad 0 \leq \langle M_n u, u \rangle, \qquad \forall u \in X.$$

Although many different stepsize rules have been investigated for methods in this general scheme (GS), the essential differences in the algorithms lie in the selection of the operator sequence $\{M_n\}$ and not in the method of choosing the stepsize. In fact, a number of papers have compared major stepsize rules (see, e.g., [4], [5], [6]), and the basic conclusion is that differences in convergence rates are minimal. In the analysis to follow, the Goldstein rule described in Chapter 2 will be used since it is prototypic of the rules for approximating line minimization (1.2). There are several good reasons for carrying out the analysis in the setting of a general Banach space; in particular, by retaining the maximum degree of flexibility at the outset it is possible to obtain sharper bounds on convergence rates for function space minimization problems later on (see Remarks 3.2, 4.2).

The methods in the (GS) have for the most part been analyzed quite thoroughly for convex differentiable functionals f with "regular" minimizers. However, when f is non-convex or when singularities exist at the minimizers, the analysis has been less thorough and in some cases sketchy. Recent work has focused on understanding the behavior of the algorithms under these less tractable conditions. The following brief review of the major results on convergence and rate of convergence of methods embedded in the (GS) will put into perspective the results of this thesis.

In Chapter 3 it will be shown that when X is a Hilbert space and $M_n = \frac{1}{\alpha_n} I$, where I is the identity operator, then the (GS) is the same as the method of gradient projection introduced by Goldstein [7] in which

$$x_{n+1} = P_\Omega(x_n - \alpha_n \nabla f(x_n)). \tag{1.5}$$

Here P_Ω is the operation of projection onto Ω and $\nabla f(x_n)$ is the Hilbert space representor of $f'(x_n)$ in X. The parameter α_n is chosen to insure convergence with stepsize parameter $\omega_n = 1$ for $n \geq 0$. Levitin and Poljak [8] first gave rate of convergence results for this method for convex f using the "threshold" rule

$$(1.6) \qquad 0 < \epsilon_1 \leq \alpha_n \leq \frac{2}{L + \epsilon_2}, \qquad \epsilon_2 > 0,$$

where L is a Lipschitz constant for f', i.e.,

$$L \geq \sup_{\substack{x,y \in \Omega \\ x \neq y}} \frac{\|f'(x) - f'(y)\|}{\|x - y\|} .$$

For functionals satisfying the uniform convexity condition

$$(1.7) \qquad \underline{\mu}\,\|u\|^2 \leq \langle f''(x)u, u\rangle \leq \overline{\mu}\,\|u\|^2, \qquad \forall x \in \Omega, \forall u \in X,$$

$$\text{and } 0 < \underline{\mu} \leq \overline{\mu} < \infty,$$

the values $r_n = f(x_n) - \inf_\Omega f$ converge linearly, i.e., $r_n = O(\lambda^n)$ for some $\lambda \in [0, 1)$. In the absence of (1.7) the convergence of $\{r_n\}$ for convex functionals will be at least like $O(\frac{1}{n})$. Similar results were obtained by Demyanov and Rubinov [9] who investigated four variations of relaxed gradient projection in which the sequence $\{\alpha_n\}$ and the sequence of stepsize parameters $\{\omega_n\}$ are selected by combinations of threshold rules like (1.6) and line minimization (1.2). Dunn [10] found that the method (1.5) with the sequence $\{\alpha_n\}$ determined by a Goldstein-like rule converges linearly if the functional grows near an optimal point, or extremal, $\xi \in \Omega$, like the square of the distance from ξ, i.e., (1.8) $$f(x) - f(\xi) \geq \gamma\|x - \xi\|^2, \qquad \forall x \in \Omega,\ \gamma > 0.$$

This occurs when (1.7) is satisfied at $x = \xi$ or when the structure of the set near the extremal is such that

(1.9) $$\langle f'(\xi), x - \xi \rangle \geq \gamma\|x - \xi\|^2, \qquad \forall x \in \Omega,\ \gamma > 0,$$

since in the case of convex functionals (1.9) implies (1.8). In fact, Dunn was able to show for a wider class of functionals which are pseudo-convex in the sense of Mangasarian [11], that a complete hierarchy of rates can be determined from the condition

(1.10) $$f(x) - f(\xi) \geq \gamma\|x - \xi\|^\nu, \qquad \forall x \in \Omega,\ \nu \in [1, \infty),\ \gamma > 0,$$

ranging from finite termination of the process (i.e., $x_N = \xi$ for some $N \geq 0$) when $\nu = 1$ to rates approaching the "worst case" rate of $O(\frac{1}{n})$ as ν assumes larger values.

The conditional gradient method [8], [9], [12] results when the operators M_n in the (GS) are the zero operator for $n \geq 0$. With steplength rules of the line minimization type, this algorithm was shown in [8] and [9] to converge at the rate $r_n = O(\frac{1}{n})$ for convex functionals with Lipschitz continuous Fréchet derivatives on convex closed bounded sets. In these investigations, however, it could not be shown that conditions of the sort (1.7) had any effect on the convergence rate of the conditional gradient method (c.f. gradient projection method); a linear convergence rate was established only under certain strong uniform convexity conditions on the set Ω when $f'(\xi) \neq 0$. In fact, an example of Canon and

Cullum [13] shows that even when (1.7) is satisfied a rate of $O(\frac{1}{n})$ can not be improved upon without imposing conditions on the set Ω. Dunn [4] proved that uniform convexity of Ω is actually a very strong sufficient condition for linear convergence of the sequence $\{r_n\}$ and that the weaker condition (1.9) will suffice. Dunn [14] has also shown that, as in [10], a hierarchy of convergence rate upper bounds exists for the conditional gradient method depending on the value of the parameter ν in the condition

$$(1.11) \qquad \langle f'(\xi), x - \xi \rangle \geq \gamma \|x - \xi\|^{\nu}, \qquad \forall x \in \Omega,\ \nu \in [1, \infty),\ \gamma > 0.$$

Conditions of this type are satisfied in various Banach spaces by "bang-bang" optimal controls [4], [17] (see Remarks 3.2, 4.2).

Allwright [15] and Barnes [16] both considered variations of the (GS) in which specific operator sequences $\{M_n\}$ are used in certain optimal control settings. Allwright specified operators which have the property

$$(1.12) \qquad 0 \leq \langle M_n u, u \rangle \leq \langle f''(x)u, u \rangle, \qquad \forall x \in \Omega,\ \forall u \in X,\ \forall n \geq 0.$$

Although he was able to prove convergence using a stepsize rule similar to Goldstein's on bounded sets with convex functionals, he established a linear convergence rate for the sequence $\{r_n\}$ only when $\{M_n\}$ satisfies

$$(1.13) \qquad \mu\|u\|^2 \leq \langle M_n u, u \rangle, \qquad \forall u \in X,\ \forall n \geq 0,\ \mu > 0.$$

which with (1.12) implies (1.7). Barnes also required condition (1.7) with operators satisfying (1.13) to achieve linear rates for $\{r_n\}$.

If f is convex the operator f'' is certainly nonnegative on Ω and the Newton methods treated by Kantorovich [18], Goldstein [19], and

Levitin and Poljak [8] are formally in the (GS) with $M_n = f''(x_n)$ for $n \geq 0$. Very little has been written about convergence rates for Newton's method in the absence of the regularity condition (1.7) or when the second derivative operator is not at least positive definite at the extremal. Levitin and Poljak [8] who first proposed the constrained version of Newton's method with $x_{n+1} = \hat{x}_n$, relied on condition (1.7) to prove superlinear convergence of the sequence $\{\|x_n - \xi\|\}$ to zero. Danilin [20] gave a proof of convergence of the method for convex functionals on bounded sets with a stepsize rule similar to Goldstein's but, once again, required condition (1.7) for rates. It was stated by Bulavskii [21] for finite dimensional spaces that condition (1.7) can be relaxed to a condition on the growth of the second order approximation to f at the extremal ξ, namely

$$(1.14) \qquad \langle f'(\xi), x - \xi \rangle + \frac{1}{2} \langle f''(\xi)(x - \xi), x - \xi \rangle \geq \gamma \|x - \xi\|^2, \quad \forall x \in \Omega, \ \gamma > 0.$$

For convex functionals this condition insures superlinear convergence of the sequence $\{\|x_n - \xi\|\}$. Dunn [22] independently formulated and proved the same result in general Banach spaces and showed that when (1.14) holds with the exponent 2 replaced by 1, then finite termination of the process occurs.

The results mentioned so far have been restricted to convex or pseudoconvex functionals. Although a number of articles have given convergence results for these methods for general non-convex functionals (e.g. [23], [5], [24]) there are very few convergence <u>rate</u> results. For projected gradient methods Goldstein [23] proved that positive definiteness of the second derivative operator at a local minimizer ξ is sufficient to give a linear rate of convergence of the sequence $\{\|x_n - \xi\|\}$ if $x_n \to \xi$; however, for constrained minimization problems, this condition is rather strong. It was shown by Bertsekas [24] that the second derivative operator does not even have to be nonnegative at an extremal to achieve linear convergence in projected gradient schemes. For certain simple sets such as orthants and cubes in $\mathbb{R}^n$, Bertsekas proved that if the first derivative at an extremal ξ is positive in coordinate directions leading into the set and the second derivative at ξ is positive definite in the subspace parallel to the manifold of active constraints, then iterates generated by the gradient projection method and passing sufficiently near the extremal will converge to the extremal at a linear rate. Similar conditions are given by Han [1] and Garcia Palomares and Mangasarian [2] for their quasi-Newton methods to achieve linear and superlinear rates of convergence for sequences coming close enough to extremals. Their methods are modifications of the (GS) as indicated earlier and are in fact included in the (GS) when Ω is defined by linear inequalities in $\mathbb{R}^n$.

In Chapter 2 of the present thesis it will be shown that no matter how the sequence $\{M_n\}$ is chosen, as long as the operators are nonnegative and uniformly bounded above, every limit point of the generated sequence will be an extremal, and if f is convex the rate of convergence of $\{r_n\}$ will be $r_n = O(n^{-1/3})$ at least.

A number of results will be established in Chapter 3 for the (GS) when $\{M_n\}$ satisfies either condition (1.12) or a condition requiring a uniform lower bound on the "condition numbers" of the operators. Note that in gradient projection methods, the operators $\frac{1}{\alpha_n} I$ have condition numbers equal to 1. The "worst case" rate of convergence for this subclass will be $O(\frac{1}{n})$ for the sequence $\{r_n\}$ when f is convex. This extends the results reported by Demyanov and Rubinov [9], who considered only bounded sequences $\{\alpha_n\}$ for the relaxed gradient projection method. Their rate of $r_n = O(\frac{1}{n})$ holds for any sequence $\{\alpha_n\}$ which is bounded below and for stepsizes determined by Goldstein's rule. A hierarchy of convergence rate upper bounds will be established for this subclass, as was done in [10] for the gradient projection method and [14] for the conditional gradient method. When ω_n is bounded away from zero the higher rates of convergence depend on the growth rate of f near ξ (see (1.10)). On the other hand, if ω_n can be arbitrarily small, then higher rates will depend on how slowly ω_n decreases, which, in turn, can be estimated in the presence of conditions on the structure of the set near the extremal, i.e., condition (1.11).

As indicated previously, results for Newton's method have been superlinear rate estimates or better for the sequence $\{\|x_n - \xi\|\}$ under regularity conditions like (1.7) or (1.14). In Chapter 4 it is shown that a hierarchy of rates for the sequence $\{r_n\}$ exists here for non-regular extremals when condition (1.11) holds with ν in the range $1 \leq \nu \leq 5$. Although somewhat incomplete these results corroborate the belief that even in nonregular cases, Newton's method outperforms the first order methods. These ideas are developed further in an example from optimal control theory.

In Chapter 5 the results in [1], [2], [23], [24] for non-convex functionals will be extended to the (GS) in Banach spaces. When $\{M_n\}$ satisfies (1.13) or when (1.9) is satisfied at ξ and the second order approximation to f at ξ satisfies

$$(1.17) \qquad \langle f'(\xi), x - \xi \rangle + \frac{1}{2}\langle f''(\xi)(x - \xi), x - \xi \rangle \geq \gamma \|x - \xi\|^2.$$

for $x \in K_\Omega(\xi) \cap B_\rho(\xi)$ for some $\rho > 0$, where $K_\Omega(\xi)$ is the tangent cone to Ω at ξ with vertex at ξ, i.e., $K_\Omega(\xi) = \{x \in X: \xi + t(x - \xi) \in \Omega$ for some $t > 0\}$, and $B_\rho(\xi)$ is a closed ball of radius ρ around ξ, then if the sequence of iterates comes sufficiently near ξ, it will converge to ξ and $f(x_n) - f(\xi) = O(\lambda^n)$ for $\lambda \in (0, 1)$. Condition (1.17) need hold only for $x \in \Omega \cap B_\rho(\xi)$ if M_n is symmetric as well as nonnegative and M_n approximates $f''(\xi)$ in one of the following four ways: either

$$(1.18) \qquad \|M_n - f''(\xi)\| < \varepsilon,$$

for ε sufficiently small and $n \geq N > 0$, or

$$(1.19) \qquad \frac{\|(M_n - f''(\xi))(x - \xi)\|}{\|x - \xi\|} < \varepsilon,$$

for ε sufficiently small and for $x \in \Omega$ and $n \geq N > 0$, or

$$(1.20) \qquad \|M_n - f''(\xi)\| \to 0, \qquad \text{as } n \to \infty,$$

or

$$(1.21) \qquad \frac{\|(M_n - f''(\xi))(x - \xi)\|}{\|x - \xi\|} \to 0, \qquad \text{for } x \in \Omega \text{ as } n \to \infty.$$

For sequences in which x_n comes close enough to ξ for $n \geq N$, x_n will converge to ξ at a linear rate of convergence when either (1.18) or (1.19) is satisfied by $\{M_n\}$, or at a superlinear rate of convergence when either (1.20) or (1.21) holds. Conditions (1.18) - (1.21) and symmetry of the operators in $\{M_n\}$ are typical conditions placed on quasi-Newton operators in the literature (e.g., [25], [2], [26]).

It will be assumed in what follows that at each step of the (GS) at least one solution to (1.1b) exists. The existence question for (1.1b) can and should be separated from the convergence rate analysis (e.g., topologies suitable for treating the former may be inappropriate for the latter). In any case the emphasis here is on convergence and rate of convergence properties of sequences in the (GS), on the assumption that such sequences exist.

2. General Convergence Results

A necessary condition for a vector x to minimize a differentiable functional f on a convex set Ω is that x be an extremal, i.e.,

$$(2.1) \qquad \langle f'(x), y - x \rangle \geq 0, \qquad \forall y \in \Omega.$$

If f is pseudoconvex (and, in particular, convex) then (2.1) is also a sufficient condition [11]. A differentiable functional f is pseudoconvex if and only if $\langle f'(x), y - x \rangle > 0$ whenever $f(x) > f(y)$ and $x, y \in \Omega$. Most of the results to follow for convex functionals can be extended to a broad subclass of pseudoconvex functionals discussed in Remark 2.4. The general scheme outlined below is designed to construct a sequence $\{x_n\}$ whose limit points are extremals of f on Ω.

Let x_n be the $n^{\underline{th}}$ approximation to the solution of (P) generated by the (GS). Recall that in this scheme, the vector $\hat{x}_n$ is determined by

$$(2.2) \qquad \hat{x}_n \in \arg \min_{y \in \Omega} Q(M_n, x_n, y),$$

where the functional $Q(M_n, x, y)$ is defined by

$$(2.3) \qquad Q(M_n, x_n, y) = \langle f'(x_n), y - x_n \rangle + \frac{1}{2} \langle M_n(y - x_n), y - x_n \rangle.$$

It will also be required that the sequence of nonnegative bounded linear operators $\{M_n\}$ be uniformly bounded above, i.e.,

$$(2.4) \qquad \|M_n\| \leq K, \qquad \forall n \geq 0, \text{ and } K < \infty.$$

The next approximation is then given by

$$x_{n+1} = x_n + \omega_n(\hat{x}_n - x_n),$$

where the stepsize parameter ω_n is chosen by Goldstein's rule [23] and is computed at each iteration as follows: Define

$$g(x, \hat{x}, \omega) = \frac{f(x) - f(x + \omega(\hat{x} - x))}{\omega\langle f'(x), x - \hat{x}\rangle}$$

when $\langle f'(x), x - \hat{x}\rangle \neq 0$. Fix $\delta \in (0, \frac{1}{2})$ and if $g(x_n, \hat{x}_n, 1) \geq \delta$ then set $\omega_n = 1$; if not, determine any $\omega \in (0, 1)$ for which

(2.5) $$\delta \leq g(x_n, \hat{x}_n, \omega) \leq 1 - \delta$$

and set $\omega_n = \omega$. If $\langle f'(x_n), x_n - \hat{x}_n\rangle = 0$, set $\omega_n = 0$.

It will always be true that $\langle f'(x_n), x_n - \hat{x}_n\rangle \geq 0$, and $\langle f'(x_n), x_n - \hat{x}_n\rangle = 0$ if and only if x_n is an extremal (see Remark 2.1 below). Also, $g(x_n, \hat{x}_n, \omega)$ is a continuous function of ω on $(0, 1]$ by the continuity of f, and $\lim_{\omega \to 0^+} g(x_n, \hat{x}_n, \omega) = 1$ since by Taylor's formula

$$f(x_n+\omega(\hat{x}_n-x_n))-f(x_n)=\omega\langle f'(x_n),\hat{x}_n-x_n\rangle + o(\omega\|\hat{x}_n-x_n\|),$$

and, therefore,

$$g(x_n, \hat{x}_n, \omega) = 1 - \frac{o(\omega)}{\omega} .$$

It follows, then, that Goldstein's rule when used in the (GS) is well defined and will determine ω_n after a finite number of calculations, since if $g(x_n, \hat{x}_n, 1) < \delta$ then (2.5) is satisfied for $\omega \in [a, b] \subset (0, 1)$ for some $a \neq b$. Normally some sort of bisection procedure is used to locate an element of such an interval.

The following two lemmas are fundamental in what follows:

Lemma 2.1. Suppose that the sequences $\{r_n\} \subset [0, \infty)$ and $\{q_n\} \subset [0, \infty)$ satisfy

$$(2.6) \qquad r_{n+1} \leq r_n - q_n r_n^k, \qquad \forall n \geq 0,$$

for k a fixed exponent in the range $(1, \infty)$. If

$$q_n \geq q > 0,$$

then

$$(2.7) \qquad r_n = O(n^{-1/(k-1)}).$$

Proof. See [10], Lemma 4.1 for the proof.

Lemma 2.2. Let f be Fréchet differentiable. Let $M:X \to X^*$ be a nonnegative operator and Ω a convex subset of a Banach space X. For any $x \in \Omega$, let $\hat{x} \in \Omega$ satisfy

$$(2.8) \qquad \hat{x} \in \arg\min_{y \in \Omega} Q(M, x, y).$$

Let

$$(2.9) \qquad \Phi(x) = \{z \in \Omega : \langle f'(x), x - z \rangle \geq 0\}.$$

Then for any $z \in \Phi(x)$

$$(2.10) \qquad \langle f'(x), x-\hat{x}\rangle \geq \begin{cases} \langle f'(x), x-z\rangle + \frac{1}{2}\langle M(x-\hat{x}), x-\hat{x}\rangle, \text{if } \langle M(x-z, x-z\rangle = 0 \\ \frac{1}{2}\min\{\langle f'(x), x-z\rangle, \frac{\langle f'(x), x-z\rangle^2}{\langle M(x-z), x-z\rangle}\} + \frac{1}{2}\langle M(x-\hat{x}), x-\hat{x}\rangle, \\ \qquad \text{if } \langle M(x-z), x-z\rangle > 0 . \end{cases}$$

<u>Proof</u>. For any $z \in \Phi(x)$ and any $\theta \in [0, 1]$ the convex combination $z_\theta = x + \theta(z - x)$ is also in Ω since Ω is convex. From (2.8) it follows that for $\theta \in [0, 1]$

$$0 \leq \langle f'(x), z_\theta - x\rangle + \frac{1}{2}\langle M(z_\theta - x), z_\theta - x\rangle - \langle f'(x), \hat{x} - x\rangle - \frac{1}{2}\langle M(\hat{x} - x), \hat{x} - x\rangle$$

or

$$\langle f'(x), x - \hat{x}\rangle \geq \langle f'(x), x - z_\theta\rangle - \frac{1}{2}\langle M(z_\theta - x), z_\theta - x\rangle + \frac{1}{2}\langle M(\hat{x} - x), \hat{x} - x\rangle .$$

By the linearity of f' and M one can write

$$(2.11) \qquad \langle f'(x), x - \hat{x}\rangle \geq \theta\langle f'(x), x - z\rangle - \frac{\theta^2}{2}\langle M(z - x), z - x\rangle + \frac{1}{2}\langle M(\hat{x} - x), \hat{x} - x\rangle .$$

The sharpest bound is obtained by maximizing the right side of (2.11) over $\theta \in [0, 1]$. If $\langle M(x - z), x - z\rangle = 0$, then letting $\theta = 1$ yields

(2.12) $$\langle f'(x), x - \hat{x} \rangle \geq \langle f'(x), x - z \rangle + \frac{1}{2} \langle M(\hat{x} - x), \hat{x} - x \rangle.$$

If $\langle M(x - z), x - z \rangle > 0$, then

$$P(\theta) = \theta \langle f'(x), x - z \rangle - \frac{1}{2} \theta^2 \langle M(x - z), x - z \rangle$$

is a quadratic polynomial with maximum value at

(2.13) $$\hat{\theta} = \frac{\langle f'(x), x - z \rangle}{\langle M(x - z), x - z \rangle} \geq 0.$$

If $\hat{\theta} < 1$, then from (2.11) with $\theta = \hat{\theta}$

(2.14) $$\langle f'(x), x - \hat{x} \rangle \geq \frac{\langle f'(x), x - z \rangle^2}{2\langle M(x - z), x - z \rangle} + \frac{1}{2} \langle M(x - \hat{x}), x - \hat{x} \rangle.$$

If $\hat{\theta} \geq 1$, then it follows from (2.13) that

$$\langle f'(x), x - z \rangle \geq \langle M(x - z), x - z \rangle,$$

and setting $\theta = 1$ in (2.11) yields

(2.15) $$\langle f'(x), x - \hat{x} \rangle \geq \langle f'(x), x - z \rangle - \frac{1}{2} \langle f'(x), x - z \rangle$$

$$+ \frac{1}{2} \langle M(x - \hat{x}), x - \hat{x} \rangle = \frac{1}{2} \langle f'(x), x - z \rangle + \frac{1}{2} \langle M(x - \hat{x}), x - \hat{x} \rangle.$$

The lower bound (2.10) follows from (2.12), (2.14), and (2.15).

QED

Remark 2.1. Lemma 2.2 shows that if at any step of the (GS) it is determined that $\langle f'(x_n), x_n - \hat{x}_n \rangle = 0$ then x_n is an extremal. Suppose it were not an extremal. Then for some $y \in \Omega$, $\langle f'(x_n), y - x_n \rangle < 0$. But that would mean that $y \in \Phi(x_n)$ and from Lemma 2.2, that

$\langle f'(x_n), x_n - \hat{x}_n \rangle \geq k \langle f'(x_n), x_n - y \rangle$ for some $k > 0$. Therefore, $\langle f'(x_n), x_n - \hat{x}_n \rangle > 0$ and this contradiction shows that $\langle f'(x_n), y - x_n \rangle \geq 0$, $\forall y \in \Omega$, whenever $\langle f'(x_n), x_n - \hat{x}_n \rangle = 0$. Also, for any $x \in \Omega$, $x \in \Phi(x)$, and letting $z = x$ in (2.11) it follows that

$$(2.16) \qquad \langle f'(x), x - \hat{x} \rangle \geq \frac{1}{2} \langle M(x - \hat{x}), x - \hat{x} \rangle \geq 0, \qquad \forall x \in \Omega.$$

If x_n is an extremal then by (2.1)

$$(2.17) \qquad \langle f'(x_n), \hat{x}_n - x_n \rangle \geq 0.$$

From (2.16) and (2.17) one can conclude that if x_n is an extremal, then $\langle f'(x_n), x_n - \hat{x} \rangle = 0$. Note also that if at any step of the process it is determined that

$$(2.18) \qquad x_n \in \arg\min_{y \in \Omega} Q(M_n, x_n, y),$$

then x_n is an extremal, since one can choose $\hat{x}_n = x_n$, and that would make $\langle f'(x_n), x_n - \hat{x}_n \rangle = 0$. If x_n is an extremal, then (2.18) holds since, if not, there exists a $y \in \Omega$ such that

$$0 < Q(M_n, x_n, x_n) - Q(M_n, x_n, y),$$

but since $Q(M_n, x_n, x_n) = 0$, it follows from the definition (2.3) of $Q(M_n, x_n, y)$ that

$$\langle f'(x_n), y - x_n \rangle < -\frac{1}{2} \langle M_n(y - x_n), y - x_n \rangle \leq 0,$$

which is a contradiction (see (2.1)). Summarizing the above remarks, one has that x_n is an extremal if and only if $\langle f'(x_n), x_n - \hat{x}_n \rangle = 0$ if and only if x_n satisfies (2.18).

Remark 2.2. Define Ω_f as the set of minimizers of f on Ω. If f is convex, then

$$\Omega_f \subset \Phi(x), \qquad \forall x \in \Omega. \tag{2.19}$$

This follows from the fact that for convex f, if $\xi \in \Omega_f$, then

$$\langle f'(x), x - \xi \rangle \geq f(x) - f(\xi) \geq 0 \qquad \forall x \in \Omega. \tag{2.20}$$

If f is pseudoconvex, then it follows from the definition of pseudoconvexity that

$$\Omega_f \subset \Phi(x) \qquad \text{for } x \in \Omega - \Omega_f. \tag{2.21}$$

For functionals f on convex bounded sets Ω the (GS) will produce sequences whose limit points are extremals; this is shown in the following theorem.

Theorem 2.1. Let $\Omega \subset X$ where X is a Banach space and Ω is convex and bounded. Let f be (Fréchet) differentiable and let f' be Lipschitz continous with Lipschitz constant L, i.e., there exists an $L > 0$ such that $\|f'(x) - f'(y)\| \leq L\|x - y\|$, $\forall x, y \in \Omega$. Then f is bounded below, $\{f(x_n)\}$ is nonincreasing and converges to some limit $\ell \geq \inf_\Omega f > -\infty$, and every limit point of a sequence $\{x_n\}$ generated by the (GS) is an extremal.

If f is also convex, then the values $r_n = f(x_n) - \inf_\Omega f$ decrease monotonically to zero at least at the rate $r_n = O(n^{-1/3})$, and limit points are minimizers.

Proof. Using Taylor's formula and the Lipschitz continuity of f' one can write

(2.22) $$f(x_n)-f(x_n+\omega_n(\hat{x}_n-x_n))=\int_0^1\langle f'(x_n+\theta(x_n+\omega_n(\hat{x}_n-x_n)-x_n)),\omega_n(x_n-\hat{x}_n)\rangle d\theta$$

$$=\omega_n\langle f'(x_n),x_n-\hat{x}_n\rangle+\omega_n\int_0^1\langle f'(x_n+\theta\omega_n(\hat{x}_n-x_n))-f'(x_n),x_n-\hat{x}_n\rangle d\theta$$

$$\geq\omega_n\langle f'(x_n),x_n-\hat{x}_n\rangle-\omega_n^2\frac{L}{2}\|x_n-\hat{x}_n\|^2.$$

From Goldstein's rule (2.5), if $\omega_n < 1$ then

$$1-\delta \geq \frac{f(x_n)-f(x_{n+1})}{\omega_n\langle f'(x_n),\ x_n-\hat{x}_n\rangle},$$

and with (2.22) there results

$$1-\delta \geq 1-\frac{\omega_n L\|x_n-\hat{x}_n\|^2}{2\langle f'(x_n),\ x_n-\hat{x}_n\rangle},$$

which gives

(2.23) $$\omega_n \geq \min\{1,\ \frac{2\delta\langle f'(x_n),\ x_n-\hat{x}_n\rangle}{L\|x_n-\hat{x}_n\|^2}\}.$$

Let $D = \text{diameter}\ \Omega = \sup_{x,y\in\Omega}\|x-y\|$. Then

(2.24) $$\omega_n \geq \min\{1,\ \frac{2\delta}{LD^2}\langle f'(x_n),\ x_n-\hat{x}_n\rangle\}.$$

Also from Goldstein's rule (2.5) one has

$$f(x_n) - f(x_{n+1}) \geq \delta\omega_n \langle f'(x_n), x_n - \hat{x}_n \rangle \tag{2.25}$$

$$\geq \delta \min\{\langle f'(x_n), x_n - \hat{x}_n \rangle, \frac{2\delta}{LD^2} \langle f'(x_n), x_n - \hat{x}_n \rangle^2\},$$

and, therefore, $f(x_n) - f(x_{n+1}) \geq 0$ since $\langle f'(x_n), x_n - \hat{x}_n \rangle \geq 0$ for $n \geq 0$. It follows easily from the Lipschitz continuity of f' and the boundedness of Ω that $\inf_{\Omega} f > -\infty$ and so

$$\lim_{n\to\infty}(f(x_n) - f(x_{n+1})) = 0,$$

which, in turn, implies that

$$\lim_{n\to\infty} \langle f'(x_n), x_n - \hat{x}_n \rangle = 0.$$

Thus, if ξ is a limit point and $\{x_{n_k}\}$ is a subsequence converging to ξ, then

$$\lim_{n\to\infty} \langle f'(x_{n_k}), x_{n_k} - \hat{x}_{n_k} \rangle = 0. \tag{2.26}$$

Suppose that ξ is not an extremal, that is, for some $z \in \Omega$

$$\langle f'(\xi), z - \xi \rangle = -a < 0.$$

Then by the continuity of f' and the fact that $x_{n_k} \to \xi$ it follows that

$$\lim_{n\to\infty} \langle f'(x_{n_k}), z - x_{n_k} \rangle = -a,$$

and, therefore, there exists an $N > 0$ such that for $n_k \geq N$

$$\langle f'(x_{n_k}), z - x_{n_k} \rangle \leq \frac{-a}{2}$$

or

$$\langle f'(x_{n_k}),\ x_{n_k} - z\rangle \geq \frac{a}{2} > 0 .$$

This implies that $z \in \Phi(x_{n_k})$ for $n_k \geq N$. Therefore, by Lemma 2.2, assuming $\langle M_{n_k}(x_{n_k} - z),\ x_{n_k} - z\rangle > 0$, one obtains

$$(2.27)\quad \langle f'(x_{n_k}),\ x_{n_k} - \hat{x}_{n_k}\rangle \geq \frac{1}{2}\min\{\langle f'(x_{n_k}), x_{n_k} - z\rangle, \frac{\langle f'(x_{n_k}), x_{n_k} - z\rangle^2}{\langle M_n(x_{n_k} - z), x_{n_k} - z\rangle}\}$$

$$\geq \frac{1}{2}\min\{\langle f'(x_{n_k}),\ x_{n_k} - z\rangle,\ \frac{\langle f'(x_{n_k}),\ x_{n_k} - z\rangle^2}{KD^2}\}$$

$$\geq \frac{1}{2}\min\{\frac{a}{2},\ \frac{a^2}{4KD^2}\} > 0 .$$

If $\langle M_{n_k}(x_{n_k} - z),\ x_{n_k} - z\rangle = 0$, then

$$(2.28)\qquad \langle f'(x_{n_k}),\ x_{n_k} - \hat{x}_{n_k}\rangle \geq \frac{a}{2} > 0 .$$

But (2.27) and (2.28) contradict (2.26) and it follows that

$$\langle f'(\xi),\ z - \xi\rangle \geq 0 , \qquad \forall z \in \Omega.$$

Let f be convex, then since any extremal is a minimizer , one can conclude that any limit point $\xi \in \Omega_f$. For every $n \geq 0$ let $z_n \in \Omega$ be such that $f(x_n) - f(z_n) \geq \frac{1}{2}(f(x_n) - \inf_{\Omega} f) = \frac{1}{2} r_n$. Then $z_n \in \Phi(x_n)$ since by the convexity of f one can write

$$(2.29)\qquad \langle f'(x_n),\ x_n - z_n\rangle \geq f(x_n) - f(z_n) \geq \frac{1}{2} r_n \geq 0.$$

Therefore, by Lemma 2.2,

$$(2.30)\quad \langle f'(x_n), x_n - \hat{x}_n \rangle \geq \frac{1}{2} \min\{\langle f'(x_n), x_n - z_n \rangle, \frac{\langle f'(x_n), x_n - z_n \rangle^2}{KD^2}\}$$

$$\geq \frac{1}{2} \min\{\frac{r_n}{2}, \frac{r_n^2}{4KD^2}\} \geq 0,$$

and (2.25) yields

$$(2.31)\qquad f(x_n) - \inf_{\Omega} f - f(x_{n+1}) + \inf_{\Omega} f$$

$$\geq \delta \min\{\frac{r_n}{4}, \frac{r_n^2}{8KD^2}, \frac{\delta r_n^2}{8LD^2}, \frac{\delta r_n^4}{32LK^2D^6}\}.$$

Since $\lim_{n\to\infty} \langle f'(x_n), x_n - \hat{x}_n \rangle = 0$, it follows from (2.30) that $r_n \to 0$, and from (2.31) with r_n sufficiently small one has

$$r_{n+1} \leq r_n - qr_n^4, \qquad \text{for } q > 0,$$

and the rate $r_n = O(n^{-1/3})$ follows from Lemma 2.1.

QED

Remark 2.3. In proving that every limit point is an extremal, the crucial fact is that

$$\lim_{n\to\infty} \langle f'(x_n), x_n - \hat{x}_n \rangle = 0.$$

As shown in the theorem, this will be true for any operator sequence $\{M_n\}$ in the (GS) provided that Ω is bounded. The condition of boundedness can be removed if it can be established that $\omega_n \geq \omega > 0$, for all $n \geq 0$, and $\inf_{\Omega} f > -\infty$. In the next chapter it will be shown that the stepsize parameters are bounded away from zero when the operator sequence $\{M_n\}$ satisfies (1.13) and for certain other methods in the (GS). In these cases the theorem is true for $\Omega = X$ provided $\inf_X f > -\infty$.

Remark 2.4. It is easy to confirm from the proof of Theorem 2.1 and Remark 2.2 that the convergence rate of $O(n^{-1/3})$ for convex functionals can be extended to pseudoconvex functionals which satisfy

$$(2.32) \qquad \langle f'(x), x - \xi \rangle \geq \kappa(f(x) - f(\xi)) \quad \forall x \in \Omega - \Omega_f, \forall \xi \in \Omega_f, \kappa > 0.$$

In [10], Dunn establishes (2.32) for a large subclass of pseudoconvex functionals which includes certain concave functionals.

3. Convergence Rates for Convex Functionals

In the previous chapter it was shown that for any sequence of nonnegative operators, bounded uniformly above, the "worst case" convergence rate of sequences $\{f(x_n) - \inf_\Omega f\}$ generated by the (GS) is $O(n^{-1/3})$ for convex f. In [8], [9] and other references, however, a "worst case" rate of $r_n = O(\frac{1}{n})$ is established for the conditional gradient method and the gradient projection method. In Theorem 3.1 it is shown that the rate $r_n = O(\frac{1}{n})$ holds for a large class of methods in the (GS) whose operator sequences satisfy either of the following two additional conditions:

$$(3.1a) \qquad \underline{\mu}_n\|u\|^2 \leq \langle M_n u, u\rangle \leq \overline{\mu}_n\|u\|^2, \qquad \forall u \in X,\ 0 < \underline{\mu}_n \leq \overline{\mu}_n < \infty, \quad \text{for } n \geq 0,$$

with

$$(3.1b) \qquad \frac{\underline{\mu}_n}{\overline{\mu}_n} \geq a > 0, \qquad \text{for } n \geq 0,$$

or, if f is twice Fréchet differentiable and

$$(3.2) \qquad 0 \leq \langle M_n u, u\rangle \leq \langle f''(x)u, u\rangle, \qquad \forall u \in X,\ \forall x \in \Omega \text{ for } n \geq 0.$$

Note that Allwright [15] specifies condition (3.2) in his method. Also, the conditional gradient method, which uses $M_n = 0$ for $n \geq 0$, is admitted by condition (3.2) for convex functionals, since $f''(x)$ is nonnegative on Ω in this case. Methods whose operator sequences satisfy (3.1) include Barnes' method [16] and the method of gradient projection in Hilbert space in which

(3.3) $\hat{x}_n = P_\Omega(x_n - \alpha_n \nabla f(x_n))$.

The operation of projection of $x_n - \alpha_n \nabla f(x_n)$ onto Ω, for $\alpha_n > 0$, is equivalent to solving

(3.4) $\hat{x}_n = \arg\min_{y \in \Omega} \{\langle \nabla f(x_n), y - x_n \rangle + \frac{1}{2\alpha_n}\|y - x_n\|^2\}$,

since (3.3) is defined as

(3.5) $\hat{x}_n = \arg\min_{y \in \Omega} \|y - (x_n - \alpha_n \nabla f(x_n))\|^2$

or

(3.6) $\hat{x}_n = \arg\min_{y \in \Omega} \{2\alpha_n \langle \nabla f(x_n), y - x_n \rangle + \|y - x_n\|^2 + \alpha_n^2 \|\nabla f(x_n)\|^2\}$,

the solution of which satisfies (3.4). The operator $M_n = \frac{1}{\alpha_n} I$ in (3.4) clearly satisfies (3.1) with $\underline{\mu}_n = \overline{\mu}_n = \frac{1}{\alpha_n}$. The relaxed gradient projection schemes in Demayanov and Rubinov [9] specify explicit upper and lower bounds for the sequence $\{\alpha_n\}$. Condition (3.1) does not need that restriction, although the requirement (2.2) that $\{M_n\}$ be bounded above imposes a lower bound on $\{\alpha_n\}$.

It is interesting to note that the method of gradient projection is imbedded in a larger family of Hilbert space variable metric gradient projection methods in which at each step the projection operation and the determination of the gradient is carried out with a new inner product. Thus, if M_n is an operator satisfying (3.1a) then as an operator in the (GS) it can be assume that M_n is symmetric, i.e., $\langle M_n x, y \rangle = \langle M_n y, x \rangle$, $\forall x, y \in X$. This is true since $\langle M_n x, x \rangle = \langle (\frac{M_n + M_n^*}{2}) x, x \rangle$ where M_n^*

operator of M_n on X and $\frac{M_n + M_n^*}{2}$ is symmetric. Therefore, one can consider the sequence $\{M_n\}$ equivalent in the (GS) to $\{\frac{M_n + M_n^*}{2}\}$. Then, with M_n symmetric and positive definite a new inner product is defined by

$$\langle x, y\rangle_{M_n} = \langle M_n x, y\rangle.$$

Although the Fréchet derivative is the same for all of the related norms, the representation of f' changes with the inner product, since

$$f'(x)[y] = \langle \nabla f(x), y\rangle = \langle M_n M_n^{-1} \nabla f(x), y\rangle = \langle M_n^{-1} \nabla f(x), y\rangle_{M_n}.$$

The variable metric version of (3.3) is now

$$\hat{x}_n = P_{M_{n_\Omega}}(x_n - M_n^{-1} \nabla f(x_n))$$

or equivalently

$$\hat{x}_n = \arg\min\{\langle \nabla f(x_n), y - x_n\rangle + \tfrac{1}{2}\langle M_n(y - x_n), y - x_n\rangle\}.$$

An example of variable metric projection which is commonly practiced in computations in $\mathbb{R}^n$ is the technique of "scaling", in which the operators M_n are represented by diagonal matrices D_n. In one scheme, for example, entries on the diagonal are second partial derivatives, $\frac{\partial^2 f}{\partial x_i^2}$, of the functional f. Although such ad hoc methods can make matters worse, they can also accelerate convergence, and on simple sets such as orthants and boxes, the process is no more difficult to carry out than "standard" gradient projection. Notice that the condition number $\frac{\underline{\mu}_n}{\bar{\mu}_n}$ of a diagonal matrix is the ratio of smallest to largest diagonal entry; therefore, if that ratio is bounded away from zero for $n \geq 0$, then the scaling procedure satisfies (3.1).

The "unrelaxed" gradient projection methods considered by Levitin and Poljak [8] and Dunn [10] in which $\omega_n = 1$ for $n \geq 0$ can be considered as part of the (GS) provided δ is sufficiently small. In both cases the methods used to select the sequence $\{\alpha_n\}$ are such that at each step Goldstein's rule will select $\omega_n = 1$ if δ is small enough. For example, Levitin and Poljak require that α_n be chosen from the interval $[\varepsilon_1, \frac{2}{L + \varepsilon_2}]$ for $\varepsilon_1, \varepsilon_2 > 0$. From the definition of $g(x_n, \hat{x}_n, \omega)$ in Goldstein's rule (2.3) and from (2.17) one obtains

$$(3.7) \qquad g(x_n, \hat{x}_n, 1) \geq 1 - \frac{L\|x_n - \hat{x}_n\|^2}{2\langle \nabla f(x_n), x_n - \hat{x}_n \rangle}$$

when $\hat{x}_n$ is not an extremal. Also, $\hat{x}_n$ minimizes the functional $Q(\frac{1}{\alpha_n}I, x_n, \cdot)$ over Ω, and is therefore an extremal of $Q(\frac{1}{\alpha_n}I, x_n, \cdot)$ satisfying (2.1). In this case (2.1) reduces to

$$\langle \nabla Q(\frac{1}{\alpha_n}I, x_n, \hat{x}_n), z - \hat{x}_n \rangle \geq 0, \qquad \forall z \in \Omega,$$

or

$$(3.8) \qquad \langle \nabla f(x_n) + \frac{1}{\alpha_n}(\hat{x}_n - x_n), z - \hat{x}_n \rangle \geq 0, \qquad \forall z \in \Omega.$$

By letting $z = x_n$ in (3.8) one can write

$$(3.9)\qquad \langle \nabla f(x_n), x_n - \hat{x}_n \rangle \geq \frac{1}{\alpha_n}\|\hat{x}_n - x_n\|^2 .$$

If $\alpha_n \leq \frac{2}{L + \varepsilon_2}$, then (3.7) and (3.9) give

$$g(x_n, \hat{x}_n, 1) \geq 1 - \frac{L\alpha_n}{2}$$

$$\geq 1 - \frac{L}{L + \varepsilon_2} .$$

Thus, Goldstein's rule yields $\omega_n = 1$ if $\delta \leq 1 - \frac{L}{L + \varepsilon_2}$. Once again the lower bound $0 < \varepsilon_1 \leq \alpha_n$ gives the uniform upper bound required by the (GS).

The following theorem gives a "worst case" convergence rate estimate for methods in the (GS) when either (3.1) or (3.2) is satisfied. As noted above, a large number of well known methods are included in this subclass of the (GS).

Theorem 3.1. Let $\Omega \subset X$ where X is a Banach space and Ω is convex and bounded with diam $\Omega = D$. Let f be convex and differentiable with f' Lipschitz continuous on Ω, and let L be a Lipschitz constant for f'. Then $\inf_\Omega f > -\infty$ and if the (GS) operator sequence $\{M_n\}$ satisfies either condition (3.1) or (3.2) then the value $r_n = f(x_n) - \inf_\Omega f$ will decrease monotonically to zero and $r_n = O(\frac{1}{n})$.

Proof. As in Theorem 2.1, lines (2.24) and (2.25) one has

$$(3.10)\qquad f(x_n) - f(x_{n+1}) \geq \delta \min\{1, \frac{2\delta\langle f'(x_n), x_n - \hat{x}_n\rangle}{L\|x_n - \hat{x}_n\|^2}\}\langle f'(x_n), x_n - \hat{x}_n\rangle$$

Also, $\inf_{\Omega} f > -\infty$ follows easily from the Lipschitz continuity of f' and the boundedness of Ω. Therefore, for every $n \geq 0$ let z_n be such that $f(x_n) - f(z_n) \geq \frac{1}{2}(f(x_n) - \inf_{\Omega} f) = \frac{1}{2} r_n$. Then, as in Theorem 2.1, since f is convex one can write

$$(3.11) \qquad \langle f'(x_n), x_n - z_n \rangle \geq f(x_n) - f(z_n) \geq \frac{1}{2} r_n \geq 0,$$

and $z_n \in \Phi(x_n)$ for all $n \geq 0$. If (3.1) holds for $\{M_n\}$, and if x_n is not an extremal, then Lemma 2.2 gives

$$(3.12) \quad \langle f'(x_n), x_n - \hat{x}_n \rangle \geq \frac{1}{2}[\min\{\langle f'(x_n), x_n - z_n \rangle, \frac{\langle f'(x_n), x_n - z_n \rangle^2}{\langle M_n(x_n - z_n), x_n - z_n \rangle}\} + \langle M_n(x_n - \hat{x}_n), x_n - \hat{x}_n \rangle].$$

With (3.11), (3.12) becomes

$$(3.13) \qquad \langle f'(x_n), x_n - \hat{x}_n \rangle \geq \frac{1}{2}[\min\{\frac{r_n}{2}, \frac{r_n^2}{4\langle M_n(x_n - z_n), x_n - z_n \rangle}\} + \langle M_n(x_n - \hat{x}_n), x_n - \hat{x}_n \rangle],$$

and since all terms in (3.13) are positive it follows that both

$$(3.14a) \qquad \langle f'(x_n), x_n - \hat{x}_n \rangle \geq \frac{1}{2} \min\{\frac{r_n}{2}, \frac{r_n^2}{4\langle M_n(x_n - z_n), x_n - z_n \rangle}\}$$

and

$$(3.14b) \qquad \langle f'(x_n), x_n - \hat{x}_n \rangle \geq \frac{1}{2} \langle M_n(x_n, \hat{x}_n), x_n - \hat{x}_n \rangle.$$

Therefore, using appropriate combinations of (3.14a) and (3.14b) one has

$$(3.15)\qquad \frac{\langle f'(x_n),\, x_n - \hat{x}_n\rangle^2}{L\|x_n - \hat{x}_n\|^2}$$

$$\geq \frac{1}{4}\min\{\frac{r_n^2}{4L\|x_n - \hat{x}_n\|^2},\ \frac{r_n^2\langle M_n(x_n - \hat{x}_n),\, x_n - \hat{x}_n\rangle}{4L\|x_n - \hat{x}_n\|^2\langle M_n(x_n - z_n),\, x_n - z_n\rangle}\}$$

$$\geq \frac{1}{16L}\min\{\frac{r_n^2}{\|x_n - \hat{x}_n\|^2},\ \frac{r_n^2\underline{\mu}_n}{\bar{\mu}_n\|x_n - z_n\|^2}\}$$

$$\geq \frac{ar_n^2}{16LD^2} = c_1 r_n^2, \qquad c_1 > 0.$$

Line (3.14a) can be written as

$$(3.16)\qquad \langle f'(x_n),\, x_n - \hat{x}_n\rangle \geq \frac{1}{2}\min\{\frac{r_n}{2},\ \frac{r_n^2}{4KD^2}\},$$

and then (3.16), (3.15), and (3.10) yield

$$(3.17)\qquad f(x_n) - f(x_{n+1}) \geq \delta\min\{\frac{r_n}{4},\ \frac{r_n^2}{8KD^2},\ 2\delta c_1 r_n^2\}.$$

If (3.2) holds, then

$$(3.18)\qquad \langle f'(x_n),\, x_n - \hat{x}_n\rangle \geq \frac{r_n}{2}.$$

This is true since with Taylor's formula and (3.2) one obtains for any $y \in \Omega$

$$f(y) = f(x_n) + \langle f'(x_n), y - x_n \rangle$$

$$+ \int_0^1 \langle f''(x_n + \theta(y - x_n))(y - x_n), y - x_n \rangle (1 - \theta) d\theta$$

$$\geq f(x_n) + \langle f'(x_n), y - x_n \rangle + \frac{1}{2} \langle M_n(y - x_n), y - x_n \rangle$$

$$= f(x_n) + Q(M_n, x_n, y)$$

$$\geq f(x_n) + Q(M_n, x_n, \hat{x}_n).$$

With $y = z_n$, it follows that

$$-Q(M_n, x_n, \hat{x}_n) \geq f(x_n) - f(z_n) \geq \frac{r_n}{2}$$

or

$$\langle f'(x_n), x_n - \hat{x}_n \rangle - \frac{1}{2} \langle M_n(x_n - \hat{x}_n), x_n - \hat{x}_n \rangle \geq \frac{r_n}{2},$$

and since M_n is a nonnegative operator, (3.18) results. Combining (3.18) with (3.10) one has

$$(3.19) \qquad f(x_n) - f(x_{n+1}) \geq \delta \min\{\frac{r_n}{2}, \frac{\delta}{2LD^2} r_n^2\}$$

By Theorem 2.1, $\{r_n\}$ decreases monotonically to zero, and with (3.17) and (3.19), for r_n sufficiently small, it follows that

$$f(x_n) - f(x_{n+1}) \geq \delta c_2 r_n^2, \qquad \text{for some } c_2 > 0.$$

Therefore,

$$r_{n+1} \leq r_n - \delta c_2 r_n^2,$$

and Lemma 2.1 gives the rate estimate $r_n = O(\frac{1}{n})$.

QED

Remark 3.1. The proof of inequality (3.18) in the theorem is due to Allwright [15].

Dunn [10] has shown that sharper convergence rate upper bounds can be determined for the gradient projection method (1.5) in the presence of conditions on the growth rate of the functional f near an extremal ξ, i.e., condition (1.10). Condition (1.11), which expresses structural properties of the set Ω near ξ, implies condition (1.10), since when f is convex,

$$(3.20) \qquad \langle f'(\xi), x - \xi \rangle \leq f(x) - f(\xi)$$

is true for any minimizer ξ. For the conditional gradient method Dunn [4], [14] requires the condition (1.11) to establish higher rates of convergence.

In Theorem 3.2 it is shown that whenever it can be established that the stepsizes are bounded away from zero, the growth rate of the functional f near ξ, i.e., condition (1.10), is enough to give a hierarchy of linear and sublinear convergence rate estimates for the sequence $\{r_n\}$. When f' is Lipschitz continuous, condition (1.13) which requires that operators in the sequence $\{M_n\}$ be uniformly positive definite, is sufficient to prove that $\omega_n \geq \omega > 0$ since line (2.23) and Lemma 2.2 yield

$$\omega_n \geq \min\{1, \frac{2\delta\langle f'(x_n), x_n - \hat{x}_n\rangle}{L\|x_n - \hat{x}_n\|^2}\}$$

$$\geq \min\{1, \frac{\delta\langle M_n(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle}{L\|x_n - \hat{x}_n\|^2}\},$$

and with condition (1.13) it follows that

$$\omega_n \geq \min\{1, \frac{\delta\mu}{L}\} > 0.$$

Condition (1.13) is not required by Dunn [10] for the gradient projection method (1.5); however, the condition $\omega_n = 1$ is inherent in the method, and upper bounds for the sequence $\{\alpha_n\}$ in the gradient projection schemes in [8] and [9] are equivalent to condition (1.13).

When f is convex, condition (1.10) implies that ξ is a unique minimizer. It is possible, however, that the set Ω_f consists of more than one vector, in which case a more appropriate condition is

(3.21a) $$f(x) - \inf_{\Omega} f \geq \gamma d(x)^{\nu}, \qquad \forall x \in \Omega,$$

where

(3.21b) $$d(x) = \inf_{y \in \Omega_f} \|x - y\|.$$

Note that conditions (1.10), (1.11), and (3.21) require that Ω_f be nonempty.

Theorem 3.2. Let $\Omega \subset X$ where X is a Banach space and Ω is convex. Let f be convex and differentiable with f' Lipschitz continuous on Ω, and let L be a Lipschitz constant for f'. Let the (GS) be such that

(3.22) $\qquad \omega_n \geq \omega > 0, \qquad \forall n \geq 0,$

and let condition (3.21) hold with $\nu \in [2, \infty)$.

If $\nu = 2$ then $\{r_n\}$ converges linearly, i.e., $r_n = O(\lambda^n)$ for some $\lambda \in (0, 1)$. If $\nu > 2$, then

(3.23) $\qquad r_n = O(n^{-\nu/(\nu-2)}).$

<u>Proof</u>. From Goldstein's rule, (2.5), and (3.22) there results

(3.24)
$$f(x_n) - f(x_{n+1}) \geq \delta\omega_n \langle f'(x_n), x_n - \hat{x}_n \rangle$$
$$\geq \delta\omega \langle f'(x_n), x_n - \hat{x}_n \rangle .$$

As stated in Remark 2.2, for convex f and any $y \in \Omega_f$, one has

$$\langle f'(x_n), x_n - y \rangle \geq r_n \geq 0.$$

Therefore, by Lemma 2.2 and for any $y \in \Omega_f$, it follows that if $x_n \neq y$

(3.25)
$$\langle f'(x_n), x_n - x_n \rangle \geq \begin{cases} r_n, & \text{if } \langle M_n(x_n - y), x_n - y \rangle = 0 \\ \frac{1}{2}\min\{r_n, \dfrac{r_n^2}{\langle M_n(x_n - y), x_n - y \rangle}\}, & \text{if } \langle M_n(x_n - y), x_n - y \rangle > 0 \end{cases}$$

$$\geq \frac{1}{2}\min\{r_n, \frac{r_n^2}{K\|x_n - y\|^2}\}.$$

Let $\{y_n\} \in \Omega_f$ be such that for $n \geq 0$

$$\|x_n - y_n\| \leq 2d(x) = 2 \inf_{y \in \Omega_f} \|x_n - y\|.$$

Then for $\nu = 2$, condition (3.21) yields

$$(3.26) \qquad \frac{r_n^{\,2}}{K\|x_n - y_n\|^2} \geq \frac{r_n^{\,2}}{4Kd(x_n)^2} \geq \frac{\gamma}{4K} r_n,$$

and for $\nu > 2$, since $r_n^{\,2/\nu} \geq \gamma^{2/\nu} d(x_n)^2$ it follows that

$$(3.27) \qquad \frac{r_n^{\,2}}{K\|x_n - y_n\|^2} \geq \frac{\gamma^{2/\nu}}{4K} r_n^{\,(2-2/\nu)}.$$

Combining (3.24) - (3.26) one has with $\nu = 2$

$$f(x_n) - f(x_{n+1}) \geq \delta\omega r_n \min\{1, \tfrac{\gamma}{4K}\} = q r_n, \text{ with } q \in (0, 1).$$

Therefore,

$$r_{n+1} \leq (1 - q) r_n$$

which implies

$$r_n = O((1 - q)^n) = O(\lambda^n).$$

If $\nu > 2$, then (3.24), (3.25) and (3.27) yield

$$f(x_n) - f(x_{n+1}) \geq (\tfrac{\delta\omega}{2}) \min\{r_n, \frac{\gamma^{2/\nu}}{4K} r_n^{\,(2-2/\nu)}\}$$

which, for r_n sufficiently small, implies

$$r_{n+1} \leq r_n - qr_n^{(2-2/\nu)}, \text{ where } q = \frac{\delta\omega\gamma^{2/\nu}}{8K}.$$

The rate (2.23) follows from Lemma 2.2.

QED

As with condition (1.10), condition (1.11) implies that the minimizer ξ is unique; however, an extension to a condition like (3.21) is not possible here.

The following lemma, which is a modification of Lemma 2.5 in [22], will be needed in the next two theorems.

Lemma 3.1. Let $\Omega \subset X$ where X is a Banach space and Ω is convex. Let f be differentiable, f' Lipschitz continuous, and let L be a Lipschitz constant for f'. Let condition (1.11) hold at ξ with $\nu \in [1, \infty)$. If $\{x_n\}$ is generated by the (GS), then

(3.28) $$\|\hat{x}_n - \xi\|^{\nu-1} \leq \left(\frac{L + K}{\gamma}\right)\|x_n - \xi\| \qquad \text{for } n \geq 0,$$

where K is the uniform bound on the norms of the operators M_n (see (2.4)).

Proof. From (2.2) one has for $n \geq 0$

$$0 \leq Q(M_n, x_n, \xi) - Q(M_n, x_n, \hat{x}_n)$$

or

(3.29) $$0 \leq \langle f'(x_n), \xi - x_n \rangle + \frac{1}{2}\langle M_n(x_n - \xi), x_n - \xi \rangle - \langle f'(x_n), \hat{x}_n - x_n \rangle$$
$$- \frac{1}{2}\langle M_n(x_n - \hat{x}_n), x_n - \hat{x}_n \rangle.$$

From the positivity and the linearity of the operators M_n, there results

$$(3.30)\quad \tfrac{1}{2}\langle M_n(x_n - \xi), x_n - \xi\rangle - \tfrac{1}{2}\langle M_n(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle$$

$$= \tfrac{1}{2}\langle M_n(x_n - \xi), x_n - \xi\rangle - \tfrac{1}{2}\langle M_n(x_n - \hat{x}_n), x_n - \xi\rangle - \tfrac{1}{2}\langle M_n(x_n - \hat{x}_n), \xi - \hat{x}_n\rangle$$

$$= \tfrac{1}{2}\langle M_n(\hat{x}_n - \xi), x_n - \xi\rangle + \tfrac{1}{2}\langle M_n(x_n - \xi), \hat{x}_n - \xi\rangle - \tfrac{1}{2}\langle M_n(\hat{x}_n - \xi), \hat{x}_n - \xi\rangle.$$

Furthermore, (3.29), (3.30) and the Lipschitz continuity of f' yield

$$(3.31)\quad \tfrac{1}{2}\langle M_n(\hat{x}_n - \xi), \hat{x}_n - \xi\rangle \le \langle f'(x_n), \xi - \hat{x}_n\rangle + \tfrac{1}{2}\langle M_n(\hat{x}_n - \xi), x_n - \xi\rangle + \tfrac{1}{2}\langle M_n(\hat{x}_n - \xi), x_n - \xi\rangle$$

or

$$(3.32)\quad \tfrac{1}{2}\langle M_n(\hat{x}_n - \xi), \hat{x}_n - \xi\rangle + \langle f'(\xi), \hat{x}_n - \xi\rangle$$

$$\le \langle f'(\xi) - f'(x_n), \hat{x}_n - \xi\rangle + K\|x_n - \xi\|\,\|\hat{x}_n - \xi\|$$

$$\le (L + K)\|x_n - \xi\|\,\|\hat{x}_n - \xi\| .$$

Finally, by the positivity of M_n and condition (1.11) one has

$$\gamma\|\hat{x}_n - \xi\|^{\nu} \le (L + K)\|x_n - \xi\|\,\|\hat{x}_n - \xi\| .$$

QED

When $\{\omega_n\}$ can decrease to zero, condition (1.11) is used with Lemma 3.1 to estimate its rate of decrease, as shown in the following two theorems:

<u>Theorem 3.3</u>. Let $\Omega \subset X$ where X is a Banach space and Ω is convex and bounded. Let f be convex and differentiable with f' Lipschitz continuous on Ω, and let L be a Lipschitz constant for f'. Let the (GS) operator sequence $\{M_n\}$ satisfy either condition (3.1) or (3.2), and let condition (1.11) hold at ξ with $\nu \in [2, \infty)$. If $\nu = 2$, then $r_n = O(\lambda^n)$ for some $\lambda \in (0, 1)$, and if $\nu > 2$, then

$$(3.33) \qquad r_n = O(n^{\left(\frac{-\nu(\nu-1)}{\nu(\nu-1)-2}\right)}).$$

<u>Proof</u>. As in the proof of Theorem 3.1, line (3.10), one can write

$$(3.34) \quad f(x_n) - f(x_{n+1}) \geq \delta \min\{\langle f'(x_n), x_n - \hat{x}_n\rangle, \frac{2\delta\langle f'(x_n), x_n - \hat{x}_n\rangle^2}{L\|x_n - \hat{x}_n\|^2}\}.$$

Clearly, (3.11) is satisfied with $z_n = \xi$, and when (3.2) holds, (3.18) and (3.34) yield

$$(3.35) \qquad f(x_n) - f(x_{n+1}) \geq \frac{\delta}{2} \min\{r_n, \frac{\delta r_n^2}{L\|x_n - \hat{x}_n\|^2}\}.$$

From (3.14a) and (3.15) in Theorem 3.1, where (3.1) holds one has

$$(3.36) \qquad f(x_n) - f(x_{n+1})$$

$$\geq \frac{\delta}{2} \min\{\frac{1}{2} r_n, \frac{r_n^2}{2K\|x_n - \xi\|^2}, \frac{\delta r_n^2}{4L\|x_n - \hat{x}_n\|^2}, \frac{\delta a r_n^2}{4L\|x_n - \xi\|^2}\}.$$

When f is convex, (1.11) implies (1.10), and as in Theorem 3.2, it follows for $\nu \geq 2$

$$(3.37) \qquad \frac{r_n^{\,2}}{\|x_n - \xi\|^2} \geq \gamma^{2/\nu} r_n^{(2-2/\nu)}.$$

Using Lemma 3.1 with the triangle inequality one can write

$$\|x_n - \hat{x}_n\| \leq \|x_n - \xi\| + \|\xi - \hat{x}_n\|$$

$$\leq \|x_n - \xi\| + \left(\frac{L + K}{\gamma}\right)^{1/(\nu-1)} \|x_n - \xi\|^{1/(\nu-1)}$$

$$= \left(\|x_n - \xi\|^{(\nu-2)/(\nu-1)} + \left(\frac{L + K}{\gamma}\right)^{1/(\nu-1)}\right)\|x_n - \xi\|^{1/(\nu-1)}.$$

Therefore,

$$(3.38) \qquad \frac{r_n^{\,2}}{\|x_n - \hat{x}_n\|^2} \geq \frac{r_n^{\,2}}{\left(D^{(\nu-2)/(\nu-1)} + \left(\frac{L + K}{\gamma}\right)^{1/(\nu-1)}\right)^2 \|x_n - \xi\|^{2/(\nu-1)}}$$

$$\geq c_3 r_n^{2-2/(\nu(\nu-1))}$$

where

$$c_3 = \frac{\delta^{2/(\nu(\nu-1))}}{\left(D^{(\nu-2)/(\nu-1)} + \left(\frac{L + K}{\gamma}\right)^{1/(\nu-1)}\right)^2} \qquad \text{and } D = \text{diam } \Omega.$$

Since

$$\left(2 - \frac{2}{\nu}\right) \leq \left(2 - \frac{2}{\nu(\nu - 1)}\right) \qquad \text{for } \nu \geq 2,$$

it follows that

(3.39) $$r_n^{2-2/\nu} \geq r_n^{2-2/\nu(\nu-1)}$$

for r_n sufficiently small. By Theorem 2.1, $r_n \to 0$ and, therefore, (3.35) and (3.36) can be written as

$$r_{n+1} \leq r_n - qr_n^{(2-2/\nu(\nu-1))}, \; q > 0,$$

for r_n sufficiently small. The result (3.33) follows from Lemma 2.1 with $\nu > 2$, and when $\nu = 2$, $r_n = O(\lambda^n)$ for some $\lambda \in (0, 1)$.

QED

Up to this point the emphasis has been placed on determining convergence rates for the sequence $\{r_n\} = \{f(x_n) - \inf_{\Omega} f\}$. It is possible that the sequence of iterates $\{x_n\}$ has no limit points, and a rate on the sequence $\{r_n\}$ is the best one can do. Also, in most applications approximating the minimum value of the functional f is the primary objective. Note that conditions (1.10) and (1.11) give convergence rates for the sequence $\{\|x_n - \xi\|\}$ when a rate for $\{r_n\}$ is known. If $r_n \geq \gamma\|x - \xi\|^\nu$ for $\nu > 2$, then if $r_n = O(n^{-k})$, it follows that $\|x_n - \xi\| = O(n^{-k/\nu})$. Similarly, if $\nu = 2$, then linear convergence of $\{r_n\}$ implies linear convergence for $\{\|x_n - \xi\|\}$.

In the following lemma, it is shown that condition (1.10) for $\nu \in [1, 2)$ implies that condition (1.11) holds at ξ. (This is also shown in [10] and [4].) In Theorem 3.4 this fact is used to show superlinear convergence or finite termination for the sequence $\{\|x_n - \xi\|\}$ for <u>any</u> operator sequence $\{M_n\}$ in the (GS).

Lemma 3.2. Let $\Omega \subset X$ where X is a Banach space and Ω is convex and bounded. Let f be differentiable with f' Lipschitz continuous on Ω and for some $\xi \subset \Omega$ let f satisfy

$$(3.40) \qquad f(x) - f(\xi) \geq \gamma\|x - \xi\|^{\nu}, \qquad \forall x \in \Omega, \text{ and } \nu \in [1, 2).$$

Then

$$(3.41) \qquad \langle f'(\xi), x - \xi \rangle \geq \tilde{\gamma}\|x - \xi\|^{\nu}, \qquad \forall x \in \Omega, \text{ and some } \tilde{\gamma} > 0.$$

Proof. By Taylor's formula and the Lipschitz continuity of f (see line (2.22)),

$$(3.42) \qquad f(x) - f(\xi) \leq \langle f'(\xi), x - \xi \rangle + R\|x - \xi\|^2, \qquad \forall x \in \Omega,\ R < \infty,$$

and with (3.40) one has

$$\langle f'(\xi), x - \xi \rangle \geq \gamma\|x - \xi\|^{\nu} - R\|x - \xi\|^2$$

$$= (\gamma - R\|x - \xi\|^{2-\nu})\|x - \xi\|^{\nu}.$$

Therefore,

$$(3.43) \qquad \langle f'(\xi), x - \xi \rangle \geq \frac{\gamma}{2}\|x - \xi\|^{\nu}, \qquad \forall x \in \Omega \cap B_{\rho}(\xi),$$

where $B_{\rho}(\xi)$ is a closed ball of radius $\rho = (\frac{\gamma}{2R})^{1/(2-\nu)}$. Let $\rho < D = \sup_{y\in\Omega} \|y - \xi\| < \infty$, and let $\tilde{\gamma} = \frac{\gamma}{2}(\frac{\rho}{D})^{\nu-1}$. By the convexity of Ω it follows that for every $x \in \Omega - B_{\rho}(\xi)$ there exists a number $r \in (\rho, D]$ and a vector $y \in \Omega$ with $\|y - \xi\| = \rho$ such that $x - \xi = \frac{r}{\rho}(y - \xi)$. Therefore, (3.43) yields for $\nu \in [1, 2)$

$$\langle f'(\xi), x - \xi \rangle = \langle f'(\xi), \frac{r}{\rho}(y - \xi) \rangle$$

$$\geq \frac{\gamma r}{2\rho} \|y - \xi\|^{\nu}$$

$$= \frac{\gamma r \rho^{\nu}}{2\rho r^{\nu}} \|\frac{r}{\rho}(y - \xi)\|^{\nu}$$

$$\geq \frac{\gamma \rho^{\nu-1}}{2D^{\nu-1}} \|x - \xi\|^{\nu}$$

$$= \tilde{\gamma} \|x - \xi\|^{\nu}.$$

Since $\frac{\gamma}{2} > \tilde{\gamma}$ the lemma is proved for all $x \in \Omega$.

QED

Theorem 3.4. Let $\Omega \subset X$ where X is a Banach space and Ω is convex and bounded. Let f be convex and differentiable with f' Lipschitz continuous and let L be a Lipschitz constant for f'. If for $\xi \in \Omega_f$,

(3.44) $\qquad f(x) - f(\xi) \geq \gamma \|x - \xi\|^{\nu}, \qquad$ for some $\nu \in [1, 2)$,

and if $\{x_n\}$ is a sequence generated by the (GS), then $\{\|x_n - \xi\|\}$ converges to zero superlinearly, i.e., either $x_n = \xi$ for some $n \geq 0$ or

(3.45) $$\lim_{n\to\infty} \frac{\|x_{n+1} - \xi\|}{\|x_n - \xi\|} = 0.$$

If $\nu = 1$ in (3.44), then $x_N = \xi$ for some $N < \infty$.

Proof. By Theorem 2.1, $r_n \to 0$ and, therefore, $x_n \to \xi$ follows from (3.43). Line (2.23) gives

$$(3.46) \qquad \omega_n \geq \min\{1, \frac{2\delta\langle f'(x_n), x_n - \hat{x}_n\rangle}{L\|x_n - \hat{x}_n\|^2}\},$$

and, as in line (3.14a) with $z_n = \xi$, one can write

$$(3.47) \qquad \langle f'(x_n), x_n - \hat{x}_n\rangle \geq \frac{1}{4}\min\{r_n, \frac{r_n^{\ 2}}{2K\|x_n - \xi\|^2}\}.$$

Therefore, (3.46) and (3.47) give

$$(3.48) \qquad \omega_n \geq \min\{1, \frac{\delta r_n}{2L\|x_n - \hat{x}_n\|^2}, \frac{\delta r_n^{\ 2}}{4LK\|x_n - \hat{x}_n\|^2\|x_n - \xi\|^2}\}.$$

By Lemma 3.2 and (3.43) one obtains

$$\langle f'(\xi), x - \xi\rangle \geq \tilde{\gamma}\|x - \xi\|^\nu, \qquad \forall x \in \Omega, \ \tilde{\gamma} > 0,$$

and Lemma 3.1 yields

$$(3.49) \qquad \|\hat{x}_n - \xi\|^{\nu-1} \leq (\frac{L + K}{\tilde{\gamma}})\|x_n - \xi\|, \qquad \text{for } n \geq 0.$$

Therefore, for $\nu \in (1, 2)$

$$\|x_n - \hat{x}_n\| \leq (\|x_n - \xi\| + \|\hat{x}_n - \xi\|)$$

$$\leq (\|x_n - \xi\| + (\frac{L + K}{\tilde{\gamma}})^{1/(\nu-1)} \|x_n - \xi\|^{1/(\nu-1)})$$

$$= (1 + (\frac{L + K}{\tilde{\gamma}})^{1/(\nu-1)} \|x_n - \xi\|^{(2-\nu)/(\nu-1)}) \|x_n - \xi\|$$

$$\leq (1 + (\frac{L + K}{\tilde{\gamma}})^{1/(\nu-1)} D^{(2-\nu)/(\nu-1)}) \|x_n - \xi\|$$

$$= c_4 \|x_n - \xi\|.$$

The lower bound (3.48) now becomes

$$\omega_n \geq \min\{1, \frac{\delta r_n}{2Lc_4{}^2\|x_n - \xi\|^2}, \frac{\delta r_n{}^2}{4LKc_4{}^2\|x_n - \xi\|^4}\},$$

and with (3.44) there results

$$\omega_n \geq \min\{1, \frac{\delta\gamma}{2Lc_4\|x_n - \xi\|^{2-\nu}}, \frac{\delta\gamma^2}{4LKc_4{}^2\|x_n - \xi\|^{2(2-\nu)}}\}.$$

Since $\|x_n - \xi\| \to 0$, it follows that $\omega_n = 1$ for $n \geq N_1$ for some $N_1 > 0$. Thus, for $n \geq N_1$, (3.49) yields

$$\frac{\|x_{n+1} - \xi\|}{\|x_n - \xi\|} \leq (\frac{L + K}{\tilde{\gamma}})^{1/\nu-1} \|x_n - \xi\|^{(2-\nu)/(\nu-1)},$$

which implies (3.45). When $\nu = 1$, the finite termination of the process follows directly from (3.49), since

$$\tilde{\gamma}\|\hat{x}_n - \xi\| \le (L + K)\|x_n - \xi\|\,\|\hat{x}_n - \xi\|$$

is true for $n \ge 0$, and if $\|x_n - \xi\| < \frac{\tilde{\gamma}}{(L + K)}$, then $\|x_n - \xi\| = 0$.

QED

Remark 3.2. Dunn [4], [17] has shown that conditions (1.11) and (1.10) can be established for certain extremals found in problems from optimal control theory. Let U be a nonempty convex set in $\mathbb{R}^m$, and let the constraint set Ω be the set of functions

$$\Omega = \{\text{measurable } u(\cdot):[0, 1] \to U\}.$$

Here f, the functional to be minimized, is defined and differentiable on a neighborhood of Ω in one of the spaces $L^p([0, 1], \mathbb{R}^m)$ and problem (P) becomes

$$\min_{u(\cdot)\in\Omega} f(u(\cdot)) .$$

Condition (1.11) at an extremal function $\xi(\cdot)$ becomes

$$(3.50) \qquad \langle f'(\xi(\cdot)), u(\cdot) - \xi(\cdot)\rangle \ge \gamma\|u(\cdot) - \xi(\cdot)\|_p^\nu ,$$

where, in this case,

$$\langle f'(\xi(\cdot)), u(\cdot) - \xi(\cdot)\rangle = \int_0^1 \gamma(t)\cdot(u(t) - \xi(t))dt$$

with $\gamma(\cdot)$ the (unique) representor of the Fréchet derivative of $f'(\xi(\cdot))$ in the conjugate space $L^q([0, 1], \mathbb{R}^m)$ with $q = p/(p - 1)$ and

$$\|u(\cdot) - \xi(\cdot)\|_p = (\int_0^1 \|u(t) - \xi(t)\|^p dt)^{1/p} .$$

In [17] it is shown that if U is a bounded polyhedron in $\mathbb{R}^m$ and if $\xi(\cdot)$, is a certain type of "bang-bang" extremal, then values for ν can be calculated directly from such factors as the number p and the growth properties of a scalar "switching function" s(t) on [0, 1] determined by $\gamma(t)$ and Ω. For example, if f is convex, U is $[-1, 1] \subset \mathbb{R}^1$, and $\gamma(t)$ is continuous, nondecreasing, and has an isolated zero at $t = \frac{1}{2}$, then

$$(3.51) \qquad \xi(t) \in \begin{cases} -1, & t \in [0, \frac{1}{2}) \\ [-1, 1], & t = \frac{1}{2} \\ +1, & t \in (\frac{1}{2}, 1], \end{cases}$$

and ν in (3.50) has the value 2 and 4 in L^1 and L^2 respectively. Thus by Theorem 3.2 the gradient projection method, which is limited to the Hilbert space L^2, would generate a sequence of iterates whose convergence rate estimate is $f(u_n(\cdot)) - f(\xi(\cdot)) = O(\frac{1}{n^2})$; a simple example with minimizer (3.51) shows that this estimate cannot be improved. On the other hand, the conditional gradient method makes sense in L^1, and for $\nu = 2$ this method converges linearly according to Theorem 3.3. Note that the L^1 analog of Hilbert space gradient projection in which $M_n = \frac{1}{\alpha_n} I$ in the (GS) is a method in which $Q_n(y) = \langle f'(x_n), y - x_n \rangle + \frac{1}{2\alpha_n} \|y - x_n\|_1^2$ is minimized at each step. This method is not formally in the (GS); however, Lemma 2.2 and Theorems 2.1, 3.1, and 3.2 could be modified to give the same results for this method with $0 < a \leq \alpha_n \leq b < \infty$, $\forall n \geq 0$, and in the example (3.51) above the convergence rate in L^1 would be at least linear. Its feasibility in L^1, however, is questionable.

Remark 3.3. The results of this chapter are readily extended to pseudo-convex functionals satisfying (2.32) in Remark 2.4, when Ω_f is nonempty.

4. Newton's Method

When f is convex, f'' is nonnegative and symmetric (see, e.g., [22]), and Newton's method, in which $M_n = f''(x_n)$, is in the (GS). The "worst case" convergence rate of $r_n = O(n^{-1/3})$ given by Theorem 2.1 when $\|f''(x)\| \leq K$, $\forall x \in \Omega$, seems far too conservative, since in the regular cases for which convergence rate estimates exist, the rates for Newton's method are clearly superior to those of such first order schemes as the gradient projection method and the conditional gradient method. On the other hand, $f''(x_n)$ need not satisfy either condition (3.1) or (3.2), and so Theorem 3.1, which gives the rate $O(\frac{1}{n})$, does not necessarily apply to Newton's method. The fact that $M_n = f''(x_n)$ in Newton's method, however, can be employed to strengthen a number of the fundamental inequalities used in previous theorems, and convergence rate estimates for Newton's method will be shown to be better than those of the first order methods in a number of less than regular cases.

The following two lemmas improve inequalities (2.23) and (3.28) when f'' is Lipschitz continuous on Ω, i.e., when there exists an $L > 0$ such that $\|f''(x) - f''(y)\| \leq L\|x - y\|$, $\forall x, y \in \Omega$.

Lemma 4.1. Let $\Omega \subset X$ be convex, where X is a Banach space. Let f be convex and twice differentiable, f'' Lipschitz continuous on Ω with L a Lipschitz constant for f'' on Ω, and let f'' satisfy $\|f''(x)\| \leq K < \infty$ $\forall x \in \Omega$. If $\{x_n\}$ is a sequence generated by the (GS) with $M_n = f''(x_n)$ then for any y belonging to the set $\Phi(x_n)$ in (2.9),

$$(4.1) \qquad \omega_n \geq \begin{cases} \min\{\delta, \left[\dfrac{\delta \min\{\langle f'(x_n), x_n - y\rangle, \dfrac{\langle f'(x_n), x_n-y\rangle^2}{\langle f''(x_n)(x_n-y), x_n-y\rangle}}{L\|x_n-\hat{x}_n\|^3}\right]^{\frac{1}{2}}\}, \\ \qquad \text{if } \langle f''(x_n)(x_n-y), x_n-y\rangle > 0 \\ \min\{\delta, \left(\dfrac{2\delta\langle f'(x_n), x_n-y\rangle}{L\|x_n-\hat{x}_n\|^3}\right)^{\frac{1}{2}}\}, \\ \qquad \text{if } \langle f''(x_n)(x_n-y), x_n-y\rangle = 0. \end{cases}$$

Proof. If $\omega_n > 1$ and x_n is not an extremal, then Goldstein's rule and Taylor's formula yield

$$(4.2) \qquad 1 - \delta \geq \frac{f(x_n) - f(x_n + \omega_n(\hat{x}_n - x_n))}{\omega_n\langle f'(x_n), x_n - \hat{x}_n\rangle}$$

$$= \frac{\omega_n\langle f'(x_n), x_n - \hat{x}_n\rangle - \frac{\omega_n^2}{2}\langle f''(\zeta_n)(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle}{\omega_n\langle f'(x_n), x_n - \hat{x}_n\rangle},$$

where

$$\zeta_n = x_n + \theta_n(x_n + \omega_n(\hat{x}_n - x_n) - x_n)$$

$$= x_n + \theta_n\omega_n(\hat{x}_n - x_n) \qquad \text{for some } \theta_n \in [0, 1].$$

It follows from the Lipschitz continuity of f" that

$$(4.3) \qquad 1 - \delta \geq 1 - \frac{\frac{\omega_n}{2}\langle f''(x_n)(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle}{\langle f'(x_n), x_n - \hat{x}_n\rangle}$$

$$+ \frac{\frac{\omega_n}{2}\langle (f''(x_n) - f''(\zeta_n))(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle}{\langle f'(x_n), x_n - \hat{x}_n\rangle}$$

$$\geq 1 - \omega_n \frac{\langle f''(x_n)(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle + L\omega_n\theta_n\|x_n - \hat{x}_n\|^3}{2\langle f'(x_n), x_n - \hat{x}_n\rangle},$$

and one obtains

$$(4.4) \qquad \omega_n \geq \frac{2\delta\langle f'(x_n), x_n - \hat{x}_n\rangle}{\langle f''(x_n)(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle + L\omega_n\theta_n\|x_n - \hat{x}_n\|^3}.$$

Let $y \in \Phi(x_n)$ and assume $\langle f''(x_n)(x_n - y), x_n - y\rangle > 0$. Then Lemma 2.2 and (4.4) yield

$$(4.5) \quad \omega_n \geq \frac{\delta \min\{\langle f'(x_n), x_n - y\rangle, \frac{\langle f'(x_n), x_n - y\rangle^2}{\langle f''(x_n)(x_n - y), x_n - y\rangle}\} + \delta\langle f''(x_n)(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle}{\langle f''(x_n)(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle + L\omega_n\theta_n\|x_n - \hat{x}_n\|^3}.$$

Suppose $\omega_n < \delta$. Then from (4.5) one has

$$(4.6) \qquad \delta\langle f''(x_n)(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle + \omega_n^2 L\|x_n - \hat{x}_n\|^3$$

$$\geq \omega_n\langle f''(x_n)(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle + \omega_n^2 L\theta_n\|x_n - \hat{x}_n\|^3$$

$$\geq \delta \min \left\{\langle f'(x_n), x_n - y\rangle, \frac{\langle f'(x_n), x_n - y\rangle^2}{\langle f''(x_n)(x_n - y), x_n - y\rangle}\right\}$$

$$+ \delta\langle f''(x_n)(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle$$

or

$$(4.7) \qquad \omega_n^2 \geq \frac{\delta \min\left\{\langle f'(x_n), x_n - y\rangle, \frac{\langle f'(x_n), x_n - y\rangle^2}{\langle f''(x_n)(x_n - y), x_n - y\rangle}\right\}}{L\|x_n - \hat{x}_n\|^3}.$$

If $\langle f''(x_n)(x_n - y), x_n - y\rangle = 0$, then using Lemma 2.2 and the above argument one obtains

$$(4.8) \qquad \omega_n^2 \geq \frac{2\delta\langle f'(x_n), x_n - y\rangle}{L\|x_n - \hat{x}_n\|^3},$$

and the result (4.1) follows from (4.7) and (4.8).

QED

<u>Lemma 4.2</u>. Let $\Omega \subset X$ be convex, where X is a Banach space. Let f be convex and twice differentiable, f'' Lipschitz continuous on Ω with L a Lipschitz constant for f'' on Ω, and let f'' satisfy $\|f''(x)\| \leq K < \infty$, $\forall x \in \Omega$. Let ξ be the unique minimizer of f in Ω and suppose (1.11) holds, i.e.,

$$\langle f'(\xi), x - \xi \rangle \geq \gamma \| x - \xi \|^\nu, \qquad \forall x \in \Omega,\ \nu \in [1, \infty).$$

If $\{x_n\}$ is a sequence generated by the (GS) with $M_n = f''(x_n)$ then

$$(4.9) \qquad \|\hat{x}_n - \xi\|^{\nu-1} \leq (\tfrac{L}{\gamma})\|x_n - \xi\|^2 .$$

<u>Proof</u>. From (3.31) in Lemma (3.1) one has with $M_n = f''(x_n)$

$$(4.10) \qquad \tfrac{1}{2}\langle f''(x_n)(\hat{x}_n - \xi), \hat{x}_n - \xi \rangle + \langle f'(\xi), \hat{x}_n - \xi \rangle$$

$$\leq \langle f'(\xi) - f'(x_n), \hat{x}_n - \xi \rangle + \tfrac{1}{2}\langle f''(x_n)(x_n - \xi), \hat{x}_n - \xi \rangle$$

$$+ \tfrac{1}{2}\langle f''(x_n)(\hat{x}_n - \xi), x_n - \xi \rangle .$$

Since the second derivative operator is symmetric and nonnegative, one obtains with the Mean Value Theorem

$$(4.11) \qquad \langle f'(\xi), \hat{x}_n - \xi \rangle \leq \langle f''(\zeta_n)(\xi - x_n), \hat{x}_n - \xi \rangle$$

$$+ \langle f''(x_n)(x_n - \xi), \hat{x}_n - \xi \rangle ,$$

where $\zeta_n = x_n + \theta_n(\xi - x_n)$ for some $\theta_n \in [0, 1]$. From (1.11) and the Lipschitz continuity of f'' there results

$$\gamma\|\hat{x} - \xi\|^\nu \leq L\|x_n - \zeta_n\| \, \|x_n - \xi\| \, \|\hat{x}_n + \xi\|$$

$$\leq L\|x_n - \xi\|^2 \, \|\hat{x}_n - \xi\| .$$

QED

Using the results of Lemma 4.1 in the proof of Theorem 2.1 one can easily obtain a new "worst case" rate of convergence of $r_n = 0(n^{-1/2})$ for Newton's method when f is convex. Lemma 4.2 can be used to prove that a hierarchy of convergence rate estimates exists as in Theorem 3.3 when condition (1.11) holds at a minimizer; however, the estimates would still be worse than those in Theorem 3.3. The following assumption on the functional f near a minimizer ξ will give a hierarchy of convergence rate estimates superior to those in Theorem 3.3, and Lemma 4.3 will prove that the assumption is actually true for a large class of convex functionals.

Assumption (A). If f is convex and ξ is a minimizer of f on Ω then for some $\rho > 0$, some $c > 0$ and all $x \in \Omega \cap B_\rho(\xi)$

$$\langle f'(x), x - \xi\rangle^2 \geq c(f(x) - f(\xi))\langle f''(x)(x - \xi), x - \xi\rangle.$$

Although (A) is not true for all convex functionals, as an example of Dunn [14] in the Hilbert space ℓ_2 shows, it is conjectured that (A) is true whenever condition (1.10) holds at a unique minimizer of f in Ω, i.e., when the functional near the extremal grows like $\|x - \xi\|^\nu$ for some $\nu \in [1, \infty)$. The following lemma supports this conjecture.

Lemma 4.3. Let $\Omega \subset X$ be convex where X is a Banach space. Let f be convex, five times differentiable with $f^{(5)}$ Lipschitz continuous on Ω with L a Lipschitz constant for $f^{(5)}$ on Ω, and suppose for some $\xi \in \Omega$

$$f(x) - f(\xi) \geq \gamma\|x - \xi\|^5, \qquad \forall x \in \Omega.$$

Then f satisfies assumption (A) at ξ.

Proof. The proof can be found after the proof of Theorem 4.1.

Remark 4.1. Similar proofs can be given for (i)-differentiable convex functionals with Lipschitz continuous $i^{\underline{th}}$ derivatives when $f(x) - f(\xi) \geq \gamma\|x - \xi\|^i$ for $i = 3, 4, 5$.

Theorem 4.1. Let $\Omega \subset X$ be convex and bounded where X is a Banach space. Let f be convex and at least twice differentiable with f" Lipschitz continuous on Ω with L a Lipschitz constant for f" on Ω, and let f" satisfy $\|f''(x)\| \leq K < \infty$, $\forall x \in \Omega$. Furthermore, suppose that assumption (A) holds at a unique minimizer $\xi \in \Omega$. Finally, let $\{x_n\}$ be a sequence generated by the (GS) with $M_n = f''(x_n)$. Then:

(i) If (1.10) holds for some $\nu \in [2, \infty)$, the values $r_n = f(x) - f(\xi)$ satisfy $r_n = O(\frac{1}{n^2})$ (at least).

(ii) If (1.11) holds for $\nu \in (3, \infty)$, then $r_n = O(n^{-2\nu(\nu-1)/(\nu(\nu-1)-6)})$.

(iii) If (1.11) holds for $\nu = 3$, then $r_n = O(\lambda^n)$ for some $\lambda \in (0, 1)$.

(iv) If (1.11) holds for $\nu \in [1, 3)$, then the sequence $\{\|x_n - \xi\|\}$ converges superlinearly, i.e., either $x_n = \xi$ beyond some N, or else

$$\lim_{n\to\infty} \frac{\|x_{n+1} - \xi\|}{\|x_n - \xi\|} = 0.$$

Proof. In all cases $r_n \to 0$ by Theorem 2.1, and since condition (1.10) or (1.11) is satisfied here, it follows that $\|x_n - \xi\| \to 0$. It can be assumed, therefore, that for some $c > 0$ the inequality

(4.12) $\langle f'(x_n), x_n - \xi\rangle^2 \geq cr_n\langle f''(x_n)(x_n - \xi), x_n - \xi\rangle$

is true uniformly in $n \geq 0$.

(i) From Lemma 2.2, Goldstein's rule (2.5), (4.12) and the convexity of f one has for $n \geq 0$

(4.13) $$f(x_n) - f(x_{n+1}) \geq \delta\omega_n\langle f'(x_n), x_n - \hat{x}_n\rangle$$

$$\geq \delta\omega_n c_1 r_n, \quad \text{where } c_1 = \frac{1}{2}\min\{1, c\}.$$

Condition (4.12) and Lemma 4.1 with $y = \xi$, $\forall n \geq 0$, now yield

(4.14) $$\omega_n \geq \min\{\delta, (\frac{2\delta c_1 r_n}{L\|x_n - \hat{x}_n\|^3})^{1/2}\}$$

$$\geq \min\{\delta, (\frac{2\delta c_1 r_n}{LD^3})^{1/2}\}, \quad \text{where } D = \text{diam } \Omega.$$

For n sufficiently large, (4.13) and (4.14) give

$$r_{n+1} \leq r_n - qr_n^{3/2}$$

with

$$q = (\frac{2}{LD^3})^{1/2}(\delta c_1)^{3/2}.$$

This implies $r_n = O(\frac{1}{n^2})$, by Lemma 2.1.

(ii) By using (1.11) with $\nu \in (3, \infty)$ in Lemma 4.2 one obtains

$$(4.15) \qquad \|\hat{x}_n - \xi\| \le (\frac{L}{\gamma})^{1/(\nu-1)}\|x_n - \xi\|^{2/(\nu-1)} \text{ for } n \ge 0.$$

Therefore, by the triangle inequality one obtains, as in Theorem 3.3,

$$(4.16) \qquad \frac{r_n}{\|x_n - \hat{x}_n\|^3} \ge \frac{r_n}{(\|x_n - \xi\| + \|\hat{x}_n - \xi\|)^3}$$

$$\ge \frac{r_n}{(\|x_n - \xi\| + (\frac{L}{\gamma})^{1/(\nu-1)}\|x_n - \xi\|^{2/(\nu-1)})^3}$$

$$\ge \frac{r_n}{(D^{(\nu-3)/(\nu-1)} + (\frac{L}{\gamma})^{1/(\nu-1)})^3\|x_n - \xi\|^{6/(\nu-1)}},$$

provided $x_n \ne \xi$. Also, since f is convex, (1.11) implies (1.10) and one can write

$$(4.17) \qquad r_n \ge r_n^{1-6/(\nu(\nu-1))}\gamma^{6/(\nu(\nu-1))}\|x_n - \xi\|^{6/(\nu-1)}.$$

The inequalities (4.16), (4.17), and (4.14) now give

$$(4.18) \qquad \omega_n \ge \min\{\delta, c_2 r_n^{(\nu(\nu-1)-6)/(2\nu(\nu-1))}\},$$

where

$$c_2 = (\frac{2\delta c_1 \nu^{6/(\nu(\nu-1))}}{L(D^{(\nu-3)/\nu-1)} + (\frac{L}{\gamma})^{1/(\nu-1)})^3})^{1/2} > 0.$$

Finally, (4.18) and (4.13) yield for n sufficiently large

$$r_{n+1} \leq r_n - q r_n^{1+(\nu(\nu-1)-6)/(2\nu(\nu-1))}$$

with

$$q = \delta c_2 c_1.$$

The desired result now follows from Lemma 2.1.

(iii) From (4.13) it is easy to see that if $\omega_n \geq \omega > 0$, $\forall n \geq 0$, then $r_n = O(\lambda^n)$ with $\lambda = \max\{0, 1 - \delta c_1 \omega\}$. When (1.11) holds with $\nu = 3$, it follows from Lemma 4.2 that

$$(4.19) \qquad \|\hat{x}_n - \xi\| \leq (\frac{L}{\gamma})^{1/2} \|x_n - \xi\|,$$

and using the triangle inequality one finds that

$$(4.20) \qquad \|x_n - \hat{x}_n\|^3 \leq (\|x_n - \xi\| + \|\hat{x}_n - \xi\|)^3 \leq (1 + (\frac{L}{\gamma})^{1/2})^3 \|x_n - \xi\|^3.$$

Since $r_n \geq \langle f'(\xi), x_n - \xi \rangle \geq \gamma \|x_n - \xi\|^3$, (4.20) and (4.14) combine to yield

$$\omega_n \geq \min\{\delta, (\frac{2\delta c_1 \gamma}{L(1 + (\frac{L}{\gamma})^{1/2})^3})^{1/2}\} = \omega > 0.$$

(iv) When (1.11) holds with $\nu \in (1, 3)$, then Lemma 4.2 states that if $x_n \neq \xi$, then

$$(4.21) \qquad \frac{\|\hat{x}_n - \xi\|}{\|x_n - \xi\|} \leq (\frac{L}{\gamma})^{1/(\nu-1)} \|x_n - \xi\|^{(3-\nu)/(\nu-1)}.$$

It will now be shown that $\omega_n = 1$ (and consequently $x_{n+1} = \hat{x}_n$) for n sufficiently large; this result, together with (4.21) implies that x_n converges superlinearly to ξ.

Goldstein's rule selects $\omega_n = 1$ if $g(x_n, \hat{x}_n, 1) \geq \delta$ for the given $\delta \in (0, \frac{1}{2})$. From the definition of $g(x, \hat{x}, \omega)$ and Taylor's formula one has

$$(4.22) \qquad g(x_n, \hat{x}_n, 1) = \frac{f(x_n) - f(\hat{x}_n)}{\langle f'(x_n), x_n - \hat{x}_n \rangle}$$

$$= 1 - \frac{\frac{1}{2}\langle f''(x_n + \theta_n(\hat{x}_n - x_n))(\hat{x}_n - x_n), \hat{x}_n - x_n \rangle}{\langle f'(x_n), x_n - \hat{x}_n \rangle}$$

for some $\theta_n \in [0, 1]$,

provided $x_n \neq \xi$. Since f" is Lipschitz continuous, it follows that

$$(4.23) \qquad g(x_n, \hat{x}_n, 1) \geq 1 - \frac{\langle f''(x_n)(x_n - \hat{x}_n), x_n - \hat{x}_n \rangle}{2\langle f'(x_n), x_n - \hat{x}_n \rangle}$$

$$- \frac{L\|x_n - \hat{x}_n\|^3}{2\langle f'(x_n), x_n - \hat{x}_n \rangle} .$$

For any operator M and each fixed $x \in \Omega$, $Q(M, x, \cdot)$ is a functional on Ω and if $\hat{x}$ minimizes $Q(M, x, \cdot)$ on Ω then $\hat{x}$ is an extremal of $Q(M, x, \cdot)$. If M is symmetric, then (2.1) yields

$$\langle (f'(x) + M)(\hat{x} - x), z - \hat{x} \rangle \geq 0 , \qquad \forall z \in \Omega.$$

In particular, for $z = x$ this gives

$$(4.24) \qquad \langle f'(x), x - \hat{x} \rangle \geq \langle M(x - \hat{x}), x - \hat{x} \rangle .$$

Since $f''(x_n)$ is symmetric and nonnegative one can use (4.24) in (4.23) to get

$$(4.25) \qquad g(x_n, \hat{x}_n, 1) \geq \frac{1}{2} - \frac{L\|x_n - \hat{x}_n\|^3}{\langle f'(x_n), x_n - \hat{x}_n \rangle} .$$

From Lemma 4.2 with assumption (A) one has

$$(4.26) \qquad \langle f'(x_n), x_n - \hat{x}_n \rangle \geq c_1 r_n \geq c_1 \gamma \|x_n - \xi\|^\nu, \text{ where } c_1 = \frac{1}{2}\min\{1, c\}.$$

Furthermore, Lemma 4.2 and the triangle inequality yield

$$(4.27) \qquad \begin{aligned} \|x_n - \hat{x}_n\|^3 &\leq (\|x_n - \xi\| + \|\hat{x}_n - \xi\|)^3 \\ &\leq (\|x_n - \xi\| + (\tfrac{L}{\gamma})^{1/(\nu-1)}\|x_n - \xi\|^{2/(\nu-1)})^3 \\ &\leq (1 + (\tfrac{L}{\gamma})^{1/(\nu-1)} D^{(3-\nu)/(\nu-1)})^3 \|x_n - \xi\|^3 . \end{aligned}$$

Together, (4.27), (4.26), and (4.25) give

$$g(x_n, \hat{x}_n, 1) \geq \frac{1}{2} - \frac{L(1 + (\frac{L}{\gamma})^{1/(\nu-1)} D^{(3-\nu)(\nu-1)})^3}{c_1 \gamma} \|x_n - \xi\|^{3-\nu},$$

and since $\|x_n - \xi\| \to 0$ one can conclude that for any fixed $\delta \in (0, \frac{1}{2})$, there is a sufficiently large $N(\delta)$ such that

$$g(x_n, \hat{x}_n, 1) \geq \delta \qquad \text{for } n \geq N(\delta).$$

If (1.11) holds with $\nu = 1$ then $x_n = \xi$ for $n \geq N$ for some $N < \infty$ by Theorem 3.4.

QED

Proof of Lemma 4.3. Let $\bar{B}_\rho(\xi) = \{x \in B_\rho(\xi): \langle f''(x)(x - \xi), x - \xi\rangle > 0\}$. Then since $\langle f'(x), x - \xi\rangle \geq f(x) - f(\xi)$ when f is convex, it is sufficient to show that

$$\frac{\langle f'(x), x - \xi\rangle}{\langle f''(x)(x - \xi), x - \xi\rangle} \geq c > 0 \qquad \forall x \in \Omega \cap \bar{B}_\rho(\xi) \quad \text{for some } \rho > 0 \tag{4.28}$$

One can expand $f(x) - f(\xi)$, $\langle f'(x), x - \xi\rangle$ and $\langle f''(x)(x - \xi), x - \xi\rangle$ with Taylor's formula and use the Lipschitz continuity of $f^{(5)}$ to obtain

$$f(x) - f(\xi) = \langle f'(\xi), x - \xi\rangle + \sum_{n=2}^{5} \frac{1}{n!}\langle f^{(n)}(\xi)(x - \xi)^{n-1}, x - \xi\rangle + r_1(\xi, x), \tag{4.29}$$

$$\langle f'(x), x - \xi\rangle = \langle f'(\xi), x - \xi\rangle + \sum_{n=2}^{5} \frac{1}{(n-1)!}\langle f^{(n)}(\xi)(x - \xi)^{n-1}, x - \xi\rangle + r_2(\xi, x), \tag{4.30}$$

$$(4.31) \qquad \langle f''(x)(x - \xi), x - \xi \rangle = \sum_{n=2}^{5} \frac{1}{(n-2)!} \langle f^{(n)}(\xi)(x - \xi)^{n-1}, x - \xi \rangle$$

$$+ r_3(\xi, x),$$

where $f^{(k)}(\xi)$ is the $k^{\underline{th}}$ differential of f at ξ (for a discussion of higher order derivatives and Taylor's formula in Banach spaces, see, e.g., Vainberg [27]). The terms $r_1(\xi, x)$, $r_2(\xi, x)$ and $r_3(\xi, x)$ satisfy

$$(4.32) \quad |r_1(\xi, x)| = \left|\frac{1}{5!} \langle (f^{(5)}(\xi + \theta_1(x - \xi)) - f^{(5)}(\xi))(x - \xi)^4, x - \xi \rangle\right|$$

$$\leq \frac{L}{5!} \|x - \xi\|^6,$$

$$(4.33) \qquad |r_2(\xi, x)| \leq \frac{L}{4!} \|x - \xi\|^6,$$

$$(4.34) \qquad |r_3(\xi, x)| \leq \frac{L}{3!} \|x - \xi\|^6.$$

Each $x \in \Omega$ can be expressed as $x = \xi + t\hat{u}$ where $\hat{u}$ is a unit vector and t is a scalar parameter. Therefore, $(x - \xi) = t\hat{u}$ and instead of (4.30), for example, one can write

$$(4.35) \qquad \langle f'(x), x - \xi \rangle = t \langle f'(\xi), \hat{u} \rangle$$

$$+ \sum_{n=2}^{5} \frac{t^n}{(n-1)!} \langle f^{(n)}(\xi)\hat{u}^{n-1}, \hat{u} \rangle + \frac{L}{4!} t^6,$$

which is valid for all pairings $(t, \hat{u})$ satisfying $\xi + t\hat{u} \in \Omega$. Using the notation $a_n(\hat{u}) = \frac{1}{n!} \langle f^{(n)}(\xi)\hat{u}^{n-1}, \hat{u} \rangle$, for $n = 1, \ldots, 5$, one obtains from (1.11), (4.35) and the convexity of f on Ω

$$(4.36) \qquad \gamma t^5 \leq f(x) - f(\xi) \leq \langle f'(x), x - \xi \rangle$$

$$\leq \sum_{n=1}^{5} na_n(\hat{u})t^n + \frac{L}{4!} t^6 .$$

Consequently, for $\xi + t\hat{u} \in \Omega$ and for t sufficiently small, say $t < \hat{t}$, one has

$$(4.37) \qquad \sum_{n=1}^{5} na_n(u)t^n \geq \frac{\gamma}{2} t^5 .$$

Also, in the simplified notation, (4.31) and (4.34) become

$$(4.38) \qquad \langle f''(x)(x - \xi), x - \xi \rangle \leq \sum_{n=2}^{5} n(n - 1)a_n(\hat{u})t^n + \frac{L}{3!} t^6 ,$$

and for $x \in \Omega \cap \overline{B}_{\hat{t}}(\xi)$, (4.28) has the lower bound

$$(4.39) \qquad \frac{\langle f'(x), x - \xi \rangle}{\langle f''(x)(x - \xi), x - \xi \rangle} \geq \frac{\sum_{n=1}^{5} na_n(\hat{u})t^n - \frac{L}{4!} t^6}{\sum_{n=2}^{5} n(n-1)a_n(\hat{u})t^n + \frac{L}{3!} t^6} \geq 0$$

in view of (4.36) and (4.37). Furthermore, using (4.37) one obtains for $x \in \Omega \cap \overline{B}_{\hat{t}}(\xi)$

$$\left| \frac{\frac{L}{4!} t^6}{\sum_{n=1}^{5} na_n(\hat{u})t^n} \right| \leq \left| \frac{\frac{L}{3!} t^6}{\sum_{n=1}^{5} na_n(\hat{u})t^n} \right| \leq c_1 t, \quad \text{with } c_1 = \frac{L}{3\gamma},$$

and (4.39) then gives

$$(4.40) \qquad \frac{\langle f'(x), x - \xi \rangle}{\langle f''(x)(x - \xi), x - \xi \rangle} \geq \frac{1 - c_1 t}{A(t, \hat{u}) + c_1 t},$$

where

$$(4.41) \qquad A(t, \hat{u}) = \frac{\sum_{n=2}^{5} n(n-1)a_n(\hat{u})t^n}{\sum_{n=1}^{5} na_n(\hat{u})t^n}.$$

If it can be shown that $A(t, \hat{u}) \leq c_2$ for some constant $c_2 < \infty$ and for all $(t, \hat{u})$ pairs satisfying $\xi + t\hat{u} \in \Omega \cap \overline{B}_{\hat{t}}(\xi)$, then (4.28) will follow from (4.40) since

$$(4.42) \qquad \frac{\langle f'(x), x - \xi \rangle}{\langle f''(x)(x - \xi), x - \xi \rangle} \geq \frac{1 - c_1 t}{c_2 + c_1 t} \geq c > 0$$

for t sufficiently small. Note that since ξ is an extremal, $a_1(\hat{u}) \geq 0$; therefore, one can write

$$(4.43) \qquad A(t, \hat{u}) \leq \frac{4a_1(u)t + \sum_{n=2}^{5} n(n-1)a_n(\hat{u})t^n}{\sum_{n=1}^{5} na_n(\hat{u})t^n}$$

$$= 4 - \frac{6a_2(\hat{u})t^2 + 6a_3(\hat{u})t^3 + 4a_4(\hat{u})t^4}{\sum_{n=1}^{5} na_n(\hat{u})t^n}.$$

It it can be established that

$$(4.44) \qquad 6a_2(\hat{u})t^2 + 6a_3(\hat{u})t^3 + 4a_4(\hat{u})t^4 \geq -c_3 t^5$$

for some $c_3 > 0$, and for $(t, \hat{u})$ satisfying $\xi + t\hat{u} \in \Omega \cap \overline{B}_{\tilde{t}}(\xi)$ with $\tilde{t} \leq \hat{t}$, then (4.43), (4.44), and (4.37) would yield

$$A(t, \hat{u}) \leq 4 + \frac{2c_3}{\gamma} \qquad \text{for } (t, \hat{u}) \text{ satisfying } \xi + t\hat{u} \in \Omega \cap \bar{B}_{\tilde{t}}(\xi).$$

To show that (4.44) is true note that convexity of f implies that

$$0 \leq \langle f''(x)(x - \xi), x - \xi \rangle$$

$$\leq \sum_{n=2}^{5} n(n - 1)a_n(\hat{u})t^n + \frac{L}{3!} t^6.$$

Also, $\|a_5(\hat{u})\| \leq \frac{1}{5!}\|f^{(5)}(\xi)\| < \infty$. Therefore,

(4.45) $$2a_2(\hat{u})t^2 + 6a_3(\hat{u})t^3 + 12a_4(\hat{u})t^4$$

$$\geq -(\frac{1}{3!}\|f^{(5)}(\xi)\| + \frac{L}{3!} t)t^5 \geq -c_5t^5, \qquad \text{for some } c_5 \geq 0,$$

for t sufficiently small, say $t \in (0, \tilde{t})$. Writing (4.45) as

(4.46) $$6a_2(\hat{u}) + 18a_3(\hat{u})t + 36a_4(\hat{u})t^2 \geq -3c_5t^3,$$

and making the change of variables $\tau = 3t$ yields

(4.47) $$6a_2(\hat{u}) + 6a_3(\hat{u})\tau + 4a_4(\hat{u})\tau^2 \geq \frac{-c_5\tau^3}{9},$$

which is valid for $\tau \in (0, 3\tilde{t})$. This proves that (4.44) is true at least for $t \in (0, \tilde{t})$ with $c_3 = \frac{c_5}{9}$ and (4.28) follows for $\rho = \min(\hat{t}, \tilde{t}, \frac{1}{2c_1})$.

QED

Remark 4.2. For certain extremals encountered in optimal control theory the exponent ν in (1.11) can be calculated; this was discussed in Remark 3.2. Let $U \in \mathbb{R}^m$ be the unit ball, and as in Remark 3.2, let the constraint set Ω be the set of functions

$$\Omega = \{\text{measurable } u(\cdot):[0, 1] \to U\} \subset L^1([0, 1], \mathbb{R}^m).$$

If $\xi(t)$ is piecewise continuous with range on the boundary of U, and if an associated switching function grows fast enough near its zeros, then it can be shown (Dunn [14]) that (1.11) (or (3.50)) holds for any ν in the range $2 < \nu \leq \infty$. For the conditional gradient method ($M_n = 0$ in the (GS)), linear convergence is not guaranteed by Theorem 3.3, and computer simulations suggest sublinear convergence for a simple example with such a minimizer. On the other hand, Theorem 4.1 and Lemma 4.3 prove superlinear convergence for Newton's method in this setting when f satisfies the hypotheses of Lemma 4.3 (see Remark 4.1 also).

5. Convergence Rates for Nonconvex Functionals

In Theorem 2.1 it was shown that limit points of a sequence x_n generated by the (GS) are extremals. This is true for differentiable, possibly nonconvex functionals f with Lipschitz continuous derivatives f'. However, convergence <u>rate</u> theorems, presented in this thesis so far, have been limited to convex functionals; the proofs of these theorems have depended heavily on the convexity property

$$(5.1) \qquad \langle f'(x), x - y \rangle \geq f(x) - f(y), \qquad \forall x, y \in \Omega,$$

although it was indicated in Remarks 2.4 and 3.3 that such theorems could be extended to a subclass of functionals which satisfy the weaker pseudo-convexity condition

$$(5.2) \qquad \langle f'(x), x - \xi \rangle \geq \kappa(f(x) - f(\xi)), \qquad \text{for some } \kappa > 0,$$

where $\xi \in \Omega_f$. In particular, linear convergence occurs for certain methods in the (GS) when (5.1) or (5.2) holds with (1.10) or (1.11) for $\nu = 2$. It will be shown in this chapter that conditions (5.2) and (1.10) with $\nu = 2$ hold near an extremal of a (possibly nonconvex) functional f if, for some $\rho > 0$, f satisfies

$$(5.3) \qquad Q(f''(\xi), \xi, x) \geq \gamma\|x - \xi\|^2 \qquad \forall x \in K_\Omega(\xi) \cap B_\rho(\xi).$$

Here $K_\Omega(\xi)$ is the tangent cone to Ω at ξ with vertex at ξ (see Chapter 1). If the operator sequence in the (GS) satisfies (1.13) or if the structure of the set Ω near ξ is such that (1.11) holds with $\nu = 2$, i.e., (5.4) $$\langle f'(\xi), x - \xi \rangle \geq \tilde{\gamma}\|x - \xi\|^2, \qquad \forall x \in \Omega, \text{ and some } \tilde{\gamma} > 0,$$

then Theorem 5.1 shows that if $\{x_n\}$ passes sufficiently near ξ, it will converge to ξ at a linear rate. When $\{M_n\}$ satisfies certain "quasi-Newton" conditions, (5.3) need hold only for $x \in \Omega \cap B_\rho(\xi)$ for some $\rho > 0$ to insure linear or superlinear convergence rates; this will be established in Theorem 5.2.

Lemma 5.1. Let $S \subset X$ be convex where X is a Banach space. Let f be twice differentiable with f'' continuous at ξ and let f satisfy (5.3) at ξ for $x \in S \cap B_\rho(\xi)$ and some $\rho > 0$. Then for some $\rho_1 > 0$

(5.5) $$f(x) - f(\xi) \geq \frac{\gamma}{2}\|x - \xi\|^2 \qquad \forall x \in S \cap B_{\rho_1}(\xi)$$

Proof. Using Taylor's formula for $x \in S \cap B_\rho(\xi)$ at ξ one has

$$\begin{aligned} f(x) - f(\xi) &= \langle f'(\xi), x - \xi \rangle + \frac{1}{2}\langle f''(\zeta)(x - \xi), x - \xi \rangle \\ &= Q(f''(\xi), \xi, x) + \frac{1}{2}\langle (f''(\zeta) - f''(\xi))(x - \xi), x - \xi \rangle \\ &\geq (\gamma - \frac{1}{2}\|f''(\zeta) - f''(\xi)\|)\|x - \xi\|^2 \end{aligned}$$

for ζ between x and ξ. By the continuity of f'' there exists a ρ_1 such that for $\|x - \xi\| < \rho_1$, $\|f''(\zeta) - f''(\xi)\| < \gamma$; (5.5) now follows for $x \in S \cap B_{\rho_1}(\xi)$.

QED

Remark 5.1. The proof of Lemma 5.1 is essentially the same as the proof of Lemma 2.4 in [22].

<u>Lemma 5.2</u>. Let $\Omega \subset X$ be convex where X is a Banach space. Let f be twice differentiable and let f" be Lipschitz continuous for $x \in K_\Omega(\xi) \cap B_\rho(\xi)$ for some $\rho > 0$. Let f satisfy (5.3) for $x \in K_\Omega(\xi) \cap B_\rho(\xi)$. Then for some $\rho_2 > 0$ and $k > 0$

(5.6) $$\langle f'(x), x - \xi \rangle \geq k(f(x) - f(\xi)), \quad \text{for } x \in K_\Omega(\xi) \cap B_{\rho_2}(\xi)$$

<u>Proof</u>. Since $K_\Omega(\xi)$ is convex it follows from Lemma 5.1 that $f(x) - f(\xi) > 0$ for $x \in K_\Omega(\xi) \cap B_{\rho_1}(\xi)$ for some $\rho_1 < \rho$. It must be shown, therefore, that for some $\rho_2 < \rho_1$, and some $k > 0$, and for $x \neq \xi$

(5.7) $$\frac{\langle f'(x), x - \xi \rangle}{f(x) - f(\xi)} \geq k \qquad \text{for } x \in K_\Omega(\xi) \cap B_{\rho_2}(\xi).$$

From Taylor's formula and the Lipschitz continuity of f", one can write for $x \neq \xi$ and $x \in K_\Omega(\xi) \cap B_{\rho_1}(\xi)$

(5.8) $$\frac{f(x) - f(\xi)}{Q(f''(\xi), \xi, x)}$$

$$\leq \frac{\langle f'(\xi), x - \xi \rangle + \frac{1}{2}\langle f''(\xi)(x - \xi), x - \xi \rangle + c_1\|x - \xi\|^3}{Q(f'(\xi), \xi, x)}$$

$$\geq 1 + c_2\|x - \xi\|,$$

where

$$c_2 = \frac{c_1}{\gamma}.$$

Similarly, one obtains

$$(5.9) \qquad \frac{\langle f'(x), x - \xi\rangle}{Q(f''(\xi), \xi, x)}$$

$$\geq \frac{\langle f'(\xi), x - \xi\rangle + \langle f''(\xi)(x - \xi), x - \xi\rangle - c_1\|x - \xi\|^3}{Q(f''(\xi), \xi, x)}$$

$$\geq 1 + \frac{\frac{1}{2}\langle f''(\xi)(x - \xi), x - \xi\rangle}{Q(f''(\xi), \xi, x)} - c_2\|x - \xi\|.$$

Let $x = \xi + t\hat{u}$ for $\hat{u}$ a unit vector in $K_\Omega(\xi)$, and let $R(t, \hat{u})$ be defined for $t \in (0, \rho]$ by

$$(5.10) \qquad R(t, \hat{u}) = \frac{\frac{1}{2} t^2\langle f''(\xi)\hat{u}, \hat{u}\rangle}{t\langle f'(\xi), \hat{u}\rangle + \frac{1}{2} t^2\langle f''(\xi)\hat{u}, \hat{u}\rangle} .$$

Whenever $\langle f''(\xi)\hat{u}, \hat{u}\rangle \geq 0$, then $R(t, \hat{u}) \geq 0$ and (5.9) yields

$$(5.11) \qquad \frac{\langle f'(x), x - \xi\rangle}{Q(f''(\xi), \xi, x)} \geq 1 - c_2 t.$$

On the other hand for any $\hat{u}$ for which $\langle f''(\xi)\hat{u}, \hat{u}\rangle < 0$, it must be shown that $|R(t, \hat{u})| < c_3 t$ for some $c_3 < \infty$. If this is so, then from (5.9) one has

$$(5.12) \qquad \frac{\langle f'(x), x - \xi\rangle}{Q(f''(\xi), \xi, x)} \geq 1 - |R(t, \hat{u})| - c_2 t \geq 1 - (c_3 + c_2)t,$$

and for t sufficiently small (5.7) will follow from (5.8), (5.11), and (5.12). For $\hat{u} \in K_\Omega(\xi)$ and for $t \in [0, \rho]$, (5.3) gives

$$t\langle f'(\xi), \hat{u}\rangle + \frac{1}{2} t^2\langle f''(\xi)\hat{u}, \hat{u}\rangle \geq \gamma t^2.$$

In particular, for $t = \rho$ and $\langle f''(\xi)\hat{u}, \hat{u}\rangle < 0$, one has

$$\langle f'(\xi), \hat{u}\rangle \geq (\gamma - \frac{1}{2}\langle f''(\xi)\hat{u}, \hat{u}\rangle)\rho > 0$$

in which case, one can write for $t \in [0, \rho]$

$$t\langle f'(\xi), \hat{u}\rangle + \frac{1}{2} t^2 \langle f''(\xi)\hat{u}, \hat{u}\rangle$$

$$\geq t\gamma\rho - \frac{1}{2} t(\rho - t)\langle f''(\xi)\hat{u}, \hat{u}\rangle \geq t(\gamma\rho).$$

Consequently, for $t \in (0, \rho]$ and $\langle f''(\xi)\hat{u}, \hat{u}\rangle < 0$, (5.10) yields

$$|R(t, \hat{u})| \leq \frac{|t^2\langle f''(\xi)\hat{u}, \hat{u}\rangle|}{2t\gamma\rho} \leq \frac{\|f''(\xi)\|}{2\gamma\rho} t = c_3 t.$$

QED

Roughly speaking, Lemmas 5.1 and 5.2 show that the functional f is "locally pseudoconvex with respect to ξ" when condition (5.3) holds.

<u>Theorem 5.1</u>. Let $\Omega \subset X$ be convex where X is a Banach space. Let f be twice differentiable with f" Lipschitz continuous for $x \in K_\Omega(\xi) \cap B_\rho(\xi)$. Let f satisfy (5.3) for $x \in K_\Omega(\xi) \cap B_\rho(\xi)$. Let either (1.13) be satisfied by the operator sequence $\{M_n\}$, or the structure of the set near ξ be such that (5.4) holds. If $\{x_n\}$, a sequence generated by the (GS), comes sufficiently near ξ, then $x_n \to \xi$ and $r_n = f(x_n) - f(\xi) = O(\lambda^n)$, for some $\lambda \in [0, 1)$.

<u>Proof</u>. By Lemmas 5.1 and 5.2 and condition (5.3) there is a ρ_1 such that

$$(5.13) \qquad \langle f'(x), x - \xi\rangle \geq k(f(x) - f(\xi)) \geq \frac{k\gamma}{2}\|x - \xi\|^2$$

$$\text{for } x \in \Omega \cap B_{\rho_1}(\xi) \subset K_\Omega(\xi) \cap B_{\rho_1}(\xi).$$

Furthermore if L_1 is a Lipschitz constant for f', and if $\{M_n\}$ satisfies (1.13), then (3.32) in Lemma 3.1 becomes

$$(5.14)\qquad \frac{1}{2}\mu\|\hat{x}_n - \xi\|^2 + \langle f'(\xi), \hat{x}_n - \xi\rangle$$
$$\leq (L_1 + K)\|x_n - \xi\|\,\|\hat{x}_n - \xi\|, \qquad \forall n \geq 0,$$

where K is a uniform bound on $\|M_n\|$. On the other hand, if (5.4) holds then (3.32) becomes

$$(5.15)\qquad \frac{1}{2}\langle M_n(\hat{x}_n - \xi), \hat{x}_n - \xi\rangle + \tilde{\gamma}\|\hat{x}_n - \xi\|^2$$
$$\leq (L_1 + K)\|x_n - \xi\|\,\|\hat{x}_n - \xi\|, \qquad \forall n \geq 0.$$

Since $f(x) - f(\xi) > 0$ in a small neighborhood around ξ, it follows that ξ is an extremal, i.e., $\langle f'(\xi), x - \xi\rangle \geq 0$, $\forall x \in \Omega$, hence (5.14) yields

$$(5.16)\qquad \|\hat{x}_n - \xi\| \leq c\|x_n - \xi\| \qquad \forall n \geq 0, \text{ and } c = 2(L_1 + K)/\mu$$

Similarly the nonnegativity of M_n in (5.15) gives (5.16) with $c = (L_1 + K)/\tilde{\gamma}$. In either case, when $0 < \|x_n - \xi\| < \rho_2 = \frac{\rho_1}{c}$, then $\hat{x}_n \in \Omega \cap B_{\rho_1}(\xi)$. Let $A = \{x \in \Omega \cap B_{\rho_1}(\xi) : f(x) - f(\xi) < \frac{\gamma}{2}\rho_2^2\}$. Then (5.13) implies that $A \subset \Omega \cap B_{\rho_2}(\xi)$. Since f is continuous there is a $\rho_3 > 0$, such that if $x \in \Omega \cap B_{\rho_3}(\xi)$ then $x \in A$. It follows then that if $x_N \in \Omega \cap B_{\rho_3}(\xi)$ for some $N > 0$, then $x_n \in \Omega \cap B_{\rho_2}(\xi)$ for $n \geq N$. This is true because $\hat{x}_N \in \Omega \cap B_{\rho_1}(\xi)$, and since x_{N+1} is a convex combination of x_N and $\hat{x}_N$, then $x_{N+1} \in \Omega \cap B_{\rho_1}(\xi)$. But from Theorem 2.1 line (2.25) and Lemma 5.1, $f(\xi) \leq f(x_{N+1}) < f(x_N)$ and, therefore, $x_{N+1} \in A \subset \Omega \cap B_{\rho_2}(\xi)$. By an induction argument $x_n \in \Omega \cap B_{\rho_2}(\xi)$ for $n \geq N$. Also, since $\langle f'(x_n), x_n - \xi\rangle \geq \frac{k\gamma}{2}\|x_n - \xi\|^2$ for $n \geq N$ it follows that $\xi \in \Phi(x_n)$, and Lemma 2.2 and (5.13) yield

$$(5.17)\quad \langle f'(x_n), x_n - \hat{x}_n \rangle \geq \begin{cases} \langle f'(x_n), x_n - \xi \rangle & \text{if } \langle M_n(x_n - \xi), x_n - \xi \rangle = 0 \\ \frac{1}{2}\min\{\langle f'(x_n), x_n - \xi \rangle, \frac{\langle f'(x_n), x_n - \xi \rangle^2}{K\|x_n - \xi\|^2}\}, & \text{if } \langle M_n(x_n - \xi), x_n - \xi \rangle > 0 \end{cases}$$

$$\geq \frac{k}{2}\min\{(f(x_n) - f(\xi)), \frac{k\gamma}{2K}(f(x_n) - f(\xi))\}$$

$$\geq c_4(f(x_n) - f(\xi))$$

where $c_4 = \frac{k}{2}\min\{1, \frac{k\gamma}{2K}\} > 0$. From (5.16), (5.17), (2.23) and the triangle inequality one now has

$$(5.18)\qquad \omega_n \geq \min\{1, \frac{2\delta c_4(f(x_n) - f(\xi))}{L_1\|x_n - \hat{x}_n\|^2}\}$$

$$\geq \min\{1, \frac{\delta c_4 k\gamma\|x_n - \xi\|^2}{L_1(1 + c)^2\|x_n - \xi\|^2}\} = \omega > 0,$$

where L_1 is a Lipschitz constant for f'. Finally, (2.25) gives

$$f(x_n) - f(\xi) - f(x_{n+1}) + f(\xi) \geq \delta\omega c_4(f(x_n) - f(\xi))$$

or equivalently

$$r_{n+1} \leq (1 - \delta\omega c_4)r_n$$

and $r_n = O((1 - \delta\omega c_4)^n) = O(\lambda^n)$. The linear convergence of $\{r_n\}$, together with (5.13) implies that $\{\|x_n - \xi\|\}$ converges to zero at a linear rate.

QED

The hypotheses of Theorem 5.1 can be weakened somewhat when the operators M_n are so called "quasi-Newton" operators. If the M_n's are symmetric and approach the second derivative operator $f''(\xi)$ in the sense of (1.18) - (1.21) then (5.3) need hold only for $x \in \Omega \cap B_\rho(\xi)$ and f'' need only be continuous at ξ to establish linear and superlinear rates of convergence near ξ.

Theorem 5.2. Let $\Omega \subset X$ be convex where X is a Banach space. Let f be twice differentiable with f'' continuous. Let f satisfy (5.3) for $x \in \Omega \cap B_\rho(\xi)$ for some $\rho > 0$ and $\xi \subset \Omega$. Let $\{M_n\}$ be a sequence of symmetric operators in the (GS), i.e.,

$$\langle M_n x, y \rangle = \langle M_n y, x \rangle, \qquad \forall x, y \in X, \quad n \geq 0.$$

Then:

(i) If (1.18) or (1.19) holds with ε sufficiently small and $n \geq N$, and if $\{x_n\}$ is a sequence generated by the (GS), there exists a $\rho_1 > 0$ such that if $x_{n_o} \in \Omega \cap B_{\rho_1}(\xi)$ for some $n_o \geq N$, then $\{\|x_n - \xi\|\}$ converges to zero at a linear rate, i.e., for some $\lambda \in (0, 1)$

$$\|x_{n+1} - \xi\| \leq \lambda \|x_n - \xi\|, \qquad \text{for } n \geq n_o.$$

(ii) If either (1.20) or (1.21) holds then $\{\|x_n - \xi\|\}$ converges superlinearly, i.e., either $x_n = \xi$ eventually, or

$$\lim_{n\to\infty} \frac{\|x_{n+1} - \xi\|}{\|x_n - \xi\|} = 0.$$

<u>Proof of (i)</u>. It will first be shown that if x_{n_o} is sufficiently close to ξ with $n_o \geq N$, then $\|\hat{x}_{n_o} - \xi\| \leq c\|x_{n_o} - \xi\|$ for some $c \in (0, 1)$, and, by induction, that $\|\hat{x}_n - \xi\| \leq c\|x_n - \xi\|$ for $n \geq n_o$. Then it will be established that $\omega_n \geq \omega > 0$ for $n \geq n_o$ and that this implies that $\|x_{n+1} - \xi\| \leq \lambda\|x_n - \xi\|$ for some $\lambda \in (0, 1)$ and for $n \geq n_o$. To show that $\|\hat{x}_n - \xi\| \leq c\|x_n - \xi\|$ for some $c \in (0, 1)$ and $n \geq n_o$, let $\overline{x}_n$ satisfy for any $n \geq 0$

$$0 \leq Q(M_n, x_n, \xi) - Q(M_n, x_n, \overline{x}_n). \tag{5.19}$$

Then inequality (3.1) in Lemma 3.1 holds with $\hat{x}_n$ replaced by $\overline{x}_n$, and with the symmetry of M_n one can write

$$\frac{1}{2}\langle M_n(\overline{x}_n - \xi), \overline{x}_n - \xi\rangle + \langle f'(\xi), \overline{x}_n - \xi\rangle$$

$$\leq \langle f'(\xi) - f'(x_n), \overline{x}_n - \xi\rangle + \langle M_n(x_n - \xi), \overline{x}_n - \xi\rangle.$$

With the Mean Value Theorem, this implies

$$\frac{1}{2}\langle (M_n - f''(\xi))(\overline{x}_n - \xi), \overline{x}_n - \xi\rangle + Q(f''(\xi), \xi, \overline{x}_n)$$

$$\leq \langle (M_n - f''(\zeta_n))(x_n - \xi), \overline{x}_n - \xi\rangle$$

$$\text{for } \zeta_n = x_n + \theta_n(\xi - x_n),\ \theta_n \in [0, 1].$$

Consequently, for $\bar{x}_n \in \Omega \cap B_\rho(\xi)$, condition (5.3) and the triangle inequality yield for $\bar{x}_n \neq \xi$,

$$(5.20)\qquad \left(\gamma - \frac{\|(M_n - f''(\xi))(\bar{x}_n - \xi)\|}{\|\bar{x}_n - \xi\|}\right)\|\bar{x}_n - \xi\|^2$$

$$\leq \left(\frac{\|(M_n - f''(\xi))(x_n - \xi)\|}{\|x_n - \xi\|} + \|f''(\xi) - f''(\zeta_n)\|\right)\|x_n - \xi\|\,\|\bar{x}_n - \xi\|$$

or

$$(5.21)\qquad (\gamma - \|M_n - f''(\xi)\|)\|\bar{x}_n - \xi\|^2$$

$$\leq (\|M_n - f''(\xi)\| + \|f''(\xi) - f''(\zeta_n)\|)\|x_n - \xi\|\,\|\bar{x}_n - \xi\| .$$

Suppose (1.19) holds. Then since f'' is continuous, there is a $\rho_1 \in (0, \rho)$ such that if $\bar{x}_n \in \Omega \cap B_\rho(\xi)$ and $x_n \in \Omega \cap B_{\rho_1}(\xi)$ for $n \geq N$, then (5.20) and (1.19) with ε sufficiently small imply

$$(5.22)\qquad \|\bar{x}_n - \xi\| \leq c\|x_n - \xi\|,$$

where

$$c = \left(\frac{\varepsilon + \|f''(\xi) - f''(\zeta_n)\|}{\gamma - \varepsilon}\right) < 1.$$

Let $x_{n_o} \in \Omega \cap B_{\rho_1}(\xi)$ for some $n_o \geq N$. Suppose that $\|\hat{x}_{n_o} - \xi\| > \rho$, and let $\bar{x}_{n_o} = \xi + \rho \dfrac{(\hat{x}_{n_o} - \xi)}{\|\hat{x}_{n_o} - \xi\|}$. Since $Q(M_{n_o}, x_{n_o}, \cdot)$ is convex and $\hat{x}_{n_o}$ minimizes $Q(M_{n_o}, x_{n_o}, \cdot)$ on Ω, one has

(5.23) $\quad Q(M_{n_o}, x_{n_o}, \hat{x}_{n_o}) \le Q(M_{n_o}, x_{n_o}, \bar{x}_{n_o}) \le Q(M_{n_o}, x_{n_o}, \xi).$

Therefore, $\bar{x}_{n_o}$ satisfies (5.19), and since $\|\bar{x}_{n_o} - \xi\| = \rho$, (5.22) holds. But (5.22) gives $\|\bar{x}_{n_o} - \xi\| \le c\|x_{n_o} - \xi\| < \rho_1 < \rho$, and this contradiction proves that $\|\hat{x}_{n_o} - \xi\| \le \rho$. From (5.23), (5.19), and (5.22) there results

$$\|\hat{x}_{n_o} - \xi\| \le c\|x_{n_o} - \xi\|, \qquad \text{for some } c \in (0, 1).$$

It follows by induction that for $n \ge n_o$, $x_n \in \Omega \cap B_{\rho_1}(\xi)$ and

(5.24) $\quad \|\hat{x}_n - \xi\| \le c\|x_n - \xi\|.$

If (1.18) holds then (5.24) follows by the same argument. To prove that $\{\|x_n - \xi\|\}$ converges at a linear rate it suffices to show that $\omega_n \ge \omega > 0$ since, one can then write for $n \ge n_o$, $x_n \ne \xi$,

(5.25)
$$\frac{\|x_{n+1} - \xi\|}{\|x_n - \xi\|} = \frac{\|x_n + \omega_n(\hat{x}_n - \xi + \xi - x_n) - \xi\|}{\|x_n - \xi\|}$$

$$\le \frac{(1 - \omega_n)\|x_n - \xi\| + \omega_n\|\hat{x}_n - \xi\|}{\|x_n - \xi\|}$$

$$\le (1 - \omega_n) + \omega_n c$$

$$= 1 - \omega_n(1 - c)$$

$$\le 1 - \omega(1 - c) < 1.$$

To prove that $\omega_n \geq \omega > 1$ observe that for $x_n \in \Omega \cap B_\rho(\xi)$

$$(5.26)\qquad -Q(M_n, x_n, \hat{x}_n) \geq -Q(M_n, x_n, \xi)$$

$$= Q(f''(\xi), \xi, x_n) - Q(f''(\xi), \xi, x_n) - Q(M_n, x_n, \xi)$$

$$\geq \gamma\|x_n - \xi\|^2 - (\langle f'(\xi) - f'(x_n), x_n - \xi\rangle$$

$$+ \frac{1}{2}\langle f''(\xi)(x_n - \xi), x_n - \xi\rangle + \frac{1}{2}\langle M_n(x_n - \xi), x_n - \xi\rangle).$$

The Mean Value Theorem, the triangle inequality, and (5.26) give for $x_n \neq \xi$

$$(5.27)\qquad \langle f'(x_n), x_n - \hat{x}_n\rangle - \frac{1}{2}\langle M_n(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle$$

$$\geq \gamma\|x_n - \xi\|^2 + \langle f''(\zeta_n)(x_n - \xi), x_n - \xi\rangle$$

$$- \frac{1}{2}\langle f''(\xi)(x_n - \xi), x_n - \xi\rangle - \frac{1}{2}\langle M_n(x_n - \xi), x_n - \xi\rangle$$

$$\geq [\gamma - \frac{1}{2}\|f''(\zeta_n) - f''(\xi)\|$$

$$- \frac{1}{2}(\frac{\|(f''(\zeta_n) - f''(\xi))(x_n - \xi)\| + \|(f''(\xi) - M_n)(x_n - \xi)\|}{\|x_n - \xi\|}]\cdot\|x_n - \xi\|^2$$

$$\geq (\gamma - \|f''(\zeta_n) - f''(\xi)\| - \frac{\frac{1}{2}\|(f''(\xi) - M_n)(x_n - \xi)\|}{\|x_n - \xi\|})\|x_n - \xi\|^2,$$

where $\zeta_n = x_n + \theta_n(\xi - x_n)$, $\theta_n \in [0, 1]$,

or

$$(5.28)\qquad \langle f'(x_n), x_n - \hat{x}_n\rangle$$

$$\geq (\gamma - \|f''(\zeta_n) - f''(\xi)\| - \frac{1}{2}\|f''(\xi) - M_n\|)\|x_n - \xi\|^2.$$

Suppose (1.19) holds with ε small enough. Once again, since f" is continuous, there is a $\rho_2 \in (0, \rho_1)$ such that if $x_n \in \Omega \cap B_{\rho_2}(\xi)$ for $n \geq N$ then (5.27) yields

$$(5.29) \qquad \langle f'(x_n), x_n - \hat{x}_n \rangle \geq \bar{\gamma}\|x_n - \xi\|^2, \qquad \text{for } \bar{\gamma} > 0.$$

From (2.23), (5.22), (5.29) and the triangle inequality one has, for $x_n \neq \xi$ and $n \geq n_o \geq N$,

$$(5.30) \qquad \omega_n \geq \min\{1, \frac{2\delta\bar{\gamma}\|x_n - \xi\|^2}{L\|x_n - \hat{x}_n\|^2}\}$$

$$\geq \min\{1, \frac{2\delta\bar{\gamma}\|x_n - \xi\|^2}{L(\|x_n - \xi\| + \|\hat{x}_n - \xi\|)^2}\}$$

$$\geq \min\{1, \frac{2\delta\bar{\gamma}}{L(1 + c)^2}\} = \omega > 0.$$

Finally, if $x_{n_o} \in \Omega \cap B_{\rho_2}(\xi)$ for some $n_o \geq N$, then from (5.25) and (5.22) it follows that $\{\|x_n - \xi\|\}$ converges to zero at a linear rate. The same result can be established when (1.18) holds using (5.28) and the same argument.

<u>Proof of (ii)</u>. Condition (1.20) implies condition (1.18); therefore, the results of (i), (1.20), (5.21) and the continuity of f" yield

$$(5.31) \qquad \|\hat{x}_n - \xi\| \leq \lambda_n\|x_n - \xi\|, \qquad \text{where } \lambda_n \to 0.$$

AD-A106 715 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH F/G 12/1
CONVERGENCE RATE ANALYSIS FOR ITERATIVE MINIMIZATION SCHEMES WI--ETC(U)
1981 G C HUGHES
UNCLASSIFIED AFIT-CI-81-17D NL

2 of 2

END
DATE
FILMED
11-81
DTIC

Also, (5.31) follows from the results of (i), (1.21), (5.20) and the continuity of f''. One need only show that after a finite number of iterations $\omega_n = 1$ for the remaining iterations. Fix $\delta \in (0, \frac{1}{2})$. Then from (4.22) one has, for x_n not an extremal,

$$(5.32)\qquad g(x_n, \hat{x}_n, 1) = 1 - \frac{\frac{1}{2}\langle f''(\zeta_n)(\hat{x}_n - x_n), \hat{x}_n - x_n\rangle}{\langle f'(x_n), x_n - \hat{x}_n\rangle}$$

$$\geq 1 - \frac{\langle M_n(x_n - \hat{x}_n), x_n - \hat{x}_n\rangle}{2\langle f'(x_n), x_n - \hat{x}_n\rangle}$$

$$- \frac{\|(f''(\zeta_n) - M_n)(\hat{x}_n - x_n)\| \, \|\hat{x}_n - x_n\|}{2\langle f'(x_n), x_n - \hat{x}_n\rangle} .$$

From the triangle inequality, (5.32) and (4.24), which is valid for symmetric operators M_n, there results

$$(5.33)\qquad g(x_n, \hat{x}_n, 1) \geq \frac{1}{2} - \frac{\|f''(\zeta_n) - f''(\xi)\| \, \|\hat{x}_n - x_n\|^2}{2\langle f'(x_n), x_n - \hat{x}_n\rangle}$$

$$- \frac{\dfrac{\|(f''(\xi) - M_n)(x_n - \xi)\|}{\|x_n - \xi\|} \|x_n - \xi\| \, \|x_n - \hat{x}_n\|}{2\langle f'(x_n), x_n - \hat{x}_n\rangle}$$

$$- \frac{\dfrac{\|(f''(\xi) - M_n)(\hat{x}_n - \xi)\|}{\|\hat{x}_n - \xi\|} \|\hat{x}_n - \xi\| \, \|x_n - \hat{x}_n\|}{2\langle f'(x_n), x_n - \hat{x}_n\rangle} .$$

Also, (5.22), (5.29) and the triangle inequality yield for $x_n \in \Omega \cap B_{\rho_2}(\xi)$,

$$(5.34) \qquad g(x_n, \hat{x}_n, 1) \geq \frac{1}{2} - \frac{(1 + c)^2}{2\tilde{\gamma}} \|f''(\zeta_n) - f''(\xi)\| - \frac{(1 + c)}{2\tilde{\gamma}} \frac{\|(f''(\xi) - M_n)(x_n - \xi)\|}{\|x_n - \xi\|} - \frac{c(1 + c)}{2\tilde{\gamma}} \frac{\|(f''(\xi) - M_n)(\hat{x}_n - \xi)\|}{\|\hat{x}_n - \xi\|}$$

or

$$(5.35) \qquad g(x_n, \hat{x}_n, 1) \geq \frac{1}{2} - \frac{(1 + c)^2}{2\tilde{\gamma}} \|f''(\zeta_n) - f''(\xi)\| - \frac{(1 + c)}{2\tilde{\gamma}} \|f''(\xi) - M_n\| - \frac{c(1 + c)}{2\tilde{\gamma}} \|f''(\xi) - M_n\|.$$

Finally, the continuity of f" and (1.20) in (5.35) or (1.21) in (5.34) give $g(x_n, \hat{x}_n, 1) \geq \delta$ for $n \geq N_1(\delta)$ with $N_1(\delta) < \infty$.

QED

Remark 5.2. The proof of Theorem 5.2 is a modification of the proofs in [22] for Newton's method.

REFERENCES

[1] S. -P. Han, "Superlinearly convergent variable metric algorithms for general nonlinear programming problems," Mathematical Programming, 11, (1976), pp. 263-282.

[2] V. M. Garcia Palomares and O. L. Mangasarian, "Superlinearly convergent quasi-Newton algorithms for nonlinearly constrained optimization problems," Mathematical Programming, 11, (1976), pp. 1-13.

[3] D. P. Bertsekas, "Projected Newton methods for optimization problems with simple constraints," (to appear).

[4] J. C. Dunn, "Rates of convergence for conditional gradient algorithms near singular and nonsingular extremals," SIAM J. Contr. Opt., 17, 2(1979), pp. 187-211.

[5] J. C. Dunn, "Convergence rates for conditional gradient sequences generated by implicit step length rules," SIAM J. Contr. Opt., 18, 5(1980), pp. 473-487.

[6] A. I. Cohen, "Stepsize analysis for descent methods," J. Optimization Theory Appl., 33, 2(1981), pp. 187-206.

[7] A. A. Goldstein, "Minimizing functions on Hilbert space," in Computer Methods in Optimization Problems, Academic Press, New York (1964).

[8] E. S. Levitin and B. T. Poljak, "Constrained minimization methods," U.S.S.R. Comput. Math. and Math. Phys., 6, 5(1966), pp. 1-50.

[9] V. F. Demyanov and A. M. Rubinov, Approximate Methods in Optimization Problems, American Elsevier, New York (1971).

[10] J. C. Dunn, "Global and asymptotic convergence rate estimates for a class of projected gradient processes," SIAM J. Contr. Opt., 19, 3(1981), pp. 368-400.

[11] O. L. Mangasarian, "Pseudo convex functions," J. SIAM Contr., Ser. A, 3, 2(1965), pp. 281-290.

[12] M. Frank and P. Wolfe, "An algorithm for quadratic programming," Nav. Res. Logist. Quart., 3(1956), pp. 95-110.

[13] M. D. Cannon and C. D. Cullum, "A tight upper bound on the rate of convergence of the Frank-Wolfe algorithm," SIAM J. Contr. Opt., 6(1968), pp. 509-516.

[14] J. C. Dunn, Private communication.

[15] J. C. Allwright, "A feasible direction algorithm for convex optimization: global convergence rates," J. Optimization Theory Appl., 30, 1(1980), pp. 1-18.

[16] E. R. Barnes, "A geometrically convergent algorithm for solving optimal control problems," SIAM J. Contr. Opt., 10(1972), pp. 434-443.

[17] J. C. Dunn, "Extremal types for L^p minimization problems with polyhedral range sets," (to appear).

[18] L. V. Kantorovich, "Functional analysis and applied mathematics," Uspekhi Math. Nauk., 3, 6(1948), pp. 89-185.

[19] A. A. Goldstein, "On Newton's method," Numer. Math., 7(1965), pp. 391-393.

[20] Y. M. Danilin, "Minimization methods based on approximation of the initial functional by a convex functional," U.S.S.R. Comp. Math. and Math. Phys., 10, 5(1970), pp. 1-18.

[21] V. A. Bulavskii, "On extensions of the convergence domain of higher precision iterative methods," Soviet Math. Dokl., 13, 4(1972), pp. 915-917.

[22] J. C. Dunn, "Newton's method and the Goldstein step length rule for constrained minimization problems," SIAM J. Contr. Opt., 18, 6(1980), pp. 659-674.

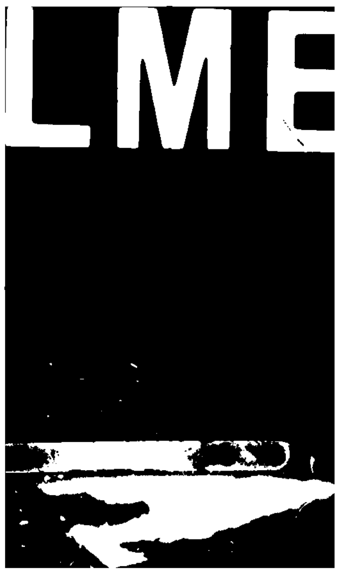
[23] A. A. Goldstein, "On steepest descent," J. SIAM Contr., Ser. A, 3, 1(1965), pp. 147-151.

[24] D. P. Bertsekas, "On the Goldstein-Levitin-Poljak gradient projection method," IEEE Trans. Auto. Contr., AC-21, 2(1976), pp. 174-184.

[25] J. E. Dennis and J. J. Moré, "Quasi-Newton methods, motivation and theory," SIAM Review, 19, 1(1977), pp. 46-89.

[26] W. T. Gruver and E. Sachs, Algorithmic Methods in Optimal Control, Pitman, London (1980).

[27] M. M. Vainberg, Variational Method and Method of Monotone Operators in the Theory of Nonlinear Equations, Wiley, New York (1973).


LME