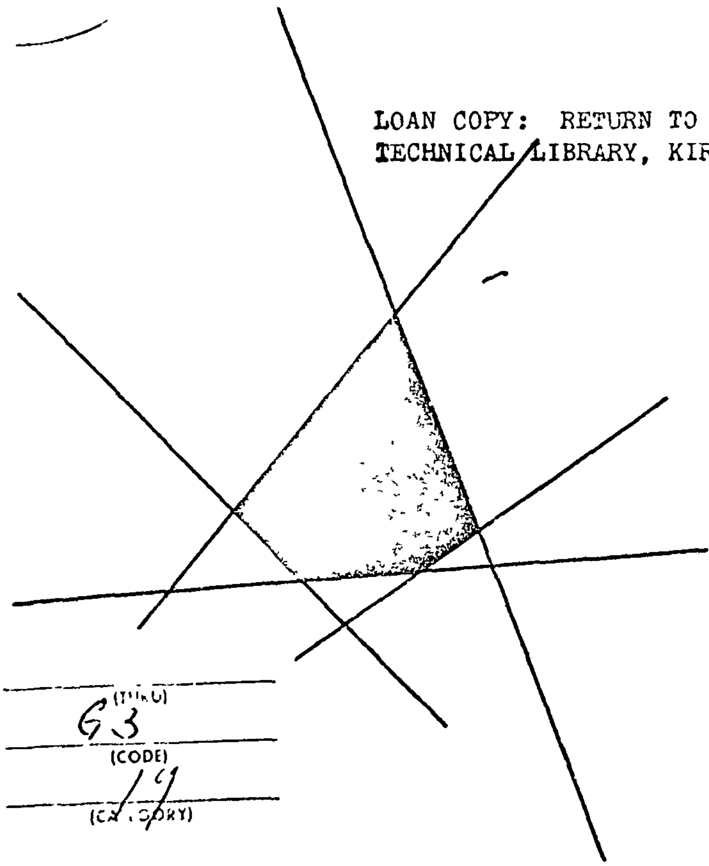
ORC 71-10
MAY 1971

CONVERGENCE RATES FOR A METHOD OF CENTERS ALGORITHM

by

ROBERT B. MIFFLIN



OPERATIONS
RESEARCH
CENTER

COLLEGE OF ENGINEERING
UNIVERSITY OF CALIFORNIA · BERKELEY

CONVERGENCE RATES FOR A METHOD OF CENTERS ALGORITHM

by

Robert B. Mifflin
Operations Research Center
University of California, Berkeley

MAY 1971

ORC 71-10

NGT-05-003-023

This research has been partially supported by the National Aeronautics and Space Administration Predoctoral Traineeship ~~NsG(T)-117~~ and the National Science Foundation under Grant GP-25081 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) University of California, Berkeley	2a REPORT SECURITY CLASSIFICATION Unclassified 2b GROUP

3 REPORT TITLE

CONVERGENCE RATES FOR A METHOD OF CENTERS ALGORITHM

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

Research Report

5 AUTHOR(S) (First name, middle initial, last name)

Robert B. Mifflin

6 REPORT DATE	7a TOTAL NO OF PAGES	7b NO OF REFS
May 1971	90	35

8a CONTRACT OR GRANT NO GP-25081 b PROJECT NO c d	9a ORIGINATOR'S REPORT NUMBER(S) ORC 71-10 9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10 DISTRIBUTION STATEMENT

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

11 SUPPLEMENTARY NOTES SEE TITLE PAGE.	12 SPONSORING MILITARY ACTIVITY NSF-GP-25081-Gale, The National Science Foundation Washington, D.C. 20550

13 ABSTRACT

SEE ABSTRACT.

DD FORM 1473 (PAGE 1)

Unclassified

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Nonlinear Programming						
Constrained Optimization						
Algorithms						
Rate of Convergence						
Sequential Unconstrained Optimization						

ACKNOWLEDGEMENTS

I would like to express my appreciation to the members of my dissertation committee, Professors Donald M. Topkis, C. Roger Glassey and Richard C. Grinold. I am especially grateful to Professor Topkis for his helpful suggestions and encouragement. I would also like to acknowledge the support of the National Aeronautics and Space Administration and the Operations Research Center of the University of California at Berkeley and to especially thank Catherine Korte and Linda Betters for the typing of the manuscript.

ABSTRACT

Convergence of a method of centers algorithm for solving nonlinear programming problems whose feasible regions have nonempty strict interiors is considered. Conditions are given under which the algorithm generates sequences of feasible points and multiplier vectors which have accumulation points satisfying the Fritz John and the Kuhn-Tucker optimality conditions. Under stronger assumptions linear convergence rates are established for the sequences of objective function, constraint function, feasible point and multiplier values.

The feasible points generated by the algorithm may be exact or approximate solutions to unconstrained maximization subproblems and in the approximate case may be found by finite step procedures. Upper bounds are derived for the number of steps required to solve each subproblem when the method of steepest ascent is employed.

1. INTRODUCTION

Consider the nonlinear programming problem of maximizing $f(x)$ subject to the constraints $g_i(x) \geq 0$ for $i = 1,2, \ldots, m$ where $g_1, \ldots, g_m$ and f are real-valued functions defined on E^n and $x = (x_1, x_2, \ldots, x_n)$. Let

$$S = \{x \mid g_i(x) \geq 0, i = 1,2, \ldots, m\} .$$

S is the set of feasible points and a point $x^* \in S$ which maximizes f over S is an optimal solution and the corresponding number $f^* = f(x^*)$ is the optimal value. Let

$$\hat{S} = \{x \mid g_i(x) > 0, i = 1,2, \ldots, m\} .$$

$\hat{S}$ is called the strict interior of S and only nonlinear programming problems with $\hat{S}$ nonempty will be considered in the sequel.

The method of centers introduced by Huard [15] is in a class of methods which solve nonlinear programming problems with $\hat{S}$ nonempty by solving a sequence of unconstrained problems. The basic idea of this approach is to consider the objective function as an additional constraint, $f(x) \geq f(x^o)$ where $x^o \in \hat{S}$, and to define an auxiliary function called a *distance function* which depends on f , $g_1, \ldots, g_m$ and x^o and is maximized by a point called a *center* in $\hat{S}^1 = \{x \mid f(x) > f(x^o), g_i(x) > 0, i = 1,2, \ldots, m\}$. If this maximization problem is solved then an $x^1 \in \hat{S}^1$ is found such that $f(x^1) > f(x^o)$. The above process is then repeated with x^1 replacing x^o . If this procedure is carried on then under certain additional assumptions an approximation to an optimal solution results. An important property of such a method is that each point generated is a feasible solution and has a better objective value than the previous point.

Examples of distance functions given by Faure and Huard [6] and Huard [15] respectively are:

$$\text{(1.1)} \qquad D(x,\alpha) = \begin{cases} (f(x) - \alpha)^p \prod_{i=1}^{m} g_i(x) & \text{for } x \in \hat{S}(\alpha) \\ 0 & \text{otherwise} \end{cases} \qquad \text{with } p > 0$$

and

$$\text{(1.2)} \qquad D(x,\alpha) = \begin{cases} \min\,[(f(x) - \alpha), g_1(x), \ldots, g_m(x)] & \text{for } x \in \hat{S}(\alpha) \\ 0 & \text{otherwise} \end{cases}$$

where $\hat{S}(\alpha) = \{x \mid f(x) > \alpha, g_i(x) > 0, i = 1,2, \ldots, m\}$ and α is a parameter determined iteratively by a method of centers algorithm. Other examples of distance function which are slight modifications of the above or mixtures of such modifications are given by Tremolières [34]. A method of centers algorithm consists of finding an x^k which approximately maximizes $D(x,\alpha^k)$ where $\alpha^k = f(x^{k-1})$ for $k = 1,2, \ldots$ starting from some $x^o \in \hat{S}$. For $k = 1,2, \ldots$ an ε_k-*center* is a point $x^k \in \hat{S}(\alpha^k)$ such that $D(x^k,\alpha^k) \geqq \tilde{D}^k - \varepsilon_k$ where $\tilde{D}^k$ is the maximum value of $D^k(x,\alpha^k)$ over $\hat{S}(\alpha^k)$ and $\{\varepsilon_k\}$, $k = 1,2, \ldots$ is a sequence of nonnegative numbers converging to zero. For a class of general distance functions Bui-Trong-Lieu and Huard [1] have shown the convergence of $f(x^k)$ to f^* where $\{x^k\}$ is a sequence of ε_k-centers essentially assuming that f is continuous and bounded on S and the closure of $\hat{S}$ is S. Tremolieres [34] has also established this result for a relaxed version of the algorithm where $\alpha^k = \alpha^{k-1} + \rho[f(x^{k-1}) - \alpha^{k-1}]$ with $0 < \rho \leqq 1$ and has given numerical results on several test problems.

The method of centers algorithm based on the minimum function given by (1.2) has been considered also by Kleibohm [18], Pironneau and Polak [28], Polak [29] and Zangwill [35]. This function suffers from a lack of differentiability even when the problem functions are differentiable and for this reason Huard [16] and Pironneau and Polak [28] developed modified algorithms with finite step subproblem procedures based upon this function. Huard's modified algorithm is closely related to a feasible directions algorithm proposed by Topkis and Veinott [33].

The following distance function is essentially the natural logarithm of the function given by (1.1) with $\beta = \frac{1}{p} > 0$.

$$(1.3)\quad D(x,\alpha) = \begin{cases} \ln\,(f(x) - \alpha) + \beta \sum_{i=1}^{m} \ln g_i(x) & \text{for } x \in \hat{S}(\alpha) \\ -\infty & \text{otherwise .} \end{cases}$$

It is similar in behavior to the following "parameter free penalty function" due to Fiacco and McCormick [10].

$$(1.4)\quad D(x,\alpha) = \begin{cases} -\frac{1}{(f(x) - \alpha)} - \sum_{i=1}^{m} \frac{1}{g_i(x)} & \text{for } x \in \hat{S}(\alpha) \\ -\infty & \text{otherwise .} \end{cases}$$

For a class of general distance functions Fiacco and McCormick [11] have shown the existence of a sequence $\{x^k\}$ of local maxima for $D(x,\alpha^k)$ over $\hat{S}(\alpha^k)$ for $k = 1,2, \ldots$ such that accumulation points of $\{x^k\}$ are local maxima for the nonlinear programming problem with objective value v^* assuming the functions $g_1, \ldots, g_m$ and f are continuous and there exists a nonempty isolated compact set of local maxima with local maximum value v^* intersecting the closure of $\hat{S}$. Fiacco [7] has demonstrated a direct relationship between the method of centers and the interior-point penalty function methods of Fiacco and McCormick [11] by showing there are corresponding classes of functions for these methods which give rise to equivalent procedures. The interior-point penalty function related to (1.3) is given by

$$(1.5) \qquad P(x,r) = \begin{cases} f(x) + r \sum_{i=1}^{m} \ln g_i(x) & \text{for } x \in \hat{S} \\ -\infty & \text{otherwise} \end{cases}$$

and the one related to (1.4) is given by

$$(1.6) \qquad P(x,r) = \begin{cases} f(x) - r \sum_{i=1}^{m} \frac{1}{g_i(x)} & \text{for } x \in \hat{S} \\ -\infty & \text{otherwise .} \end{cases}$$

The associated algorithmic procedure consists of sequentially maximizing $P(x,r_k)$ for a decreasing sequence of positive r_k which tends to zero. The function given by (1.5) was first proposed by Frisch [12,13] and later used by Parisot [27] for solving linear programming problems and by Lootsma [21,22] for nonlinear problems. The one given by (1.6) was first proposed by Carroll [2] and extensively developed by Fiacco and McCormick [8,9].

The logarithmic distance function $d^k(x) = D(x,\alpha^k)$ with convergence rate parameter β given by (1.3) will be considered here along with the assumption that $g_1, \ldots, g_m$ and f are continuously differentiable *in order to obtain convergence rate results. The sequence of points* $\{x^k\}$, $k = 1,2, \ldots$ generated by the algorithm is defined by $x^k \in \hat{S}^k = \hat{S}(\alpha^k)$ satisfying $||\nabla d^k(x^k)|| \leq \varepsilon$ for $k = 1,2, \ldots$ where $\varepsilon \geq 0$ is a subproblem termination parameter. For the case when $\epsilon > 0$, if an algorithm used to maximize $d^k(x)$ over $\hat{S}^k$ has the property that any accumulation point $\tilde{x}$ satisfies $\nabla d^k(\tilde{x}) = 0$, then only a finite number of subproblem steps will be required to find x^k. This definition of an approximate center does not depend on the usually unknown maximum value $\bar{D}^k$ used to define an ϵ_k-center.

In Section 2 the logarithmic method of centers algorithm is defined and under differentiability assumptions it is shown that accumulation points of the sequence of feasible points $\{x^k\}$, $k = 1,2, \ldots$ generated by the algorithm satisfy the Fritz John [17] optimality conditions for the nonlinear programming problem. With the addition of pseudo-concavity [25] assumptions on the constraint functions it is shown that the algorithm also generates a bounded multiplier sequence $\left\{\left(u_1^k, u_2^k, \ldots, u_m^k\right)\right\}$, $k = 1,2, \ldots$ such that accumulation points of this sequence and the feasible point sequence satisfy the Kuhn-Tucker [19] optimality conditions. For the special case when $\epsilon = 0$, Lootsma [23] and Fiacco and McCormick [11] have also established this type of result for general classes of differentiable distance functions under concavity assumptions on all the functions. If the objective function is also pseudo-concave then accumulation points of the feasible point sequence are shown to be optimal solutions to the nonlinear programming problem. The relation to Huard's original method of centers algorithm for the case of concave objective and constraint functions is demonstrated by showing that the approximate centers x^k defined here are ϵ_k-centers with respect to the distance function $\exp(d^k(x))$ which is a member of the class of distance functions for which Huard [15] proved under concavity assumptions on all the functions that accumulation points of an ϵ_k-center sequence are optimal solutions.

In Section 3 all functions are assumed to be concave and p^* is defined to be the number of positive components in a Kuhn-Tucker multiplier vector which has the largest number of positive components among such vectors and q^* is defined to be the number of positive constraint values for an optimal solution which has the largest number of positive constraint values among optimal solutions. It is shown that all the accumulation points of the feasible point sequence have the same q^* positive constraints and all the accumulation points of the multiplier vector sequence have the same p^* positive components. It is also shown in general that $f^* - f(x^k)$ is bounded above by a decreasing exponential function of k and for the special case when $q^* = m$ which implies $p^* = 0$ there exists an upper bound which is a product of k fractions where the k^{th} fraction converges to zero as k tends to infinity. For the case when $p^* > 0$ which implies $q^* < m$ it is shown that $f^* - f(x^k)$ and $||x^* - x^k||$ for any optimal point x^* are bounded from below by decreasing exponential functions of k which have the same rates. It is also demonstrated that $g_i(x^k)$ for any i such that $u_i^* > 0$ for some Kuhn-Tucker multiplier vector $\left(u_1^*, u_2^*, \ldots, u_m^*\right)$ and that u_j^k for any j such that $g_j(x^*) > 0$ for some optimal point x^* converge to zero with the same type of convergence bounds as $f^* - f(x^k)$. These results are established in part by finding an upper bound on $\left(\frac{f^* - f(x^k)}{f^* - f(x^{k-1})}\right)$ for all $k \geqq 1$ which for the special case when $\varepsilon = 0$ is equal to $\left(\frac{\beta m}{1 + \beta m}\right)$ and is the same as the bound found under stronger assumptions on the nonlinear programming problem by Funk [5] for linear functions and by Tremolières [34] for general concave functions. Actually Tremolières' bound depends on the relaxation parameter $\rho \in (0,1]$ and is smallest and equals the one obtained here when $\rho = 1$ which is the case of no relaxation. It is also shown here that the sequence $\left\{\left(\frac{f^* - f(x^k)}{f^* - f(x^{k-1})}\right)\right\}$ has all of its accumulation points in the interval $\left[\left(\frac{\beta p^*}{1 + \beta p^*}\right), \left(\frac{\beta(m - q^*)}{1 + \beta(m - q^*)}\right)\right]$. This asymptotic result is independent of the value of the subproblem termination parameter ε and justifies calling β a convergence rate parameter. For the special case when $p^* + q^* = m$ it agrees with the result stated by Faure and Huard [6] and proved for $\varepsilon = 0$ under assumptions which imply the problem has a unique nondegenerate optimal point and Kuhn-Tucker multiplier vector pair by Faure [5] for linear objective and constraint functions and by Lootsma [24] for concave problem functions. Under this uniqueness assumption with exact centers Lootsma found the limit of $\left(\frac{f^* - f(x^k)}{f^* - f(x^{k-1})}\right)$ for a general class of differentiable distance functions and showed that the logarithmic distance function is the only member of this class for which the limit is independent of the value of the Kuhn-Tucker multiplier vector. For the nondifferentiable minimum function defined by (1.2) assuming a unique optimal point and exact centers Pironneau and Polak [28] demonstrated that $\left(\frac{f^* - f(x^k)}{f^* - f(x^{k-1})}\right)$ converges to a fraction with a value depending on the set of Kuhn-Tucker multiplier vectors.

In Section 4 the Lagrangian function $f(x) + \sum_{i=1}^{m} u_i^* g_i(x)$ for some Kuhn-Tucker multiplier vector $\left(u_1^*, u_2^*, \ldots, u_m^*\right)$ is assumed to be strongly concave [20] in a neighborhood of an optimal solution x^*. It is shown that $||x^* - x^k||$ and $|g_i(x^k) - g_i(x^*)|$ for $i = 1,2, \ldots, m$ are bounded above by decreasing exponential functions of k having rates which are one half the rate for the exponential function which bounds $f^* - f(x^k)$ from above.

This result represents a typical way of obtaining a rate for $x^k \to x^*$ given a rate for $f(x^k) \to f^*$. For example Pironneau and Polak [28] established this type of result for their modified method of centers algorithm based upon the minimum function defined by (1.2) under the slightly stronger assumptions of twice continuously differentiable problem functions and f having a negative definite matrix of second partial derivatives in a ball about an optimal point. If in addition to the strongly concave Lagrangian, it is assumed that the first partial derivatives of the objective and constraint functions satisfy Lipschitz conditions, the gradient vectors of the constraint functions which are active at x^* are linearly independent and $u_i^* > 0$ for all constraints i which are active at x^* then it is shown here that the above rates may be improved by a factor of two and that $|u_i^k - u_i^*|$ for $i = 1,2, \ldots, m$ is also bounded above by a decreasing exponential function of k which has the same rate as the one bounding $f^* - f(x^k)$ from above.

The convergence of the method of steepest ascent [3,4,14,29,33,35] on the subproblems for the case when the subproblem termination parameter ϵ is positive is considered in Section 5. The number of steepest ascent steps required to find an approximate center x^k starting from x^{k-1} for each $k \geq 1$ is shown to be bounded above by an increasing function of k. Combined with the results of Section 3 this leads to an upper bounding function of t for the total number of steepest ascent steps required to find a feasible point x^k starting from x^o such that $f^* - f(x^k) \leq t$ where t is a termination parameter for the algorithm.

2. DEFINITION AND GENERAL CONVERGENCE PROPERTIES OF THE ALGORITHM

In order to define the algorithm and establish its convergence properties certain assumptions will be required. The following two conditions will be assumed to hold throughout:

(2.1) There exists an $x^0 \in \hat{S} = \{x \mid g_i(x) > 0 \ , \ i = 1,2, \ldots, m\}$ such that $S^1 = \left\{x \mid f(x) \geqq f^0 \ , \ g_i(x) \geqq 0 \ , \ i = 1,2, \ldots, m\right\}$ is bounded where $f^0 = f(x^0)$.

(2.2) f and g_i for $i = 1,2, \ldots, m$ are continuously differentiable on S^1 .

If $S = \{x \mid g_i(x) \geqq 0 \ , \ i = 1,2, \ldots, m\}$ is a closed convex set, f is a concave and upper semi-continuous function on S and the set of optimal points that maximize f over S is bounded then Topkis [32] has shown that S^1 is bounded. Similar results which imply Assumption (2.1) for $\hat{S}$ non-empty are contained in Rockafellar [30] and Fiacco and McCormick [11]. Assumption (2.1) implies that if x^* is an optimal solution to the nonlinear programming problem and $f^* = f(x^*)$ is the optimal value then $x^* \in S^1$ and $f^* \geqq f^0$.

Define the norm of $y \in E^p$ by

$$||y|| = \left(y_1^2 + y_2^2 + \ldots + y_p^2\right)^{\frac{1}{2}}$$

and define the gradient vector of partial derivatives of a differentiable function d defined on a subset of E^p by

$$\nabla d(y) = \left(\frac{\partial d(y)}{\partial y_1}, \frac{\partial d(y)}{\partial y_2}, \ldots, \frac{\partial d(y)}{\partial y_p}\right) .$$

Algorithm:

Choose numbers $\varepsilon \geqq 0$ and $\beta > 0$. Given $x^{k-1} \varepsilon \hat{S}$ for any integer $k \geqq 1$ terminate the algorithm with x^{k-1} if $\nabla f(x^{k-1}) = 0$. Otherwise define

$$(2.3) \qquad f^{k-1} = f(x^{k-1}) ,$$

$$(2.4) \qquad \hat{S}^k = \left\{x \mid f(x) > f^{k-1} , g_i(x) > 0 , i = 1,2, \ldots, m\right\}$$

and

$$(2.5) \qquad d^k(x) = \ln (f(x) - f^{k-1}) + \beta \sum_{i=1}^{m} \ln g_i(x) \qquad \text{for} \quad x \varepsilon \hat{S}^k$$

and find $x^k \varepsilon \hat{S}^k$ such that

$$(2.6) \qquad ||\nabla d^k(x^k)|| \leqq \varepsilon$$

where by Assumption (2.2)

$$(2.7) \qquad \nabla d^k(x) = \frac{\nabla f(x)}{(f(x) - f^{k-1})} + \beta \sum_{i=1}^{m} \frac{\nabla g_i(x)}{g_i(x)} \qquad \text{for} \quad x \varepsilon \hat{S}^k .$$

It should be noted that a starting point x^0 exists by Assumption (2.1) and if x^k exists for some $k \geqq 1$ then $f^k > f^{k-1}$ and $\hat{S}^{k+1} \subset \hat{S}^k \subset S^1$ by Definitions (2.4) and (2.3). The finding of x^k is to be accomplished by a subroutine which maximizes $d^k(x)$ or, equivalently, $\exp (d^k(x))$ over $\hat{S}^k$. Due to the behavior of $d^k(x)$ at the boundary of $\hat{S}^k$ this subproblem optimization is essentially unconstrained.

The following two lemmas justify the statement of the algorithm. The first lemma shows that if the algorithm does not terminate at x^{k-1} then the next set $\hat{S}^k$ is nonempty.

<u>Lemma 2.1</u>:

If $x^{k-1} \varepsilon \hat{S}$ exists for some $k \geqq 1$ and $\nabla f(x^{k-1}) \neq 0$ then $\hat{S}^k$ is nonempty.

<u>Proof</u>:

Since $x^{k-1} \varepsilon \hat{S}$,

$$g_i(x^{k-1}) > 0 \qquad \text{for } i = 1,2, \ldots, m .$$

Let $x(\lambda) = x^{k-1} + \lambda \nabla f(x^{k-1})$ where λ is a real number. Since g_i for $i = 1,2, \ldots, m$ is continuous on S^1 , there exists a $\bar{\lambda} > 0$ such that for $i = 1,2, \ldots, m$

$$g_i(x(\lambda)) > 0 \qquad \text{for } 0 \leqq \lambda \leqq \bar{\lambda} .$$

Since $\nabla f(x^{k-1}) \neq 0$ there exists $\tilde{\lambda} \varepsilon (0,\bar{\lambda}]$ such that

$$f(x(\lambda)) > f(x^{k-1}) \qquad \text{for } 0 < \lambda \leqq \tilde{\lambda} .$$

Therefore, $\hat{S}^k$ is nonempty.||

If $\nabla f(x^{k-1}) \neq 0$, then Lemma 2.1 shows that $\nabla f(x^{k-1})$ is a feasible direction from x^{k-1} in which to start subproblem k maximization even though ∇d^k is undefined at x^{k-1} . In fact, $\nabla f(x^{k-1})$ multiplied by any positive definite matrix will suffice. The next lemma which is a slight modification of an existence result given by Fiacco and McCormick [10] shows that if $\hat{S}^k$ is nonempty then there exists a point maximizing $d^k(x)$ over $\hat{S}^k$.

<u>Lemma 2.2</u>:

If $\hat{S}^k$ is nonempty for some $k \geqq 1$, then there exists an $\tilde{x} \varepsilon \hat{S}^k$ which maximizes $d^k(x)$ over $\hat{S}^k$ and thus $\nabla d^k(\tilde{x}) = 0$

Proof:

Let $S^k = \left\{x \mid f(x) \geqq f^{k-1}, g_i(x) \geqq 0, i = 1,2, \ldots, m\right\}$. S^k is bounded by Assumption (2.1) since $f^{k-1} \geqq f^o$ implies $S^k \subseteq S^1$ and S^k is nonempty by hypothesis since $\hat{S}^k \subset S^k$. S^k is closed since f and g_i for $i = 1,2, \ldots, m$ are continuous on $S^1 \supseteq S^k$ by Assumption (2.2). Let $D^k(x) = (f(x) - f^{k-1}) \prod_{i=1}^{m} g_i(x)^{\beta}$ and let $\tilde{x}$ maximize the continuous function $D^k(x)$ over the nonempty compact set S^k. Since $D^k(x) > 0$ for $x \varepsilon \hat{S}^k$ and $D^k(x) = 0$ for $x \varepsilon S^k - \hat{S}^k$, $\tilde{x} \varepsilon \hat{S}^k$. Since $d^k(x) = \ln D^k(x)$ and $\tilde{x}$ maximizes $D^k(x)$ over $\hat{S}^k \subset S^k$, $\tilde{x}$ must maximize $d^k(x)$ over $\hat{S}^k$. The continuity of f and g_i for $i = 1,2, \ldots, m$ implies that $\hat{S}^k$ is an open set and Assumption (2.2) implies $d^k(x)$ is differentiable on $\hat{S}^k$. Therefore $\nabla d^k(\tilde{x}) = 0$. ||

If $\varepsilon > 0$ and subproblem k is solved by an unconstrained maximization algorithm which has the property that any accumulation point $\tilde{x}$ generated by it satisfies $\nabla d^k(\tilde{x}) = 0$, then a point x^k such that $||\nabla d^k(x^k)|| \leqq \varepsilon$ will be found in a finite number of subproblem steps since d^k is continuously differentiable on $\hat{S}^k$. For general discussions of unconstrained maximization algorithms which have the above property see Fiacco and McCormick [11], Polak [29], Topkis and Veinott [33] and Zangwill [35].

The next result which is a general property of method of centers algorithms when S^1 is compact and f is continuous on S^1 has been essentially demonstrated by Huard [15].

Lemma 2.3:

Assume the algorithm does not terminate in a finite number of iterations. Then

$$f^k - f^{k-1} > 0 \qquad \text{for } k = 1,2, \ldots \tag{2.8}$$

and

(2.9) $$\lim_{k\to\infty} (f^k - f^{k-1}) = 0 .$$

Proof:

Since $x^k \in \hat{S}^k$, $f^k > f^{k-1}$ by (2.4) and (2.3). The monotone increasing sequence $\{f^k\}$, $k = 1,2, \ldots$ is bounded above since f is continuous on S^1 by Assumption (2.2) and S^1 is closed and bounded by Assumptions (2.1) and (2.2). Then (2.9) follows since there exists an $\bar{f}$ such that

$$\lim_{k\to\infty} f^k = \bar{f} . ||$$

The following theorem shows that accumulation points of the sequence $\{x^k\}$, $k = 1,2, \ldots$ generated by this method of centers algorithm satisfy the Fritz John [17] optimality conditions for the nonlinear programming problem.

Theorem 2.4:

Either the algorithm terminates in a finite number of iterations with a point $x^k \in S$ such that $\nabla f(x^k) = 0$ or the sequence $\{x^k\}$, $k = 1,2, \ldots$ has at least one accumulation point and for each accumulation point $\bar{x}$ there exist multipliers $\bar{v}_i \geqq 0$ for $i = 0,1, \ldots, m$ not all zero such that

(2.10) $$\bar{v}_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{v}_i \nabla g_i(\bar{x}) = 0 ,$$

(2.11) $$\bar{v}_i g_i(\bar{x}) = 0 \qquad \text{for } i = 1,2, \ldots, m .$$

and

$$g_i(\bar{x}) \geqq 0 \qquad \text{for } i = 1,2, \ldots, m . \tag{2.12}$$

Proof:

Either the algorithm terminates in a finite number of iterations with a point $x^k \in \hat{S}^k \subset S^1$ such that $\nabla f(x^k) = 0$ or by Lemmas 2.1 and 2.2 applied inductively the algorithm generates a sequence $\{x^k\}$, $k = 1,2, \ldots$ such that

$$x^k \in \hat{S}^k \subset S^1 , \tag{2.13}$$

$$g_i(x^k) > 0 \qquad \text{for } i = 1,2, \ldots, m \tag{2.14}$$

and

$$||\nabla d^k(x^k)|| \leqq \varepsilon . \tag{2.15}$$

By Assumptions (2.1) and (2.2) S^1 is closed and bounded and therefore by (2.13) $\{x^k\}$, $k = 1,2, \ldots$ has an accumulation point $\bar{x} \in S^1$. Let K_1 be an infinite subset of $\{1,2, \ldots\}$ such that $\lim_{k \in K_1} x^k = \bar{x}$. Then by Assumption (2.2)

$$\lim_{k \in K_1} \nabla f(x^k) = \nabla f(\bar{x}) , \tag{2.16}$$

$$\lim_{k \in K_1} \nabla g_i(x^k) = \nabla g_i(\bar{x}) \qquad \text{for } i = 1,2, \ldots, m , \tag{2.17}$$

$$\lim_{k \in K_1} g_i(x^k) = g_i(\bar{x}) \qquad \text{for } i = 1,2, \ldots, m$$

and by (2.14)

$$g_i(\bar{x}) \geqq 0 \quad \text{for} \quad i = 1,2, \ldots, m$$

which establishes (2.12). Let

(2.18) $$g_o^k = \beta(f^k - f^{k-1}) \ ,$$

(2.19) $$g_i^k = g_i(x^k) \quad \text{for} \quad i = 1,2, \ldots, m$$

and

(2.20) $$h^k = \min \left[g_o^k, g_1^k, \ldots, g_m^k \right] \qquad \text{for} \quad k = 1,2, \ldots \ .$$

Then by (2.8), (2.14), (2.18) and (2.19)

(2.21) $$h^k > 0 \qquad \text{for} \quad k = 1,2, \ldots$$

and by Lemma 2.3

(2.22) $$\lim_{k \to \infty} h^k = 0 \ .$$

Multiplying $\nabla d^k(x^k)$ by h^k and using (2.7) yields

(2.23) $$h^k \nabla d^k(x^k) = \left(\frac{h^k}{f^k - f^{k-1}} \right) \nabla f(x^k) + \sum_{i=1}^{m} \left(\frac{\beta h^k}{g_i^k} \right) \nabla g_i(x^k) \qquad \text{for} \quad k = 1,2, \ldots$$

Let

(2.24) $$v_i^k = \frac{\beta h^k}{g_i^k} \quad \text{for} \quad i = 0,1,2, \ldots, m \quad \text{and} \quad k = 1,2, \ldots \ .$$

Then by (2.20) and (2.21)

$$(2.25) \qquad 0 < v_i^k \leqq \beta \quad \text{for } i = 0,1,2, \ldots, m \text{ and } k = 1,2, \ldots .$$

By (2.23), (2.24) and (2.18)

$$(2.26) \qquad h^k \nabla d^k(x^k) = v_o^k \nabla f(x^k) + \sum_{i=1}^{m} v_i^k \nabla g_i(x^k) \qquad \text{for } k = 1,2, \ldots .$$

Choose an infinite subset $K_2 \subset K_1$ such that

$$(2.27) \qquad \lim_{k \in K_2} v_i^k = \bar{v}_i \qquad \text{for } i = 0,1, \ldots, m$$

which is possible by (2.25). Then choose $K_3 \subset K_2$ such that for some $j \in \{0,1, \ldots, m\}$

$$(2.28) \qquad h^k = g_j^k \qquad \text{for all } k \in K_3 .$$

This is possible since there are a finite number of indices i and at least one must identify the minimal g_i^k infinitely often. By (2.24) and (2.28), $v_j^k = \beta$ for all $k \in K_3$ and, therefore,

$$(2.29) \qquad \bar{v}_j = \beta .$$

By (2.16), (2.17) and (2.27)

$$(2.30) \qquad \lim_{k \in K_2} \left[v_o^k \nabla f(x^k) + \sum_{i=1}^{m} v_i^k \nabla g_i(x^k) \right] = \bar{v}_o \nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{v}_i \nabla g_i(\bar{x}) .$$

By (2.15) and (2.22)

$$(2.31) \qquad \lim_{k\to\infty} t^k u^k(x^k) = 0 .$$

Therefore by (2.26), (2.30) and (2.31)

$$\bar{v}_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{v}_i \nabla g_i(\bar{x}) = 0$$

which establishes (2.10). If $g_i(\bar{x}) > 0$ for some $i \in \{1,2,\ldots,m\}$, then $\bar{v}_i = 0$ by (2.24), (2.22) and (2.27) since $\lim_{k\in K_2} g_i(x^k) > 0$. This establishes (2.11). By (2.25) $\bar{v}_i \geqq 0$ for $i = 0,1,\ldots,m$ and the multipliers are not all zero by (2.29) which completes the proof.||

Under stronger assumptions on the constraint functions the algorithm generates a bounded multiplier sequence $\left\{\left(u_1^k, u_2^k, \ldots, u_m^k\right)\right\}$, $k = 1,2,\ldots$ for which the combined sequence $\left\{\left(x^k, u_1^k, u_2^k, \ldots, u_m^k\right)\right\}$, $k = 1,2,\ldots$ has accumulation points satisfying the Kuhn-Tucker [19] optimality conditions for the nonlinear programming problem.

<u>Definition:</u>

A real-valued function g is *pseudo-concave* [25] on a convex set $T \subset E^n$ if g is differentiable on T and $\nabla g(y)\cdot(x - y) \leqq 0$ for $x,y \in T$ implies $g(x) \leqq g(y)$. It can be shown that a differentiable concave function is pseudo-concave and that pseudo-concave functions have the property that local maxima are global maxima.

Combining the results of Theorem 2.4 with pseudo-concavity assumptions on the constraint functions and defining $u_i^k = \dfrac{\epsilon(t^k - t^{k-1})}{g_i(x^k)}$ for $i = 1,2,\ldots,m$ and $k = 1,2,\ldots$ yields the following theorem.

Theorem 2.5.

Assume that g_i for $i = 1,2, \ldots, m$ are pseudo-concave on a convex set containing S^1 and that the algorithm does not terminate in a finite number of iterations. Then there exists a positive number b such that

$$0 < u_i^k = \frac{\varepsilon(f^k - f^{k-1})}{g_i(x^k)} \leq b \qquad \text{for } i = 1,2, \ldots, m \text{ and } k = 1,2, \ldots . \tag{2.32}$$

Furthermore the combined sequence $\left\{\left(x^k, u_1^k, u_2^k, \ldots, u_m^k\right)\right\}$, $k = 1,2, \ldots$ has at least one accumulation point and each accumulation point $(\bar{x}, \bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m)$ satisfies the following conditions.

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \nabla g_i(\bar{x}) = 0 \, . \tag{2.33}$$

$$\bar{u}_i g_i(\bar{x}) = 0 \qquad \text{for } i = 1,2, \ldots, m \, , \tag{2.34}$$

$$\bar{u}_i \geqq 0 \qquad \text{for } i = 1,2, \ldots, m \tag{2.35}$$

and

$$g_i(\bar{x}) \geqq 0 \qquad \text{for } i = 1,2, \ldots, m \, . \tag{2.36}$$

Proof:

Let v_i^k for $i = 0,1, \ldots, m$ and $k = 1,2, \ldots$ be as in the proof of Theorem 2.4. If $\liminf_{k\to\infty} v_o^k = 0$ then there exists an infinite subset $K_o \subset \{1,2, \ldots\}$ with $\lim_{k\in K_o} v_i^k = \bar{v}_i$ for $i = 0,1, \ldots, m$ such that $\bar{v}_o = 0$. Choose $K_1 \subset K_o$ with $\lim_{k\in K_1} x^k = \bar{x}$. Then (2.10) reduces to

(2.37) $$\sum_{i=1}^{m} \bar{v}_i \nabla g_i(\bar{x}) = 0 .$$

Let $y \in \hat{S}$ which is nonempty by Assumption (2.1). Then $g_i(y) > 0$ for $i = 1,2, \ldots, m$. If $\bar{v}_i > 0$ then $g_i(\bar{x}) = 0$ by (2.11). Therefore

$$g_i(y) > g_i(\bar{x}) \quad \text{for all } i \text{ such that } \bar{v}_i > 0 .$$

Since $y,\bar{x} \in S^1$ and g_i for $i = 1,2, \ldots, m$ are pseudo-concave on a convex set containing S^1 ,

(2.38) $$\nabla g_i(\bar{x})\cdot(y - \bar{x}) > 0 \quad \text{for all } i \text{ such that } \bar{v}_i > 0 .$$

Since $\bar{v}_o = 0$ and not all the $\bar{v}_i$ are zero in Theorem 2.4, it must be true that $\bar{v}_i > 0$ for some $i \geqq 1$. Therefore by (2.38)

$$\sum_{i=1}^{m} \bar{v}_i \nabla g_i(\bar{x})\cdot(y - \bar{x}) > 0$$

which contradicts (2.37). Therefore $\liminf_{k\to\infty} v_o^k > 0$, and since $v_o^k > 0$ for $k = 1,2, \ldots$ there exists a positive number a such that $v_o^k \geqq a$ for $k = 1,2, \ldots$. By the definition in (2.32) and (2.24)

(2.39) $$u_i^k = \frac{\beta(f^k - f^{k-1})}{g_i^k} = \frac{g_o^k}{g_i^k} = \frac{v_i^k}{v_o^k} \quad \text{for } i = 1,2, \ldots, m \text{ and } k = 1,2, \ldots .$$

Therefore by (2.25)

$$0 < u_i^k \leqq \frac{\beta}{a} \quad \text{for } i = 1,2, \ldots, m \text{ and } k = 1,2, \ldots$$

Letting $b = \frac{\beta}{a}$ establishes the upper bound of (2.32). Now let

$$\lim_{k\varepsilon K_1} \left(x^k, u_1^k, u_2^k, \ldots, u_m^k\right) = (\bar{x}, \bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m)$$

which is possible by Theorem 2.4 and relation (2.32). Choose $K_2 \subset K_1$ such that $\lim_{k\varepsilon K_2} v_i^k = \bar{v}_i$ for $i = 0,1, \ldots, m$. Then $\bar{x}$ and $\bar{v}_i \geq 0$ for $i = 0,1, \ldots, m$ satisfy (2.10) to (2.12) with $\bar{v}_o > 0$. By (2.39) $\bar{u}_i = \lim_{k\varepsilon K_2} u_i^k = \frac{\bar{v}_i}{\bar{v}_o}$ for $i = 1,2, \ldots, m$ and therefore $(\bar{x}, \bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m)$ satisfy (2.33) to (2.36).||

The assumptions that the feasible set has a nonempty strict interior and the constraint functions are pseudo-concave constitute Slater's weak constraint qualification [26] for the nonlinear programming problem. For the case when $\varepsilon = 0$ the results of Theorem 2.5 have been obtained by Lootsma [23] and Fiacco and McCormick [11] under concavity assumptions on the functions f and g_i for $i = 1,2, \ldots, m$.

For reference in the sequel a vector $\bar{u} = (\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m) \varepsilon E^m$ which satisfies relations (2.33), (2.34) and (2.35) for some $\bar{x} \varepsilon S$ will be called a Kuhn-Tucker multiplier vector.

In order to show that accumulation points of $\{x^k\}$, $k = 1,2, \ldots$ are optimal solutions to the nonlinear programming problem an additional assumption on the objective function f will be required.

Theorem 2.6:

Assume that f and g_i for $i = 1,2, \ldots, m$ are pseudo-concave on a convex set containing S^1. Either the algorithm terminates in a finite number of iterations with an optimal solution to the nonlinear programming problem or every accumulation point $\bar{x}$ of the sequence $\{x^k\}$, $k = 1,2, \ldots$ is an optimal solution.

Proof:

Under the above assumptions the Kuhn-Tucker conditions, (2.33) through (2.36) of Theorem 2.5, are sufficient to imply optimality by Theorem 10.1.1 of Mangasarian [26].||

The above result was first established by Huard [15] under concavity assumptions on the objective and constraint functions. For the distance function $D^k(x) = \exp(d^k(x))$ Huard's algorithm is to find ε_k-centers y^k such that $D^k(y^k) \geqq D^k(\tilde{x}^k) - \varepsilon_k$ for $k = 1,2,\ldots$ where $\tilde{x}^k$ maximizes $D^k(x)$ over $\hat{S}^k$ and $\{\varepsilon_k\}$ is a sequence of nonnegative numbers converging to zero. The following analysis will show that the sequence $\{x^k\}$, $k = 1,2,\ldots$ generated by the algorithm discussed here is a sequence of ε_k-centers if f and g_i for $i = 1,2,\ldots,m$ are concave functions on a convex set containing S^1. By the mean value theorem for all $k \geqq 1$

(2.40) $$D^k(\tilde{x}^k) - D^k(x^k) = \nabla D^k(\xi^k)\cdot(\tilde{x}^k - x^k)$$

where

(2.41) $$\xi^k = x^k + \lambda^k(\tilde{x}^k - x^k) \quad \text{and} \quad 0 < \lambda^k < 1 .$$

From the concavity assumptions it is easy to see that $d^k(x)$ is a concave function on the convex set $\hat{S}^k$. Then

$$(\nabla d^k(\xi^k) - \nabla d^k(x^k))\cdot(\xi^k - x^k) \leqq 0$$

which implies since $\lambda^k > 0$

$$\left(\frac{1}{\lambda^k}\right)\nabla d^k(\xi^k)\cdot(\xi^k - x^k) \leqq \left(\frac{1}{\lambda^k}\right)\nabla d^k(x^k)\cdot(\xi^k - x^k)$$

or by (2.41)

$$\nabla d^k(\xi^k)\cdot(\tilde{x}^k - x^k) \leqq \nabla d^k(x^k)\cdot(\tilde{x}^k - x^k) \ . \tag{2.42}$$

By (2.40), the definition of $D^k(x)$ and (2.42)

$$D^k(\tilde{x}^k) - D^k(x^k) = D^k(\xi^k)\nabla d^k(\xi^k)\cdot(\tilde{x}^k - x^k) \leqq D^k(\xi^k)\nabla d^k(x^k)\cdot(\tilde{x}^k - x^k)$$

which implies by the definition of $\tilde{x}^k$ and the Cauchy-Schwarz inequality

$$D^k(\tilde{x}^k) - D^k(x^k) \leqq D^k(\tilde{x}^k)||\nabla d^k(x^k)||\ ||\tilde{x}^k - x^k|| \ .$$

Then since $||\nabla d^k(x^k)|| \leqq \varepsilon$

$$D^k(\tilde{x}^k) - D^k(x^k) \leqq D^k(\tilde{x}^k)\varepsilon\gamma$$

where $\gamma = \sup_{x,y\in S^1} ||x - y||$. Defining $\varepsilon_k = D^k(\tilde{x}^k)\varepsilon\gamma$ for all $k \geqq 1$ yields

$$D^k(x^k) \geqq D^k(\tilde{x}^k) - \varepsilon_k$$

nd

$$\lim_{k\to\infty} \varepsilon_k = 0$$

since $D^k(\tilde{x}^k) = (f(\tilde{x}^k) - f^{k-1}) \prod_{i=1}^{m} g_i(\tilde{x}^k)^\beta$, g_i for $i = 1,2, \ldots, m$ is continuous on the compact set S^1 , $f^{k-1} < f(\tilde{x}^k) \leqq f^*$ and $\lim_{k\to\infty} f^{k-1} = f^*$ by Theorem 2.6. Thus, x^k is an ε_k-center for each $k \geqq 1$, but here the definition of an approximate center x^k does not depend on the unknown maximum value $D^k(\tilde{x}^k)$.

These concavity assumptions will be used in the next section to derive convergence rate results.

3. CONVERGENCE RATE RESULTS REQUIRING CONCAVITY

For purposes of establishing convergence rate results the following condition in addition to Assumptions (2.1) and (2.2) will be assumed to hold.

(3.1) f and g_i for $i = 1,2, \ldots, m$ are concave functions on a convex set containing S^1.

It should be noted that this assumption implies that S^1 is a convex set and together with (2.2) implies that f and g_i for $i = 1,2, \ldots, m$ are pseudo-concave functions on S^1.

It will also be assumed throughout the sequel that the algorithm does not terminate in a finite number of iterations so that a feasible point sequence $\{x^k\}$, $k = 1,2, \ldots$ and a multiplier sequence $\left\{\left(u_1^k, u_2^k, \ldots, u_m^k\right)\right\}$, $k = 1,2, \ldots$ as defined in Section 2 are generated. The stronger assumption that $\nabla f(x) \neq 0$ for all $x \in S^1$ will be explicitly stated where needed for additional results.

The following lemma is a direct consequence of the concavity and differentiability of the problem functions.

Lemma 3.1:

For $k = 1,2, \ldots$

$$f(x) - f^k \leqq (f^k - f^{k-1})\left[\beta m - \beta \sum_{i=1}^{m} \frac{g_i(x)}{g_i(x^k)} + \varepsilon||x - x^k||\right] \quad \text{for all } x \in S^1.$$

Proof:

By the concavity and differentiability of f and g_i for $i = 1,2, \ldots, m$ on S^1

$$(3.2) \qquad f(x) \leq f(x^k) + \nabla f(x^k)\cdot(x - x^k)$$

and

$$(3.3) \qquad g_i(x) \leqq g_i(x^k) + \nabla g_i(x^k)\cdot(x - x^k) \qquad \text{for } i = 1,2, \ldots, m$$

for all $x \in S^1$ since $x^k \in S^1$ for $k \geqq 1$. Multiplying the i^{th} inequality of (3.3) by $\left(\frac{\beta(f^k - f^{k-1})}{g_i(x^k)}\right) > 0$ and adding the resultant inequalities to (3.2) yields

$$f(x) + \sum_{i=1}^{m} \left(\frac{\beta(f^k - f^{k-1})}{g_i(x^k)}\right) g_i(x) \leqq f(x^k) + \beta(f^k - f^{k-1})m +$$

$$+ \left[\nabla f(x^k) + \sum_{i=1}^{m} \left(\frac{\beta(f^k - f^{k-1})}{g_i(x^k)}\right)\nabla g_i(x^k)\right]\cdot(x - x^k) \qquad \text{for all } x \in S^1$$

which is equivalent to

$$f(x) - f^k \leqq \beta(f^k - f^{k-1})m - \beta(f^k - f^{k-1}) \sum_{i=1}^{m} \frac{g_i(x)}{g_i(x^k)} +$$

$$+ (f^k - f^{k-1})\nabla d^k(x^k)\cdot(x - x^k) \qquad \text{for all } x \in S^1$$

since

$$\nabla d^k(x^k) = \frac{\nabla f(x^k)}{(f^k - f^{k-1})} + \beta \sum_{i=1}^{m} \frac{\nabla g_i(x^k)}{g_i(x^k)} .$$

The result then follows since

$$\nabla d^k(x^k)\cdot(x - x^k) \leqq ||\nabla d^k(x^k)||\ ||x - x^k|| \leqq \varepsilon||x - x^k||$$

by the Cauchy-Schwartz inequality and the definition of x^k for $k = 1,2, \ldots$. ||

A well-known [19] consequence of the concavity of the problem functions is the following:

(3.4) $$f^* - f(x) \geqq \sum_{i=1}^{m} u_i^* g_i(x) \qquad \text{for all } x \in S^1$$

where $u^* = \left(u_1^*, \ldots, u_m^*\right)$ is any Kuhn-Tucker multiplier vector associated with an optimal solution to the nonlinear programming problem and f^* is the optimal objective value.

By combining Lemma 3.1 with the above result bounds on $f^* - f^k$ can be obtained. The following lemma is the key lemma from which most of the results of this section are derived. It will require some preliminary definitions which will be used throughout the sequel. Let X^* be the set of optimal solutions to be nonlinear programming problem and U^* be the set of Kuhn-Tucker multiplier vectors associated with optimal solutions. Let

(3.5) $$\gamma = \sup_{x,y \in S^1} ||x - y||$$

which is a finite number since S^1 is assumed to be bounded.

Lemma 3.2:

Let $x^* \in X^*$ and $u^* \in U^*$. Then

$$0 \leqq \sum_{i=1}^{m} \frac{u_i^*}{u_i^k} \leqq \frac{(f^* - f^k)}{\beta(f^k - f^{k-1})} \leqq m - \sum_{i=1}^{m} \frac{g_i(x^*)}{g_i(x^k)} + \left(\frac{\varepsilon}{\beta}\right)||x^* - x^k|| \leqq$$

$$\leqq m + \left(\frac{\varepsilon}{\beta}\right)\gamma \qquad \text{for } k = 1,2, \ldots .$$

Proof:

Any optimal solution x^* is in S^1 and, therefore, by the result of Lemma 3.1 with $x = x^*$

$$\frac{(f(x^*) - f^k)}{\beta(f^k - f^{k-1})} \leqq m - \sum_{i=1}^{m} \frac{g_i(x^*)}{g_i(x^k)} + \left(\frac{\varepsilon}{\beta}\right)||x^* - x^k|| .$$

Furthermore, since $x^k \varepsilon S^1$ for $k = 1,2, \ldots$

$$- \sum_{i=1}^{m} \frac{g_i(x^*)}{g_i(x^k)} + \left(\frac{\varepsilon}{\beta}\right)||x^* - x^k|| \leqq \left(\frac{\varepsilon}{\beta}\right)\gamma \qquad \text{for } k = 1,2, \ldots$$

by the definitions of S^1 and γ . Thus, the last two of the desired inequalities are established. From (3.4) with $x = x^k$ for $k = 1,2, \ldots$

$$f(x^*) - f^k \geqq \beta(f^k - f^{k-1}) \sum_{i=1}^{m} \frac{u_i^*}{u_i^k}$$

since

$$u_i^k = \frac{\beta(f^k - f^{k-1})}{g_i(x^k)} > 0 \qquad \text{for } i = 1,2, \ldots, m .$$

These last two relations establish the first two desired inequalities.||

This lemma shows that the convergence of $f^* - f^k$ to zero is at least as fast as $f^k - f^{k-1}$ which converges to zero by Lemma 2.3.

The next lemma which gives a basic convergence result also requires some preliminary definitions. Let $q(x)$ be the number of indices $i \varepsilon \{1,2, \ldots, m\}$ such that $g_i(x) > 0$ for $x \varepsilon S$ and $p(u)$ be the number of indices $i \in \{1,2, \ldots, m\}$ such that $u_i > 0$ for $u = (u_1,u_2, \ldots, u_m) \geqq 0$. Define

$$(3.6) \qquad q^* = \max_{x \in X^*} q(x)$$

and

$$(3.7) \qquad p^* = \max_{u \in U^*} p(u) \; .$$

It should be noted that if $u^* \in U^*$ and $x^* \in X^*$ then (x^*,u^*) satisfy the Kuhn-Tucker conditions and $p(u^*) + q(x^*) \leqq m$ since $u_i^* g_i(x^*) = 0$, $u_i^* \geqq 0$ and $g_i(x^*) \geqq 0$ for $i = 1,2, \ldots, m$. If $p(u^*) + q(x^*) = m$, then the pair (x^*,u^*) is said to be nondegenerate.

<u>Lemma 3.3</u>:

For $k = 1,2, \ldots$ and $i = 1,2, \ldots, m$

$$(3.8) \qquad g_i(x^k) \geqq \left(\frac{1}{m + \left(\frac{\varepsilon}{\beta}\right)\gamma}\right) \sup_{x \in X^*} g_i(x)$$

and

$$(3.9) \qquad u_i^k \geqq \left(\frac{1}{m + \left(\frac{\varepsilon}{\beta}\right)\gamma}\right) \sup_{u \in U^*} u_i \; ;$$

and if $\bar{x}$ is an accumulation point of the sequence $\{x^k\}$, $k = 1,2, \ldots$ and $\bar{u} = (\bar{u}_1,\bar{u}_2, \ldots, \bar{u}_m)$ is an accumulation point of the sequence $\left\{\left(u_1^k,u_2^k, \ldots, u_m^k\right)\right\}$, $k = 1,2, \ldots$, then

$$(3.10) \qquad q(\bar{x}) = q^*$$

and

$$(3.11) \qquad p(\bar{u}) = p^* .$$

Proof:

The results of Lemma 3.2 imply that for any $x^* \in X^*$ and any $u^* \in U^*$

$$\sum_{i=1}^{m} \left(\frac{u_i^*}{u_i^k} + \frac{g_i(x^*)}{g_i(x^k)} \right) \leqq m + \left(\frac{\varepsilon}{\beta}\right)\gamma \qquad \text{for } k = 1,2, \ldots .$$

Then (3.8) and (3.9) follow immediately from this inequality. From (3.8) and (3.9) and Definitions (3.6) and (3.7)

$$q(\bar{x}) \geqq q^*$$

and

$$p(\bar{u}) \geqq p^* .$$

Furthermore,

$$q(\bar{x}) \leqq q^*$$

and

$$p(\bar{u}) \leqq p^*$$

since $\bar{x} \in X^*$ and $\bar{u} \in U^*$ by Theorems 2.5 and 2.6. Thus, (3.10) and (3.11) are established.||

This lemma combined with Theorems 2.5 and 2.6 shows that there are q^* constraint indices i satisfying $\liminf_{k\to\infty} g_i(x^k) > 0$ and $\lim_{k\to\infty} u_i^k = 0$,

p^* constraint indices i satisfying $\lim\limits_{k\to\infty} g_i(x^k) = 0$ and $\liminf\limits_{k\to\infty} u_i^k > 0$ and $m - q^* - p^*$ constraint indices i satisfying $\lim\limits_{k\to\infty} g_i(x^k) = 0$ and $\lim\limits_{k\to\infty} u_i^k = 0$.

The next lemma combines the results of Lemmas 3.2 and 3.3 to show that the sequence $\left\{\frac{(f^* - f^k)}{\beta(f^k - f^{k-1})}\right\}$, $k = 1,2, \ldots$ has accumulation points in the interval $[p^*, m - q^*]$ and if there exists a nondegenerate optimal solution and Kuhn-Tucker multiplier vector pair, then the limiting value is $p^* = m - q^*$. Note that X^* is bounded by Assumption (2.1) and U^* is bounded by Lemma 3.3. Define

$$p_k = \sup_{u\in U^*} \left[\sum_{i=1}^{m} \frac{u_i}{u_i^k}\right] \qquad \text{for } k = 1,2, \ldots \tag{3.12}$$

and

$$s_k = \inf_{x\in X^*} \left[m - \sum_{i=1}^{m} \frac{g_i(x)}{g_i(x^k)} + \left(\frac{\varepsilon}{\beta}\right)\|x - x^k\|\right] \qquad \text{for } k = 1,2, \ldots . \tag{3.13}$$

<u>Lemma 3.4</u>:

For $k = 1,2, \ldots$

$$p_k \leq \frac{(f^* - f^k)}{\beta(f^k - f^{k-1})} \leq s_k , \tag{3.14}$$

$$p^* \leq \liminf_{k\to\infty} p_k \leq \limsup_{k\to\infty} p_k \leq m - q^* , \tag{3.15}$$

and

$$(3.16)\qquad p^* \leqq \liminf_{k\to\infty} s_k \leqq \limsup_{k\to\infty} s_k \leqq m - q^* .$$

Furthermore, if there exists an $x^* \varepsilon X^*$ and a $u^* = \left(u_1^*, u_2^*, \ldots, u_m^*\right) \varepsilon U^*$ such that $p(u^*) + q(x^*) = m$ then

$$(3.17)\qquad \lim_{k\to\infty} p_k = \lim_{k\to\infty} \frac{(f^* - f^k)}{\beta(f^k - f^{k-1})} = \lim_{k\to\infty} s_k = m - q^* = p^* .$$

<u>Proof:</u>

Relation (3.14) follows immediately from Lemma 3.2 and Definitions (3.12) and (3.13). Let $\bar{x}$ be any accumulation point of the sequence $\{x^k\}$, $k = 1,2, \ldots$ and $\bar{u} = (\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m)$ be any accumulation point of the sequence $\left\{\left(u_1^k, u_2^k, \ldots, u_m^k\right)\right\}$, $k = 1,2, \ldots$. By Definitions (3.12) and (3.13)

$$p_k \geqq \sum_{i=1}^{m} \frac{\bar{u}_i}{u_i^k} \qquad \text{for } k = 1,2, \ldots$$

and

$$s_k \leqq m - \sum_{i=1}^{m} \frac{g_i(\bar{x})}{g_i(x^k)} + \left(\frac{\varepsilon}{\beta}\right)||\bar{x} - x^k|| \qquad \text{for } k = 1,2, \ldots$$

since for any such $\bar{u}$ and $\bar{x}$, $\bar{u} \varepsilon U^*$ and $\bar{x} \varepsilon X^*$ by Theorems 2.5 and 2.6. Let K_1 be an infinite subset of $\{1,2, \ldots\}$ such that $\lim_{k\varepsilon K_1} p_k = \bar{p}$ and choose $K_2 \subset K_1$ such that $\lim_{k\varepsilon K_2} u^k = \bar{u}$. Then by Lemma 3.3

$$\bar{p} = \lim_{k\varepsilon K_2} p_k \geqq \lim_{k\varepsilon K_2} \sum_{i=1}^{m} \frac{\bar{u}_i}{u_i^k} = p(\bar{u}) = p^* .$$

Therefore, $\liminf_{k\to\infty} p_k \geqq p^*$. Now let K_3 be an infinite subset of $\{1,2, \ldots\}$ such that $\lim_{k\in K_3} s_k = \bar{s}$ and choose $K_4 \subset K_3$ such that $\lim_{k\in K_4} x^k = \bar{x}$. Then by Lemma 3.3

$$\bar{s} = \lim_{k\in K_4} s_k \leqq \lim_{k\in K_4} \left[m - \sum_{i=1}^{m} \frac{g_i(\bar{x})}{g_i(x^k)} + \left(\frac{\varepsilon}{\beta}\right) ||\bar{x} - x^k|| \right] =$$

$$= m - q(\bar{x}) = m - q^* .$$

Therefore, $\limsup_{k\to\infty} s_k \leqq m - q^*$ which together with $\liminf_{k\to\infty} p_k \geqq p^*$ and (3.14) implies (3.15) and (3.16). Now suppose there exists an $x^* \varepsilon X^*$ and a $u^* \varepsilon U^*$ such that $p(u^*) + q(x^*) = m$. Then by the definitions of p^* and q^*

$$p^* + q^* \geqq m .$$

But by the remarks preceding Lemma 3.3

$$p^* + q^* \leqq m .$$

Therefore, $p^* + q^* = m$ which together with (3.14), (3.15) and (3.16) establishes the final result (3.17).||

It should be noted that if $q^* = m$ then this lemma implies that

$$\lim_{k\to\infty} \frac{(f^{k+1} - f^k)}{(f^k - f^{k-1})} = \lim_{k\to\infty} \frac{(f^* - f^k)}{(f^k - f^{k-1})} = 0 \quad \text{since} \quad f^{k+1} \leqq f^* \quad \text{for all} \quad k \geqq 0 .$$

To show that $(f^* - f^k)$ does not converge to zero any faster than $(f^k - f^{k-1})$ requires the existence of a positive number $\tilde{p}$ which bounds p_k below for all k . This in turn requires an assumption which implies

$p^* > 0$ and in order to obtain an expression for $\tilde{p}$ requires upper bounds on the multiplier values u_i^k for $i = 1,2, \ldots, m$ and $k = 1,2, \ldots$. The next lemma which follows from Lemma 3.1 provides these bounds along with lower bounds and upper and lower bounds on the constraint function values $g_i(x^k)$ for $i = 1,2, \ldots, m$ and $k = 1,2, \ldots$. Define

$$\tilde{g}_i = \sup_{x \in S^1} g_i(x) \qquad \text{for } i = 1,2, \ldots, m \tag{3.18}$$

and

$$\tilde{u}_i = (1 + \beta m + \varepsilon\gamma)\left(\frac{f^* - f^o}{g_i(x^o)}\right) \qquad \text{for } i = 1,2, \ldots, m . \tag{3.19}$$

<u>Lemma 3.5</u>:

For $k = 1,2, \ldots$ and $i = 1,2, \ldots, m$

$$0 < \frac{\beta(f^k - f^{k-1})}{\tilde{u}_i} \leqq g_i(x^k) \leqq \tilde{g}_i \tag{3.20}$$

and

$$0 < \frac{\beta(f^k - f^{k-1})}{\tilde{g}_i} \leqq u_i^k \leqq \tilde{u}_i . \tag{3.21}$$

<u>Proof</u>:

Since S^1 is assumed to be compact and g_i for $i = 1,2, \ldots, m$ are assumed to be continuous on S^1 the quantities $\tilde{g}_i$ defined by (3.18) are finite numbers and the upper bound of (3.20) follows immediately since $x^k \in S^1$ for all $k \geqq 1$. Since $x^o \in S^1$, Lemma 3.1 implies

$$f^{o} - f^{k} \leqq (f^{k} - f^{k-1})\left[\beta m - \beta \sum_{j=1}^{m} \frac{g_j(x^o)}{g_j(x^k)} + \varepsilon||x^o - x^k||\right]$$

for $k = 1,2, \ldots$.

Rearranging this expression yields

$$\sum_{j=1}^{m} \frac{g_j(x^o)}{g_j(x^k)} \leqq \left(\frac{1}{\beta(f^k - f^{k-1})}\right)\Big[(f^k - f^o) + (f^k - f^{k-1})\cdot$$

$$\cdot(\beta m + \varepsilon||x^o - x^k||)\Big] \qquad \text{for } k = 1,2, \ldots .$$

Then for $i = 1,2, \ldots, m$ and $k = 1,2, \ldots$

(3.22) $$\frac{g_i(x^o)}{g_i(x^k)} \leqq \sum_{j=1}^{m} \frac{g_j(x^o)}{g_j(x^k)} \leqq \left(\frac{1}{\beta(f^k - f^{k-1})}\right)\Big[(f^* - f^o)(1 + \beta m + \varepsilon\gamma)\Big]$$

since $f^* > f^k > f^{k-1} \geqq f^o$ and $||x^o - x^k|| \leqq \gamma$ for all $k \geqq 1$. Then the remaining bounds of (3.20) follow from (3.22) and (3.19). Since

$$u_i^k = \frac{\beta(f^k - f^{k-1})}{g_i(x^k)} \qquad \text{for } i = 1,2, \ldots, m \text{ and } k = 1,2, \ldots ,$$

(3.21) follows from (3.20).||

The existence of upper bounds for all the multiplier values u_i^k has been shown in Theorem 2.5 under pseudo-concavity assumptions on the constraint functions. Here the stronger concavity assumptions of this section specify these bounds. The next corollary uses these bounds to provide a lower bound on the sequence $\{p_k\}$, $k = 1,2, \ldots$. Define

$$(3.23) \qquad \tilde{p} = \sup_{u \in U^*} \sum_{i=1}^{m} \frac{u_i}{\tilde{u}_i}$$

The following is immediate from (3.12) and (3.21).

Corollary 3.6:

For $k = 1,2, \ldots$

$$p_k \geqq \sup_{u \in U^*} \sum_{i=1}^{m} \frac{u_i}{\tilde{u}_i} = \tilde{p} .$$

The next lemma gives a sufficient condition for the existence of a positive number $\tilde{p}$ which bounds $\frac{(f^* - f^k)}{\beta(f^k - f^{k-1})}$ from below for all k.

Lemma 3.7:

If $\nabla f(x) \neq 0$ for all $x \in S^1$, then

$$0 < \tilde{p} \leqq \frac{(f^* - f^k)}{\beta(f^k - f^{k-1})} \qquad \text{for } k = 1,2, \ldots .$$

Proof:

Choose $u^* \in U^*$ and $x^* \in X^*$. Then

$$\nabla f(x^*) + \sum_{i=1}^{m} u_i^* \nabla g_i(x^*) = 0 .$$

Since $\nabla f(x^*) \neq 0$ there exists an $i \in \{1,2, \ldots, m\}$ such that $u_i^* > 0$.

Therefore, $\tilde{p} = \sup_{u \in U^*} \sum_{i=1}^{m} \frac{u_i}{\tilde{u}_i} > 0$ and the desired result follows from Corollary 3.6 and Lemma 3.4. ||

The results of Lemma 3.4 may be used to find upper and lower bounds on the ratio $\frac{(f^* - f^k)}{(f^* - f^{k-1})}$ for all k. This result provides an objective value convergence rate for the algorithm.

Theorem 3.8:

For $k = 1,2, \ldots$

$$\frac{\beta p_k}{1 + \beta p_k} \leqq \frac{(f^* - f^k)}{(f^* - f^{k-1})} \leqq \frac{\beta s_k}{1 + \beta s_k} \leqq \frac{\beta m + \varepsilon\gamma}{1 + \beta m + \varepsilon\gamma} .$$

Proof:

$$\frac{(f^* - f^k)}{(f^* - f^{k-1})} = \frac{(f^* - f^k)}{(f^* - f^k + f^k - f^{k-1})}$$

$$= \frac{1}{1 + \frac{(f^k - f^{k-1})}{(f^* - f^k)}} \qquad \text{for } k = 1,2, \ldots .$$

Then from (3.14) when $p_k > 0$

$$\frac{(f^* - f^k)}{(f^* - f^{k-1})} \geqq \frac{1}{1 + \frac{1}{\beta p_k}} \qquad \text{for } k = 1,2, \ldots$$

and from (3.14) and Lemma 3.2

$$\frac{(f^* - f^k)}{(f^* - f^{k-1})} \leqq \frac{1}{1 + \frac{1}{\beta s_k}} \leqq \frac{1}{1 + \frac{1}{\beta m + \varepsilon\gamma}} \qquad \text{for } k = 1,2, \ldots . ||$$

For the case when $\varepsilon = 0$ the upper bound result $\frac{(f^* - f^k)}{(f^* - f^{k-1})} \leqq \frac{\beta m}{1 + \beta m}$ for $k = 1,2, \ldots$ has been established by Tremolieres [34] under the a unique optimal solution x^* with $g_i(x^*) = 0$ for $i = 1,2, \ldots, n$ and the constraint gradient vectors $\nabla g_i(x)$ are linearly independent for all $x \in S^1$. Under similar assumptions with linear objective and constraint functions this result has been established with equality holding by Faure [5].

To obtain a nonzero lower bound on $\frac{(f^* - f^k)}{(f^* - f^{k-1})}$ for $k = 1,2, \ldots$ requires the assumption of Lemma 3.7 which implies there exists a nonzero $u^* \in U^*$, i.e., $p^* > 0$. The following is an immediate consequence of Corollary 3.6 and Theorem 3.8.

Corollary 3.9:

If $\nabla f(x) \neq 0$ for all $x \in S^1$, then

$$0 < \frac{\beta\tilde{p}}{1 + \beta\tilde{p}} \leqq \frac{(f^* - f^k)}{(f^* - f^{k-1})} \qquad \text{for} \quad k = 1,2, \ldots .$$

By combining the results of Lemma 3.4 and Theorem 3.8 the asymptotic behavior of $\frac{(f^* - f^k)}{(f^* - f^{k-1})}$ can be determined.

Theorem 3.10:

$$\frac{\beta p^*}{1 + \beta p^*} \leqq \liminf_{k\to\infty} \frac{(f^* - f^k)}{(f^* - f^{k-1})} \leqq \limsup_{k\to\infty} \frac{(f^* - f^k)}{(f^* - f^{k-1})} \leqq \frac{\beta(m - q^*)}{1 + \beta(m - q^*)} .$$

Furthermore, if there exists an $x^* \in X^*$ and a $u^* \in U^*$ such that $p(u^*) + q(x^*) = m$ then

$$\lim_{k\to\infty} \frac{(f^* - f^k)}{(f^* - f^{k-1})} = \frac{\beta p^*}{1 + \beta p^*} = \frac{\beta(m - q^*)}{1 + \beta(m - q^*)} .$$

It should be noted that if $q^* = m$ then $\lim_{k\to\infty} \frac{(f^* - f^k)}{(f^* - f^{k-1})} = 0$, i.e., the sequence $\{f^k\}$, $k = 1,2, \ldots$ converges to f^* superlinearly.

For the case when $\varepsilon = 0$ the concluding result of Theorem 3.10 has been stated by Faure and Huard [6]. It has also been proved for this case under assumptions which imply the problem has a unique nondegenerate optimal solution and Kuhn-Tucker multiplier vector pair by Faure [5] for linear objective and constraint functions and by Lootsma [24] for general concave problem functions. Theorem 3.10 shows that the asymptotic rate of convergence of the algorithm is independent of ε and is better for smaller values of β . For example, if $\beta = \frac{1}{m}$ then $\frac{\beta p^*}{1 + \beta p^*} \leq \frac{1}{2}$.

The following corollary is the result of inductively applying Theorem 3.8 and gives upper and lower bounds on $f^* - f^k$ for $k = 1,2, \ldots$ in terms of products of k fractions and gives an upper bounding exponential function of k .

Corollary 3.11:

For $k = 1,2, \ldots$

$$\prod_{j=1}^{k} \left(\frac{\beta p_j}{1 + \beta p_j}\right) \leq \frac{(f^* - f^k)}{(f^* - f^o)} \leq \prod_{j=1}^{k} \left(\frac{\beta s_j}{1 + \beta s_j}\right) \leq \left(\frac{\beta m + \varepsilon\gamma}{1 + \beta m + \varepsilon\gamma}\right)^k .$$

This corollary can be used to obtain a lower bound on the number of iterations k which is sufficient for $f^* - f^k \leq t$ where t is a termination parameter for the algorithm.

Corollary 3.12:

If

$$k \geqq \frac{\ln\left(\frac{f^* - f^o}{t}\right)}{\ln\left(\frac{1 + \beta m + \varepsilon\gamma}{\beta m + \varepsilon\gamma}\right)} \qquad \text{for} \quad t > 0 ,$$

then

$$f^* - f^k \leqq t .$$

<u>Proof</u>:

Suppose

$$k \geqq \frac{\ln\left(\frac{f^* - f^o}{t}\right)}{\ln\left(\frac{1 + \beta m + \varepsilon\gamma}{\beta m + \varepsilon\gamma}\right)} \qquad \text{for} \quad t > 0 .$$

Then since $\frac{1 + \beta m + \varepsilon\gamma}{\beta m + \varepsilon\gamma} > 1$

$$k \ln\left(\frac{1 + \beta m + \varepsilon\gamma}{\beta m + \varepsilon\gamma}\right) \geqq \ln\left(\frac{f^* - f^o}{t}\right)$$

or

$$\left(\frac{1 + \beta m + \varepsilon\gamma}{\beta m + \varepsilon\gamma}\right)^k > \left(\frac{f^* - f^o}{t}\right)$$

which implies

$$t \geqq (f^* - f^o)\left(\frac{\beta m + \varepsilon\gamma}{1 + \beta m + \varepsilon\gamma}\right)^k .$$

Then the conclusion follows from Corollary 3.11.||

It should be noted that an upper bound on $(f^* - f^o)$ is known after one iteration of the algorithm provided an upper bound on γ is known since

$f^* - f^1 \leqq (f^1 - f^o)(\beta m + \varepsilon\gamma)$ by Lemma 3.2 which implies $f^* - f^o \leqq (f^1 - f^o) \cdot$ $\cdot (1 + \beta m + \varepsilon\gamma)$. Thus, a lower bound on the number of iterations k which is sufficient for $f^* - f^k \leqq t$ may be determined from Corollary 3.12 after one iteration of the algorithm.

Another interesting feature of this particular method of centers algorithm is that it is possible to choose values of the algorithm parameters ε and β such that $f^* - f^1 \leqq t$.

Corollary 3.13.

If $\beta > 0$ and $\varepsilon \geqq 0$ are such that

$$\beta m + \varepsilon\gamma \leqq \frac{t}{(f^* - f^o) - t}$$

where

$$0 < t < (f^* - f^o) ,$$

then

$$f^* - f^1 \leqq t .$$

Proof:

$$\beta m + \varepsilon\gamma \leqq \frac{t}{(f^* - f^o) - t}$$

implies

$$\frac{1}{\beta m + \varepsilon\gamma} \geqq \frac{(f^* - f^o)}{t} - 1$$

or

$$\left(\frac{1 + \beta m + \varepsilon\gamma}{\beta m + \varepsilon\gamma}\right) \geqq \frac{(f^* - f^o)}{t} .$$

Then the result follows from Corollary 3.11 with $k = 1$. ||

The above result has been observed by Lootsma [21] and Fiacco and McCormick [11] for the case when $\varepsilon = 0$.

The next corollary which follows from inductive application of Corollary 3.9 provides an exponential function of k which bounds $f^* - f^k$ from below for $k = 1,2, \ldots$.

Corollary 3.14:

If $\nabla f(x) \neq 0$ for all $x \in S^1$, then

$$0 < \left(\frac{\beta\tilde{p}}{1 + \beta\tilde{p}}\right)^k \leqq \frac{f^* - f^k}{f^* - f^o} \qquad \text{for } k = 1,2, \ldots .$$

A lower bound on the number of iterations k which is necessary for $f^* - f^k \leqq t$ can be derived from the previous corollary.

Corollary 3.15:

If $\nabla f(x) \neq 0$ for all $x \in S^1$ and

$$f^* - f^k \leqq t ,$$

then

$$k \geqq \frac{\ln\left(\frac{f^* - f^o}{t}\right)}{\ln\left(\frac{1 + \beta\tilde{p}}{\beta\tilde{p}}\right)} .$$

Proof:

Since the algorithm does not terminate in a finite number of iterations $f^* - f^k > 0$ for all $k \geq 1$. Therefore, if $f^* - f^k \leq t$ then $t > 0$ and by Corollary 3.14

$$0 < \left(\frac{\delta\bar{p}}{1 + \delta\bar{p}}\right)^k \leq \frac{t}{(f^* - f^0)} .$$

Then

$$\left(\frac{1 + \delta\bar{p}}{\delta\bar{p}}\right)^k \geq \frac{(f^* - f^0)}{t}$$

or

$$k \ln\left(\frac{1 + \delta\bar{p}}{\delta\bar{p}}\right) \geq \ln\left(\frac{f^* - f^0}{t}\right)$$

which implies the result since $\left(\frac{1 + \delta\bar{p}}{\delta\bar{p}}\right) > 1$.||

Corollary 3.14 can also be used to obtain an exponential function of k which bounds $||x^* - x^k||$ from below where x^* is any optimal solution to the nonlinear programming problem. Combined with Lemma 3.5 it also yields lower bounding exponential functions of k for all of the constraint function values $g_i(x^k)$ and all of the multiplier values u_i^k . Define

$$(3.24) \qquad \Delta_o = \sup_{x \in S^1} ||\nabla f(x)|| .$$

Theorem 3.16:

If $\nabla f(x) \neq 0$ for all $x \in S^1$, then for $k = 1,2, . .$

$$(3.25) \qquad 0 < \left(\frac{f^* - f^o}{\Delta_o}\right)\left(\frac{\beta\bar{p}}{1 + \beta\bar{p}}\right)^k \leqq \inf_{x \in X^*} ||x - x^k|| \ ,$$

$$(3.26) \qquad 0 < \left[\frac{(f^* - f^o)}{\bar{u}_i\left(m + \left(\frac{\varepsilon}{\beta}\right)\gamma\right)}\right]\left(\frac{\beta\bar{p}}{1 + \beta\bar{p}}\right)^k \leqq g_i(x^k) \qquad \text{for } i = 1,2, \ldots, m$$

and

$$(3.27) \qquad 0 < \left[\frac{(f^* - f^o)}{\bar{g}_i\left(m + \left(\frac{\varepsilon}{\beta}\right)\gamma\right)}\right]\left(\frac{\beta\bar{p}}{1 + \beta\bar{p}}\right)^k \leqq u_i^k \qquad \text{for } i = 1,2, \ldots, m \ .$$

Proof:

Since f is continuously differentiable on S^1 and S^1 is compact, $\Delta_o = \sup_{x \in S^1} ||\nabla f(x)||$ is finite. Then by the mean value theorem

$$f(x^*) - f(x^k) \leqq \Delta_o ||x^* - x^k|| \qquad \text{for any } x^* \in X^* \text{ and for } k = 1,2, \ldots \ .$$

By assumption $\Delta_o > 0$ which implies

$$(3.28) \qquad \frac{(f^* - f^k)}{\Delta_o} \leqq \inf_{x \in X^*} ||x - x^k|| \qquad \text{for } k = 1,2, \ldots \ .$$

From Lemma 3.2

$$\frac{(f^* - f^k)}{(\beta m + \varepsilon\gamma)} \leqq f^k - f^{k-1} \qquad \text{for } k = 1,2, \ldots$$

which combined with Lemma 3.5 yields for $k = 1,2, \ldots$

$$\frac{\beta(f^* - f^k)}{\tilde{u}_i(\beta m + \varepsilon\gamma)} \leqq g_i(x^k) \qquad \text{for } i = 1,2, \ldots, m \tag{3.29}$$

and

$$\frac{\beta(f^* - f^k)}{\tilde{g}_i(\beta m + \varepsilon\gamma)} \leqq u_i^k \qquad \text{for } i = 1,2, \ldots, m \,. \tag{3.30}$$

Then (3.25), (3.26) and (3.27) follow from Corollary 3.14 and (3.28), (3.29) and (3.30) respectively.||

It should be recalled that positive lower bounds on the constraint function value and multiplier value sequences which have positive accumulation points are given in Lemma 3.3.

As demonstrated by the next two theorems, upper bounds which converge to zero are available for constraint function values $g_i(x^k)$ with i such that $\sup_{u \in U^*} u_i > 0$ and multiplier values u_j^k with j such that $\sup_{x \in X^*} g_j(x) > 0$.

Theorem 3.17:

For all $i \in \{1,2, \ldots, m\}$ such that $\sup_{u \in U^*} u_i > 0$

$$g_i(x^k) \leqq \frac{(f^* - f^k)}{\left(\sup_{u \in U^*} u_i\right)} \qquad \text{for } k = 1,2, \ldots \tag{3.31}$$

and

$$g_i(x^k) \leqq \left[\frac{(f^* - f^o)}{\left(\sup_{u \in U^*} u_i\right)}\right]\left(\frac{\beta m + \varepsilon\gamma}{1 + \beta m + \varepsilon\gamma}\right)^k \qquad \text{for } k = 1,2, \ldots \,. \tag{3.32}$$

Proof:

By Relation (3.4) for any $u^* = \left(u_1^*, u_2^*, \ldots, u_m^*\right) \varepsilon\, U^*$

$$f^* - f(x^k) \geqq \sum_{j=1}^{m} u_j^* g_j(x^k) \geqq u_i^* g_i(x^k) \qquad \text{for } i = 1,2, \ldots, m$$

since $x^k \varepsilon\, S^1$ for $k = 1,2, \ldots$. Then (3.31) follows for any i such that $\sup\limits_{u \varepsilon U^*} u_i > 0$, and (3.32) follows from (3.31) by Corollary 3.11.||

Theorem 3.18:

For all $i \varepsilon \{1,2, \ldots, m\}$ such that $\sup\limits_{x \varepsilon X^*} g_i(x) > 0$

$$(3.33) \qquad u_i^k \leqq \frac{(\beta m + \varepsilon\gamma)(f^k - f^{k-1})}{\left(\sup\limits_{x \varepsilon X^*} g_i(x)\right)} \qquad \text{for } k = 1,2, \ldots$$

and

$$(3.34) \qquad u_i^k \leqq \left[\frac{(f^* - f^0)(1 + \beta m + \varepsilon\gamma)}{\left(\sup\limits_{x \varepsilon X^*} g_i(x)\right)}\right]\left(\frac{\beta m + \varepsilon\gamma}{1 + \beta m + \varepsilon\gamma}\right)^k \qquad \text{for } k = 1,2, \ldots .$$

Proof:

By (3.8) for $i = 1,2, \ldots, m$ and $k = 1,2, \ldots$

$$\left(\frac{1}{m + \left(\frac{\varepsilon}{\beta}\right)\gamma}\right) \sup_{x \varepsilon X^*} g_i(x) \leqq g_i(x^k) = \frac{\beta(f^k - f^{k-1})}{u_i^k} .$$

Then (3.33) follows for all i such that $\sup\limits_{x \varepsilon X^*} g_i(x) > 0$, and (3.34) follows from (3.33) by Corollary 3.11 since $f^k - f^{k-1} \leqq f^* - f^{k-1}$ for all $k \geqq 1$.||

Upper bounds on $||x^* - x^k||$, $|g_i(x^k) - g_i(x^*)|$ for indices i such that $g_i(x^*) > 0$ and $|u_j^k - u_j^*|$ for indices j such that $u_j^* > 0$ where x^* is an optimal solution and $u^* = \left(u_1^*, u_2^*, \ldots, u_m^*\right)$ is a Kuhn-Tucker multiplier vector require stronger assumptions on the problem functions. Such assumptions will be considered in the next section in order to obtain further convergence rate results.

4. CONVERGENCE RATE RESULTS REQUIRING STRONG CONCAVITY

In order to obtain further convergence rate results such as upper bounding functions of k for $||x^* - x^k||$, and $|u_i^k - u_i^*|$ and $|g_i(x^k) - g_i(x^*)|$ for all $i \in \{1,2, \ldots, m\}$ where x^* is an optimal solution and $u^* = \left(u_1^*, u_2^*, \ldots, u_m^*\right)$ is a Kuhn-Tucker multiplier vector assumptions stronger than concavity and continuous differentiability will be required. It is for this reason that the following definition is considered.

Definition:

A real-valued function L is *strongly concave* [20] on a convex set $T \subseteq E^n$ if there exists a $\lambda > 0$ such that

$$L\left(\frac{1}{2}(x + y)\right) \geqq \frac{1}{2} L(x) + \frac{1}{2} L(y) + \frac{\lambda}{2}||x - y||^2 \qquad \text{for all } x,y \in T .$$

It can be shown that if T is compact, L has continuous second partial derivatives on T and the matrix of second partial derivatives of L is negative definite on T, then L is strongly concave on T.

In addition to Assumptions (2.1), (2.2), (3.1) and nonfinite termination of the algorithm it will be assumed throughout this section that

(4.1) there exists an $x^* \in X^*$ and a $u^* \in U^*$ such that

(a) $L(x) = f(x) + \sum_{i=1}^{m} u_i^* g_i(x)$ is strongly concave on S^1 with the corresponding constant $\lambda > 0$.†

(b) $\nabla g_i(x^*)$ for $i \in A(x^*) = \left\{i \mid g_i(x^*) = 0 , i \in \{1,2, \ldots, m\}\right\}$ are linearly independent vectors.

---

†Actually this assumption only need hold in the insection of S^1 and a ball about x^*. The stronger condition is assumed for convenience of exposition. It also implies that S^1 is a bounded set which is part of Assumption (2.1).

(c) $p(u^*) + q(x^*) = m$.

It is a well-known saddle point result [19] that any optimal solution to the nonlinear programming problem maximizes $L(x)$ over S^1 . Assumption (4.1.a) implies that x^* is the only point maximizing $L(x)$ on S^1 and therefore x^* is the unique optimal solution to the nonlinear programming problem. It is easy to see from the Kuhn-Tucker conditions (2.33) to (2.36) of Theorem 2.5 that Assumption (4.1.b) implies u^* is the only Kuhn-Tucker multiplier vector. Therefore, under these assumptions Theorems 2.5 and 2.6 imply $\lim_{k\to\infty} x^k = x^*$ and $\lim_{k\to\infty} \left(u_1^k, u_2^k, \ldots, u_m^k\right) = \left(u_1^*, u_2^*, \ldots, u_m^*\right) = u^*$.

Assumption (4.1.c) is a nondegeneracy assumption which implies that $A(x^*)$ has $p^* = p(u^*)$ elements, i.e., $u_i^* > 0$ for all $i \in A(x^*)$. If the index set $Q(x^*)$ is defined by

$$Q(x^*) = \{1,2, \ldots, m\} - A(x^*)$$

then $Q(x^*)$ has $q^* = q(x^*)$ elements. Since it is implicitly assumed that $m \geq 1$, at least one of the index sets $A(x^*)$ or $Q(x^*)$ is nonempty and

$$\delta^* = \min \left[\min_{i\in A(x^*)} u_i^* , \min_{i\in Q(x^*)} g_i(x^*) \right] \tag{4.2}$$

is a finite positive number where the minimum over the empty set is defined to be $+\infty$.

In addition to the above, it will be assumed in this section that the following Lipschitz conditions are satisfied:

(4.3) there exists a positive number μ such that for all $x,y \in S^1$

$$||\nabla f(x) - \nabla f(y)|| \leqq \mu||x - y||$$

and

$$||\nabla g_i(x) - \nabla g_i(y)|| \leqq \mu||x - y|| \qquad \text{for } i = 1,2, \ldots, m .$$

Since S^1 is assumed to be bounded this latter assumption will hold if f and g_i for $i = 1,2, \ldots, m$ have continuous second partial derivatives on S^1 by the generalized mean value theorem [14]. Similar bounds exist for the function values since f and g_i for $i = 1,2, \ldots, m$ are assumed to be continuously differentiable on S^1. That is, for all $x,y \in S^1$

$$|f(x) - f(y)| \leqq \Delta_o||x - y||$$

where by (3.24)

$$\Delta_o = \sup_{x \in S^1} ||\nabla f(x)||$$

and

(4.4) $$|g_i(x) - g_i(y)| \leqq \Delta||x - y|| \qquad \text{for } i = 1,2, \ldots, m$$

where

(4.5) $$\Delta = \max_{1 \leqq i \leqq m} \left[\sup_{x \in S^1} ||\nabla g_i(x)||\right] .$$

The following lemma uses strong concavity to provide a second order extension of Relation (3.4).

Lemma 4.1:

For all $x \in S^1$,

$$||x^* - x||^2 \leqq \left(\frac{1}{\lambda}\right)\left[f^* - f(x) - \sum_{i=1}^{m} u_i^* g_i(x)\right] .$$

Proof:

Since $x^* \varepsilon S^1$ and $L(x)$ is strongly concave on S^1

(4.6) $$L\left(\frac{1}{2}(x^* + x)\right) \geqq \frac{1}{2} L(x^*) + \frac{1}{2} L(x) + \frac{\lambda}{2}||x^* - x||^2 \quad \text{for all } x \varepsilon S^1 .$$

Since S^1 is a convex set, $\frac{1}{2}(x^* + x) \varepsilon S^1$ for all $x \varepsilon S^1$. By the remark following Assumption (4.1) x^* maximizes $L(x)$ on S^1 and, therefore,

(4.7) $$L(x^*) \geqq L\left(\frac{1}{2}(x^* + x)\right) \quad \text{for all } x \varepsilon S^1 .$$

Inequalities (4.6) and (4.7) imply by the definition of $L(x)$ that

$$\frac{1}{2}\left[f(x^*) + \sum_{i=1}^{m} u_i^* g_i(x^*) - f(x) - \sum_{i=1}^{m} u_i^* g_i(x)\right] \geqq \frac{\lambda}{2}||x^* - x||^2 \quad \text{for all } x \varepsilon S^1 .$$

Then the desired result follows since $\sum_{i=1}^{m} u_i^* g_i(x^*) = 0$.||

It should be noted that the uniqueness of x^* follows immediately from this lemma. Combining the result of this lemma with the sequences $\{x^k\}$ and $\left\{\left(u_1^k, u_2^k, \ldots, u_m^k\right)\right\}$, $k = 1,2, \ldots$ generated by the algorithm yields the following lemma. By (3.12) and the uniqueness of u^*

(4.8) $$p_k = \sum_{i=1}^{m} \frac{u_i^*}{u_i^k} \quad \text{for } k = 1,2, \ldots .$$

Define

$$q_k = \sum_{i=1}^{m} \frac{g_i(x^*)}{g_i(x^k)} \qquad \text{for } k = 1,2, \dots . \tag{4.9}$$

Lemma 4.2:

For $k = 1,2, \dots$

$$||x^* - x^k||^2 \leq (f^k - f^{k-1})\left(\frac{1}{\lambda}\right)\left[\beta(m - p_k - q_k) + \varepsilon||x^* - x^k||\right] .$$

Proof:

Since $x^k \in S^1$ for $k = 1,2, \dots$, Lemma 4.1 implies

$$||x^* - x^k||^2 \leq \left(\frac{1}{\lambda}\right)\left[f^* - f^k - \sum_{i=1}^{m} u_i^* g_i(x^k)\right] \qquad \text{for } k = 1,2, \dots . \tag{4.10}$$

By Lemma 3.2 and (4.9)

$$f^* - f^k \leq (f^k - f^{k-1})\left[\beta(m - q_k) + \varepsilon||x^* - x^k||\right] \qquad \text{for } k = 1,2, \dots . \tag{4.11}$$

Combining (4.10) and (4.11) with (4.8) yields the desired result since

$$g_i(x^k) = \frac{\beta(f^k - f^{k-1})}{u_i^k} \qquad \text{for } i = 1,2, \dots, m \text{ and } k = 1,2, \dots . ||$$

From this lemma it is easy to see that the convergence of $||x^* - x^k||^2$ to zero is at least as fast as $(f^k - f^{k-1})$ since $m - p_k - q_k \leq m$ and $||x^* - x^k|| \leq \gamma = \sup_{x,y \in S^1} ||x - y||$ for all $k \geq 1$. The next corollary shows that it is even faster due to the nondegeneracy assumption (4.1.c).

Corollary 4.3:

$$\lim_{k\to\infty} \frac{||x^* - x^k||^2}{(f^k - f^{k-1})} = 0 .$$

Proof:

By (4.8) and (4.9)

$$\lim_{k\to\infty} (m - p_k - q_k) = m - p(u^*) - q(x^*) . \tag{4.12}$$

Then by (4.12) and Assumption (4.1.c)

$$\lim_{k\to\infty} \left[\beta(m - p_k - q_k) + \varepsilon||x^* - x^k||\right] = 0$$

and the desired result follows from Lemma 4.2.||

In fact as the remainder of this section will show, a result stronger than Corollary 4.3 is true. The next lemma begins this development by providing an algebraic equivalent for the expression $(m - p_k - q_k)$ which appears in Lemma 4.2.

Lemma 4.4:

For $k = 1,2, \ldots$

$$m - \sum_{i=1}^{m} \frac{u_i^*}{u_i^k} - \sum_{i=1}^{m} \frac{g_i(x^*)}{g_i(x^k)} = \sum_{i=1}^{m} \left(\frac{u_i^k - u_i^*}{u_i^k}\right)\left(\frac{g_i(x^k) - g_i(x^*)}{g_i(x^k)}\right) .$$

Proof:

By the assumptions on x^* and $u^* = \left(u_1^*, u_2^*, \ldots, u_m^*\right)$

$$u_i^* g_i(x^*) = 0 \qquad \text{for } i = 1,2, \ldots, m .$$

Thus,

$$m - \sum_{i=1}^{m} \frac{u_i^*}{u_i^k} - \sum_{i=1}^{m} \frac{g_i(x^*)}{g_i(x^k)} = \sum_{i=1}^{m} \left[\frac{u_i^k g_i(x^k)}{u_i^k g_i(x^k)} - \frac{u_i^* g_i(x^k)}{u_i^k g_i(x^k)} - \frac{u_i^k g_i(x^*)}{u_i^k g_i(x^k)} + \frac{u_i^* g_i(x^*)}{u_i^k g_i(x^k)}\right]$$

$$= \sum_{i=1}^{m} \left[\frac{\left(u_i^k - u_i^*\right) g_i(x^k) - \left(u_i^k - u_i^*\right) g_i(x^*)}{u_i^k g_i(x^k)}\right]$$

which is equivalent to the desired result.||

An upper bound for $(m - p_k - q_k)$ can be found by combining this lemma with the nondegeneracy assumption (4.1.c), the positive lower bounds on u_i^k for $i \in A(x^*)$ and $g_i(x^k)$ for $i \in Q(x^*)$ provided by Lemma 3.3 and the definition of δ^*.

Lemma 4.5:

For $k = 1,2, \ldots$

$$m - p_k - q_k \leq \left(\frac{1}{\delta^*}\right)\left(m + \left(\frac{\varepsilon}{\delta}\right)\right)\left[\sum_{i \in A(x^*)} |u_i^k - u_i^*| + \sum_{i \in Q(x^*)} |g_i(x^k) - g_i(x^*)|\right]$$

where summation over an empty index set is assumed to be zero.

Proof:

Since $g_i(x^*) = 0$ for all $i \in A(x^*)$, $u_i^* = 0$ for all $i \in Q(x^*)$ and $A(x^*) \cup Q(x^*) = \{1,2, \ldots, m\}$ the result of Lemma 4.4 implies

$$m - p_k - q_k = \sum_{i \in A(x^*)} \frac{\left(u_i^k - u_i^*\right)}{u_i^k} + \sum_{i \in Q(x^*)} \frac{\left(g_i(x^k) - g_i(x^*)\right)}{g_i(x^k)} \quad \text{for } k = 1,2, \ldots$$

where summation over an empty index set is assumed to be zero. Since $u_i^* > 0$ for all $i \in A(x^*)$ by Assumption (4.1.c) and $g_i(x^*) > 0$ for all

$i \in Q(x^*)$, the lower bound results of Lemma 3.3 imply that

$$m - p_k - q_k \leq \left(m + \binom{\cdot}{\cdot}\right)\left[\sum_{i\in A(x^*)} \frac{|u_i^k - u_i^*|}{u_i^*} + \sum_{i\in Q(x^*)} \frac{|g_i(x^k) - g_i(x^*)|}{g_i(x^*)}\right]$$

for $k = 1,2, \ldots$.

Then the desired result follows from (4.2).||

In order to proceed further it is necessary to bound the expressions $\sum_{i\in A(x^*)} |u_i^k - u_i^*|$ and $\sum_{i\in Q(x^*)} |g_i(x^k) - g_i(x^*)|$ from above by functions of $(f^k - f^{k-1})$ and $||x^* - x^k||$. The latter can be accomplished by using (4.4) and the former will be considered after a preliminary result depending on Assumption (4.1.b) is established.

For a $p \times q$ matrix H denote the transpose of H by H^T and define the norm of H using the Euclidean norm for the vectors $y \in L^q$ and $Hy \in L^p$ by

(4.13) $$||H|| = \sup_{||y||=1} ||Hy|| .$$

If $p^* > 0$ renumber the constraint functions, if necessary, so that $A(x^*) = \{1,2, \ldots, p^*\}$ and for $x \in S^1$ let $H^*(x)$ be the $p^* \times n$ matrix whose i^{th} row is $\nabla g_i(x)$ for each $i \in A(x^*)$.

Lemma 4.6:

If $p^* > 0$, then there exist positive numbers $\bar{\rho}$ and η such that $[H^*(x)H^*(x)^T]^{-1}$ exists and

$$||[H^*(x)H^*(x)^T]^{-1}|| \leq \frac{1}{\bar{\rho}} \qquad \text{for all } x \in B_\eta(x^*) \cap S^1$$

$$B_\eta(x^*) = \{x \mid ||x - x^*|| \leqq \eta\} .$$

Proof:

Since $\nabla g_i(x)$ for $i = 1,2, \ldots, m$ is continuous on S^1, $\rho(x) = \min\limits_{||y||=1} y[H^*(x)H^*(x)^T]y$ is continuous on S^1. By Assumption (4.1.b) $H^*(x)$ has full row rank p^* and, therefore, $\rho(x^*)$ is positive. Thus, there exist positive numbers $\bar{\rho}$ and η such that $[H^*(x)H^*(x)^T]^{-1}$ exists and $\rho(x) \geqq \bar{\rho} > 0$ for all $x \in B_\eta(x^*) \cap S^1$. It can be shown that $\rho(x)$ is the minimum eigenvalue of $[H^*(x)H^*(x)^T]$ and, therefore, $\frac{1}{\rho(x)}$ is the maximum eigenvalue of $[H^*(x)H^*(x)^T]^{-1}$. Then

$$\max_{||y||=1} y[H^*(x)H^*(x)^T]^{-1}y = \frac{1}{\rho(x)} \leqq \frac{1}{\bar{\rho}}$$

for all $x \in B_\eta(x^*) \cap S^1$ and the desired result follows since as in Goldstein [14; p. 22]

$$||[H^*(x)H^*(x)^T]^{-1}|| = \max_{||y||=1} ||[H^*(x)H^*(x)^T]^{-1}y|| =$$

$$= \max_{||y||=1} y[H^*(x)H^*(x)^T]^{-1}y \; .||$$

By combining the result of this lemma with bounds provided by Assumptions (2.2) and (4.3) an upper bound on $\sum\limits_{i\in A(x^*)} |u_i^k - u_i^*|$ for $k = 1,2, \ldots$ can be found in terms of $(f^k - f^{k-1})$ and $||x^* - x^k||$.

Lemma 4 7:

If $p^* > 0$, then there exists a positive number ρ such that for $k = 1,2, \ldots$

$$\sum_{i\varepsilon A(x^*)} |u_i^k - u_i^*| \leq p^* \Lambda\left(\frac{1}{\rho}\right)\left\{\left[\varepsilon + \Lambda\left(\frac{q^*}{\delta^*}\right)(2m + \varepsilon\gamma)\right](f^k - f^{k-1}) + \right.$$

$$\left. + \mu\left(1 + \sum_{i=1}^{m} u_i^*\right)||x^* - x^k||\right\} .$$

<u>Proof:</u>

By the definitions of x^* and u^*

$$\nabla f(x^*) + \sum_{i=1}^{m} u_i^* \nabla g_i(x^*) = 0 \tag{4.14}$$

and by the definitions of $\nabla d^k(x^k)$ and u_i^k for $i = 1,2, \ldots, m$

$$\nabla f(x^k) + \sum_{i=1}^{m} u_i^k \nabla g_i(x^k) = (f^k - f^{k-1})\nabla d^k(x^k) \qquad \text{for } k = 1,2, \ldots . \tag{4.15}$$

Subtracting (4.14) from (4.15) yields

$$\nabla f(x^k) - \nabla f(x^*) + \sum_{i=1}^{m} u_i^*\left(\nabla g_i(x^k) - \nabla g_i(x^*)\right) + \sum_{i=1}^{m} \left(u_i^k - u_i^*\right)\nabla g_i(x^k) =$$

$$= (f^k - f^{k-1})\nabla d^k(x^k) \qquad \text{for } k = 1,2, \ldots$$

and by rearranging terms

$$\sum_{i\varepsilon A(x^*)} \left(u_i^k - u_i^*\right)\nabla g_i(x^k) = - \sum_{i\varepsilon Q(x^*)} \left(u_i^k - u_i^*\right)\nabla g_i(x^k) + (f^k - f^{k-1})\nabla d^k(x^k) +$$

$$+ \nabla f(x^*) - \nabla f(x^k) + \sum_{i=1}^{m} u_i^*\left(\nabla g_i(x^*) - \nabla g_i(x^k)\right) \qquad \text{for } k = 1,2, \ldots$$

which implies by the triangle inequality

$$(4.16)\qquad \Big\| \sum_{i \in A(x^*)} \left(u_i^k - u_i^*\right) \nabla g_i(x^k) \Big\| \leq \sum_{i \in Q(x^*)} u_i^k \|\nabla g_i(x^k)\| + (f^k - f^{k-1}) \|\nabla d^k(x^k)\| + \|\nabla f(x^*) - \nabla f(x^k)\| + \sum_{i=1}^{m} u_i^* \|\nabla g_i(x^*) - \nabla g_i(x^k)\| \qquad \text{for } k = 1,2, \ldots$$

since $u_i^* = 0$ for all $i \in Q(x^*)$. Since $Q(x^*)$ has q^* elements, $\|\nabla g_i(x^k)\| \leq \ell$ for $i = 1,2, \ldots, m$ by (4.5), $g_i(x^k) \geq \left(\frac{1}{m + \left(\frac{\ell}{\varepsilon}\right)\gamma}\right) g_i(x^*)$ for $i = 1,2, \ldots, m$ by lemma 3.3 and $\delta^* \leq \min_{i \in Q(x^*)} g_i(x^*)$

$$(4.17)\qquad \sum_{i \in Q(x^*)} u_i^k \|\nabla g_i(x^k)\| = \sum_{i \in Q(x^*)} \left(\frac{\beta(f^k - f^{k-1})}{g_i(x^k)}\right) \|\nabla g_i(x^k)\| \leq \ell\left(\frac{q^*}{\delta^*}\right)(\beta m + \varepsilon\gamma)(f^k - f^{k-1}) \qquad \text{for } k = 1,2, \ldots .$$

By Assumption (4.3)

$$(4.18)\qquad \|\nabla f(x^*) - \nabla f(x^k)\| + \sum_{i=1}^{m} u_i^* \|\nabla g_i(x^*) - \nabla g_i(x^k)\| \leq \mu\left(1 + \sum_{i=1}^{m} u_i^*\right) \|x^* - x^k\| \qquad \text{for } k = 1,2, \ldots .$$

Combining (4.16), (4.17) and (4.18) with $\|\nabla d^k(x^k)\| \leq \varepsilon$ yields

$$(4.19)\qquad \Big\| \sum_{i \in A(x^*)} \left(u_i^k - u_i^*\right) \nabla g_i(x^k) \Big\| \leq \left[\varepsilon + \ell\left(\frac{q^*}{\delta^*}\right)(\beta m + \varepsilon\gamma)\right](f^k - f^{k-1}) + \mu\left(1 + \sum_{i=1}^{m} u_i^*\right) \|x^* - x^k\| \qquad \text{for } k = 1,2, \ldots .$$

Now let w^k be a p^* vector for $k = 1,2, \ldots$ with

(4.20) $$w_i^k = u_i^k - u_i^* \quad \text{for} \quad i \in A(x^*) = \{1,2, \ldots, p^*\} .$$

Then for $k = 1,2, \ldots$

$$w^k H^*(x^k)H^*(x^k)^T = \left[\sum_{i\in A(x^*)} \left(u_i^k - u_i^*\right)\nabla g_i(x^k)\right] H^*(x^k)^T .$$

Since $\lim_{k\to\infty} x^k = x^*$, Lemma 4.6 implies there exists an integer $\bar{k}$ and a positive number $\bar{\rho}$ such that $[H^*(x^k)H^*(x^k)^T]^{-1}$ exists and

(4.21) $$\|[H^*(x^k)H^*(x^k)^T]^{-1}\| \leq \frac{1}{\bar{\rho}} \qquad \text{for all} \quad k \geq \bar{k} .$$

Then

$$w^k = \left[\sum_{i\in A(x^*)} \left(u_i^k - u_i^*\right)\nabla g_i(x^k)\right] H^*(x^k)^T [H^*(x^k)H^*(x^k)^T]^{-1} \qquad \text{for all} \quad k \geq \bar{k}$$

and by the generalized Cauchy-Schwarz inequality [31; p. 185]

(4.22) $$\|w^k\| \leq \left\|\sum_{i\in A(x^*)} \left(u_i^k - u_i^*\right)\nabla g_i(x^k)\right\| \, \|H^*(x^k)^T\| \, \|[H^*(x^k)H^*(x^k)^T]^{-1}\| \quad \text{for all} \quad k \geq \bar{k} .$$

By a matrix norm property [31; p. 188] and the definitions of Δ and p^*

(4.23) $$\|H^*(x^k)^T\| \leq \left[\sum_{i\in A(x^*)} \sum_{j=1}^{n} \left(\frac{\partial g_i(x^k)}{\partial x_j}\right)^2\right]^{1/2} \leq (p^*)^{1/2}\Delta \qquad \text{for} \quad k = 1,2, \ldots .$$

Then combining (4.19), (4.21), (4.23) and (4.22) yields

$$||w^k|| \leq (p^*)^{\frac{1}{2}} \iota\left(\frac{1}{\rho}\right)\left\{\left[\varepsilon + \iota\left(\frac{q^*}{\delta^*}\right)(\bar{\varepsilon} m + \varepsilon\gamma)\right](f^k - f^{k-1}) + \right. \tag{4.24}$$

$$\left. + \mu\left(1 + \sum_{i=1}^{m} u_i^*\right)||x^* - x^k||\right\} \qquad \text{for all} \quad k \geq \bar{k} \ .$$

By (4.20)

$$\sum_{i \in A(x^*)} |u_i^k - u_i^*| \leq (p^*)^{\frac{1}{2}}||w^k|| \qquad \text{for} \quad k = 1,2, \ldots \ . \tag{4.25}$$

Then from (4.24) and (4.25) there exists a positive number $\rho \leq \bar{\rho}$ such that the desired result holds.||

In order to combine Lemmas 4.2, 4.5 and 4.7 to obtain an upper bound on $||x^* - x^k||$ in terms of $(f^k - f^{k-1})$ a lemma not depending on problem assumptions will be required.

Lemma 4.8:

Let a , b , c and d be nonnegative numbers such that

$$a^2 - bad - cd^2 \leq 0 \ . \tag{4.26}$$

Then

$$a \leq \frac{1}{2}[b + (b^2 + 4c)^{\frac{1}{2}}]d \ .$$

Proof:

The result is trivial if $a = 0$ so suppose $a > 0$. Then clearly

$$a - \frac{1}{2}[b - (b^2 + 4c)^{\frac{1}{2}}]d > 0 \ . \tag{4.27}$$

Relation (4.26) is equivalent to

$$(4.28)\qquad \left[a - \tfrac{1}{2}[b + (b^2 + 4c)^{\frac{1}{2}}]d\right]\left[a - \tfrac{1}{2}[b - (b^2 + 4c)^{\frac{1}{2}}]d\right] \leqq 0 .$$

Then the desired result follows from (4.27) and (4.28).||

Now all of the previous results may be combined to show that $||x^* - x^k||$ and $\sum_{i\epsilon A(x^*)} |u_i^k - u_i^*|$ are bounded above by linear functions of $(f^k - f^{k-1})$

<u>Lemma 4.9</u>:

Suppose $p^* > 0$ and let ρ be as in Lemma 4.7. Then for $k = 1,2, \ldots$

$$(4.29)\qquad ||x^* - x^k|| \leqq \left(\tfrac{1}{2}\right)\left[b_1 + \left(b_1^2 + 4b_2\right)^{\frac{1}{2}}\right](f^k - f^{k-1})$$

and

$$(4.30)\qquad \sum_{i\epsilon A(x^*)} |u_i^k - u_i^*| \leqq b_3(f^k - f^{k-1})$$

where

$$(4.31)\qquad b_1 = \left(\frac{\varepsilon}{\lambda}\right) + \left(\frac{\Lambda}{\lambda}\right)\left[\left(\frac{q^*}{\delta^*}\right) + \left(\frac{\mu}{\rho}\right)\left(1 + \sum_{i=1}^{m} u_i^*\right)\left(\frac{p^*}{\delta^*}\right)\right](\beta m + \varepsilon\gamma) ,$$

$$(4.32)\qquad b_2 = \left(\frac{1}{\rho}\right)\left(\frac{\Lambda}{\lambda}\right)\left(\frac{p^*}{\delta^*}\right)\left[\varepsilon + \Delta\left(\frac{q^*}{\delta^*}\right)(\beta m + \varepsilon\gamma)\right](\beta m + \varepsilon\gamma)$$

and

$$(4.33)\qquad b_3 = p^*\left(\frac{\Lambda}{\rho}\right)\left[\varepsilon + \Delta\left(\frac{q^*}{\delta^*}\right)(\beta m + \varepsilon\gamma) + \left(1 + \sum_{i=1}^{m} u_i^*\right)\left(\frac{\mu}{2}\right)\left[b_1 + \left(b_1^2 + 4b_2\right)^{\frac{1}{2}}\right]\right] .$$

Proof.

By combining the results of Lemma 4.2 and Lemma 4.5

$$(4.34)\quad \|x^* - x^k\|^2 \leqq (f^k - f^{k-1})\left(\frac{1}{\lambda}\right)\left\{\left(\frac{1}{\delta^*}\right)(\beta m + \varepsilon\gamma)\left[\sum_{i\in A(x^*)} |u_i^k - u_i^*| + \sum_{i\in Q(x^*)} |g_i(x^k) - g_i(x^*)|\right] + \varepsilon\|x^* - x^k\|\right\} \quad \text{for } k = 1,2,\ldots.$$

by (4.4) and the definition of q^*

$$(4.35)\quad \sum_{i\in Q(x^*)} |g_i(x^k) - g_i(x^*)| \leqq q^*\Delta\|x^k - x^*\| \quad \text{for } k = 1,2,\ldots.$$

By combining (4.34) and (4.35) and the result of Lemma 4.7

$$\|x^* - x^k\|^2 \leqq (f^k - f^{k-1})\left(\frac{1}{\lambda}\right)\left\{(\beta m + \varepsilon\gamma)\left(\frac{p^*}{\delta^*}\right)\left(\frac{1}{\mu}\right)\left[\varepsilon + \mu\left(\frac{q^*}{\delta^*}\right)(\beta m + \varepsilon\gamma)\right](f^k - f^{k-1}) + \left[(\beta m + \varepsilon\gamma)\left(\frac{p^*}{\delta^*}\right)\left(\frac{1}{\mu}\right)\cdot\left(1 + \sum_{i=1}^{m} u_i^*\right) + (\beta m + \varepsilon\gamma)\left(\frac{q^*}{\delta^*}\right)\cdot + \varepsilon\right]\|x^* - x^k\|\right\}$$

$$\text{for } k = 1,2,\ldots.$$

Defining b_1 and b_2 by (4.31) and (4.32) yields

$$(4.36)\quad \|x^* - x^k\| \leqq b_2(f^k - f^{k-1})^2 + b_1(f^k - f^{k-1})\|x^* - x^k\|$$

$$\text{for } k = 1,2,\ldots.$$

Then (4.29) follows immediately from (4.36) and Lemma 4.8 with $a = \|x^* - x^k\|$, $b = b_1$, $c = b_2$ and $d = (f^k - f^{k-1})$. Then (4.30) follows directly from Lemma 4.7 and (4.29) when b_3 is defined by (4.33).||

Combining the above result with Corollary 3.11 yields the following upper bounding decreasing exponential functions of k for $||x^* - x^k||$, $\sum_{i \in A(x^*)} |u_i^k - u_i^*|$ and $|g_i(x^k) - g_i(x^*)|$ for $i = 1,2, \ldots, m$.

Theorem 4.10

Suppose $p^* > 0$ and let $a_1 = \left(\frac{1}{2}\right)\left[b_1 + \left(b_1^2 + 4b_2\right)^{1/2}\right]$. Then for $k = 1,2, \ldots$

$$(4.37) \qquad ||x^* - x^k|| \leqq a_1(f^* - f^o)\left(\frac{\beta m + \varepsilon\gamma}{1 + \beta m + \varepsilon\gamma}\right)^{k-1},$$

$$(4.38) \qquad \sum_{i \in A(x^*)} |u_i^k - u_i^*| \leqq b_3(f^* - f^o)\left(\frac{\beta m + \varepsilon\gamma}{1 + \beta m + \varepsilon\gamma}\right)^{k-1}$$

and

$$(4.39) \qquad |g_i(x^k) - g_i(x^*)| \leqq \Delta a_1(f^* - f^o)\left(\frac{\beta m + \varepsilon\gamma}{1 + \beta m + \varepsilon\gamma}\right)^{k-1} \quad \text{for } i = 1,2, \ldots, m.$$

Proof:

By Corollary 3.11

$$(4.40) \qquad f^k - f^{k-1} \leqq f^* - f^{k-1} \leqq (f^* - f^o)\left(\frac{\beta m + \varepsilon\gamma}{1 + \beta m + \varepsilon\gamma}\right)^{k-1} \quad \text{for } k = 1,2, \ldots .$$

Then (4.37) and (4.38) follow from (4.40) and (4.29) and (4.30) of Lemma 4.9, respectively. The final result (4.39) follows from (4.4) and (4.37).||

For the case when $p^* = 0$ corresponding upper bounds can be given in terms of products of $k - 1$ fractions where the fractions converge to zero.

Theorem 4.11:

Suppose $p^* = 0$ and let $a_2 = \left(\frac{\varepsilon}{\lambda}\right) + \left(\frac{\Delta}{\lambda}\right)\left(\frac{m}{\delta^*}\right)(\beta m + \varepsilon\gamma)$. Then for $k = 1,2, \ldots$

$$(4.41) \qquad ||x^* - x^k|| \leqq a_2(f^* - f^o) \prod_{j=1}^{k-1} \left(\frac{\beta s_j}{1 + \beta s_j}\right)$$

and

$$(4.42) \qquad |g_i(x^k) - g_i(x^*)| \leqq \Delta a_2(f^* - f^o) \prod_{j=1}^{k-1} \left(\frac{\beta s_j}{1 + \beta s_j}\right) \quad \text{for } i = 1,2, \ldots, m$$

where

$$\lim_{j\to\infty} s_j = 0 .$$

Proof.

From (4.34) and (4.35) with $A(x^*)$ empty

$$||x^* - x^k||^2 \leqq (f^k - f^{k-1})\left(\frac{1}{\lambda}\right)\left[\left(\frac{1}{\delta^*}\right)(\beta m + \varepsilon\gamma)q^*\Delta + \varepsilon\right]||x^* - x^k|| \quad \text{for } k = 1,2, \ldots .$$

Then since $q^* = m$

$$(4.43) \qquad ||x^* - x^k|| \leqq a_2(f^k - f^{k-1}) \qquad \text{for } k = 1,2, \ldots .$$

By Corollary 3.11

$$(4.44) \quad f^k - f^{k-1} \leqq f^* - f^{k-1} \leqq (f^* - f^o) \prod_{j=1}^{k-1} \left(\frac{\beta s_j}{1 + \beta s_j}\right) \quad \text{for } k = 1,2, \ldots .$$

Then (4.41) follows from (4.43) and (4.44) and $\lim_{j\to\infty} s_j = 0$ by Lemma 3.4 since $m - q^* = p^* = 0$. Then (4.42) follows from (4.4) and (4.41).||

The convergence rate given by (4.37) of Theorem 4.10 is an improvement by a factor of 2 over the following convergence rate result which represents the usual way of getting a rate for $x^k \to x^*$ given a rate for $f^k \to f^*$.[†] This result follows directly from Lemma 4.1 and Corollary 3.11 and does not require Assumptions (4.1.b), (4.1.c) or (4.3).

Theorem 4 12:

For $k = 1,2, \ldots$

$$||x^* - x^k|| \leqq \left(\frac{f^* - f^o}{\lambda}\right)^{\frac{1}{2}} \left(\frac{\beta m + \varepsilon\gamma}{1 + \beta m + \varepsilon\gamma}\right)^{k/2} .$$

Proof:

From Lemma 4.1 with $x = x^k \in S^1$ for $k = 1,2, \ldots$

$$||x^* - x^k||^2 \leqq \left(\frac{1}{\lambda}\right)\left[f^* - f(x^k) - \sum_{i=1}^{m} u_i^* g_i(x^k)\right] \leqq \left(\frac{1}{\lambda}\right)(f^* - f^k) .$$

Then the desired result follows from Corollary 3.11.||

[†] For example see [20], [28], [29] and [32].

5. SUBPROBLEM CONVERGENCE

In this section the convergence of Cauchy's [3] method of steepest ascent for each subproblem k will be studied and an upper bound on the number of steepest ascent steps required to find x^k from x^{k-1} will be derived. Combined with the result of Corollary 3.12, this will lead to an upper bound on the total number of steepest ascent steps required to find a point x^k starting from x^o such that $f^* - f(x^k) \leqq t$ where t is a termination parameter for the algorithm.

In addition to Assumptions (2.1) and (3.1) which imply S^1 is bounded and convex, it will be assumed throughout this section that

(5.1) f and g_i for $i = 1,2, \ldots, m$ are twice continuously differentiable on S^1 ,

$$\nabla f(x) \neq 0 \qquad \text{for all } x \in S^1 \tag{5.2}$$

and

$$\varepsilon > 0 . \tag{5.3}$$

The $n \times n$ symmetric matrices of second partial derivatives of the respective problem functions which exist and are continuous by Assumption (5.1) will be denoted by

$$H_o(x) = \left[\frac{\partial^2 f(x)}{\partial x_j \partial x_\ell}\right] \tag{5.4}$$

and

$$H_i(x) = \left[\frac{\partial^2 g_i(x)}{\partial x_j \partial x_\ell}\right] \qquad \text{for } i = 1,2, \ldots, m . \tag{5.5}$$

Assumption (5.1) implies $g_1, \ldots, g_m$ and f are continuously differentiable on S^1 which is assumption (2.2) and together with the assumption that S^1 is bounded and the generalized mean value theorem [14, p. 20] implies that the Lipschitz condition of Assumption (4.3) holds when μ is defined by

$$(5.6) \qquad \mu = \max_{0 \leq i \leq m} \left[\sup_{x \in S^1} \|H_i(x)\| \right]$$

where the matrix norm is as defined in (4.13). The concavity of f and g_i for $i = 1, 2, \ldots, m$ implies that the matrices $H_i(x)$ for $i = 0, 1, \ldots, m$ are negative semidefinite for all $x \in S^1$. Then as in [14, p. 22]

$$(5.7) \qquad \|H_i(x)\| = \sup_{\|y\|=1} y[-H(x)]y \quad \text{for } i = 0, 1, \ldots, m \text{ and all } x \in S^1 .$$

Combining (5.6) and (5.7) gives the useful result, that for all $x \in S^1$ and all $y \in E^n$

$$(5.8) \qquad y[-H_i(x)]y \leq \mu \|y\|^2 \qquad \text{for } i = 0, 1, \ldots, m .$$

Assumption (5.2) implies the algorithm does not terminate in a finite number of iterations and together with (5.1) implies that $\sigma > 0$ where

$$(5.9) \qquad \sigma = \inf_{x \in S^1} \|\nabla f(x)\| .$$

It will be convenient to define a function $G(x)$ which gives the smallest constraint value for feasible points x by

$$(5.10) \qquad G(x) = \min_{1 \leq i \leq m} g_i(x) \qquad \text{for } x \in S .$$

In addition to the parameters defined by (3.5), (3.24), (4.5), (5.6) and (5.9) the following will be used in this section:

(5.11) $$\tilde{g} = \max_{1 \leq i \leq m} \left[\sup_{x \in S^1} g_i(x)\right]$$

(5.12) $$\eta = 2 \max \left[\mu, 2\Delta_o^2, 2u^2\right]$$

(5.13) $$\theta = \left(\frac{\eta}{\sigma^2}\right) \max \left[\left(\frac{f - f^o}{G(x^o)}\right), \left(\frac{\tilde{g}}{G(x^o)}\right)\right]$$

(5.14) $$e(\beta,\varepsilon) = \max\left[\tilde{g}, (\eta + \varepsilon\gamma)(1 + \beta m + \varepsilon\gamma)\right]$$

(5.15) $$b_1(\beta) = \left[\theta \max\left[G(x^o), 1\right]\right]^{\beta m+1}$$

(5.16) $$b(\beta,\varepsilon) = \left[\theta\left(\frac{e(\beta,\varepsilon)}{\mu}\right)\right]^{\beta m+1}$$

and

(5.17) $$\alpha(\beta,\varepsilon) = \left(\frac{1 + \beta\tilde{p}}{\beta\tilde{p}}\right)$$

where $\tilde{p}$ is defined by (3.23) and depends on $\tilde{u}_i$ for $i = 1,2, \ldots, m$ defined by (3.19) which is a function of the algorithm parameters β and ε.

As will be shown in the sequel, Assumption (5.3) guarantees that only a finite number of subproblem steps will be required to find each x^k when a slight modification of the following algorithm is used to solve each subproblem.

Method of Steepest Ascent with Optimal Step Size:[†]

Let d be a real-valued function defined on E^n and $z_o \in E^n$ be a starting point. Assume that $T = \{z \mid d(z) \geq d(z_o)\}$ is bounded and that d is continuously differentiable on $T^\bullet$.

[†]For general algorithms of this type see Topkis and Veinott [33].

For $j = 1,2, \ldots$ let λ_{j-1} be a positive number satisfying

$$d(z_{j-1} + \lambda_{j-1}\nabla d(z_{j-1})) = \max_{\lambda \geqq 0} d(z_{j-1} + \lambda \nabla d(z_{j-1}))$$

and let

$$z_j = z_{j-1} + \lambda_{j-1}\nabla d(z_{j-1})$$

starting from z_o and stopping if $\nabla d(z_{j-1}) = 0$ for some $j \geqq 1$. Curry [4] has shown that if $\bar{z}$ is an accumulation point of the sequence $\{z_j\}$, $j = 1,2, \ldots$ then $\nabla d(\bar{z}) = 0$.

For solving subproblem k the first step of this algorithm will have to be modified in order to take into account that d^k is to be maximized over an open set $\hat{S}^k$ starting from a point x^{k-1} on the boundary of $\hat{S}^k$ where d^k is not defined. By employing the result of Lemma 2.1 a step of optimal size may be made from $z_o^k = x^{k-1}$ in the direction $\nabla f(x^{k-1})$ to find a point $z_1^k \in \hat{S}^k$ and a set

$$T^k = \left\{x \mid x \in \hat{S}^k \text{ , } d^k(x) \geqq d^k\left(z_1^k\right)\right\} \tag{5.18}$$

on which to carry out the remainder of the steepest ascent steps. The modified algorithm essentially defines $\nabla d^k\left(z_o^k\right)$ to be $\nabla f(x^{k-1})$.

For each integer $k \geqq 1$ let $\left\{z_j^k\right\}$, $j = 1,2, \ldots$ be the sequence of points generated by the modified steepest ascent algorithm starting from $z_o^k = x^{k-1}$. Since ∇d^k is continuous on T^k and T^k is compact by the continuity of d^k on the bounded set $\hat{S}^k \supset T^k$, each accumulation point $\bar{z}^k$ of $\left\{z_j^k\right\}$, $j = 1,2, \ldots$ satisfies $\nabla d^k(\bar{z}^k) = 0$ and therefore since $\varepsilon > 0$ there exists an integer j such that $\left|\left|\nabla d^k\left(z_j^k\right)\right|\right| \leqq \varepsilon$. Let $\ell(k)$ be the smallest integer j such that $\left|\left|\nabla d^k\left(z_j^k\right)\right|\right| \leqq \varepsilon$ and set $x^k = z_{\ell(k)}^k$.

Then $\ell(k)$ is the number of steps required to solve subproblem k and, thus, find a starting point for subproblem $k+1$.

The development to bound $\ell(k)$ begins with the following lemma which is an extension of Lemma 2.1 dealing with a step from x^{k-1} in the direction $\nabla f(x^{k-1})$ to a point $x(\tilde{\lambda})$ in S^k. It not only shows the existence of $x(\cdot)$ but uses second order information to provide positive lower bounds for $f(x(\tilde{\lambda})) - f^{k-1}$ and $g_i(x(\tilde{\lambda}))$ for $i = 1,2, \ldots, m$.

Lemma 5.1

For each integer $k \geqq 1$ there exists a positive number $\tilde{\lambda}$ depending on k such that

$$(5.19)\quad f(x^{k-1} + \tilde{\lambda}\nabla f(x^{k-1})) - f^{k-1} \geqq \left(\frac{\sigma^2}{\mu}\right) \min \{G(x^{k-1}),1\} > 0$$

and for each $i \in \{1,2, \ldots, m\}$

$$(5.20)\quad g_i(x^{k-1} + \tilde{\lambda}\nabla f(x^{k-1})) \geqq \left(\frac{\sigma^2}{\mu}\right) \min \left|g_i(x^{k-1}),1\right| > 0 .$$

Proof:

For some $k \geqq 1$ let

$$(5.21)\qquad x(\lambda) = x^{k-1} + \lambda\nabla f(x^{k-1})$$

$$(5.22)\qquad h_o(\lambda) = \lambda||\nabla f(x^{k-1})||^2 - \frac{1}{2}\lambda^2\mu||\nabla f(x^{k-1})||^2$$

and

$$(5.23)\qquad h(\lambda) = G(x^{k-1}) - \lambda\Delta||\nabla f(x^{k-1})|| - \frac{1}{2}\lambda^2\mu||\nabla f(x^{k-1})||^2$$

for $\lambda \geq 0$ such that $x(\lambda) \in S^1$ where μ , Δ and G are defined by (5.6), (4.5) and (5.10) respectively. Assumption (5.1) and (5.21) imply by the second order Talyor's theorem [31] that

$$(5.24) \quad f(x(\lambda)) = f(x^{k-1}) + \lambda\nabla f(x^{k-1})\cdot\nabla f(x^{k-1}) + \frac{1}{2}\lambda^2\nabla f(x^{k-1})H_0\left(\xi_0^{k-1}\right)\cdot\nabla f(x^{k-1})$$

and

$$(5.25) \quad \begin{aligned} g_i(x(\lambda)) = g_i(x^{k-1}) + \lambda\nabla g_i(x^{k-1})\cdot\nabla f(x^{k-1}) + \frac{1}{2}\lambda^2\nabla f(x^{k-1})H_i\left(\xi_i^{k-1}\right)\cdot \\ \cdot\nabla f(x^{k-1}) \qquad \text{for } i = 1,2, \ldots, m \end{aligned}$$

where ξ_i^{k-1} lies on the line segment connecting x^{k-1} and $x(\lambda)$ for $i = 0,1, \ldots, m$. Then (5.24), (5.8) and (5.22) imply

$$(5.26) \quad f(x(\lambda)) - f^{k-1} \leq h_0(\lambda) \qquad \text{for all } \lambda \geq 0 \text{ such that } x(\lambda) \in S^1 .$$

Similarly (5.25), the Cauchy-Schwarz inequality and the definitions of Δ and μ imply that for each $i \in \{1,2, \ldots, m\}$

$$g_i(x(\lambda)) \geq g_i(x^{k-1}) - \lambda\Delta||\nabla f(x^{k-1})|| - \frac{1}{2}\lambda^2\mu||\nabla f(x^{k-1})||^2$$

or

$$(5.27) \quad \begin{aligned} \left(\frac{G(x^{k-1})}{g_i(x^{k-1})}\right)g_i(x(\lambda)) \geq G(x^{k-1}) - \left(\frac{G(x^{k-1})}{g_i(x^{k-1})}\right)\cdot \\ \cdot\left[\lambda\Delta||\nabla f(x^{k-1})|| + \frac{1}{2}\lambda^2\mu||\nabla f(x^{k-1})||^2\right] . \end{aligned}$$

Then since $G(x^{k-1}) \geq g_i(x^{k-1})$ by the definition of G , (5.21) and (5.27) imply that for each $i \in \{1,2, \ldots, m\}$

so

(5.28) $$\left(\frac{G(x^{k-1})}{g_1(x^{k-1})}\right) g_1(x(\lambda)) \geqq h(\lambda) \qquad \text{for all } \lambda \geqq 0 \text{ such that } x(\lambda) \in S^1$$

Note that $h_o(0) = 0$ and $\frac{dh_o(0)}{d\lambda} = ||\nabla f(x^{k-1})||^2 > 0$ and $h(0) = G(x^{k-1}) > 0$ and $\frac{dh(0)}{d\lambda} = -\Delta||\nabla f(x^{k-1})||^2 < 0$ since $||\nabla f(x^{k-1})|| > 0$ by Assumption (5.2) and $\Delta > 0$ by the boundedness of S^1. Consider increasing λ from zero until either $h_o(\lambda) = h(\lambda)$ or $h_o(\lambda)$ is maximized whichever occurs first. If $\mu > 0$, let $\bar{\lambda}$ maximize $h_o(\lambda)$ for $\lambda \geqq 0$ which exists since $h_o(\lambda)$ is strictly concave for $\lambda \geqq 0$. Then

(5.29) $$\bar{\lambda} = \frac{1}{\mu}$$

and

(5.30) $$h_o(\bar{\lambda}) = \left(\frac{1}{2\mu}\right)||\nabla f(x^{k-1})||^2 .$$

If $\mu = 0$, define $\bar{\lambda} = +\infty$. Define $\hat{\lambda}$ by $h_o(\hat{\lambda}) = h(\hat{\lambda})$ so

$$\hat{\lambda}||\nabla f(x^{k-1})||^2 - \frac{1}{2}\hat{\lambda}^2\mu||\nabla f(x^{k-1})||^2 = G(x^{k-1}) - \hat{\lambda}\Delta||\nabla f(x^{k-1})||^2 - \frac{1}{2}\hat{\lambda}^2\mu||\nabla f(x^{k-1})||^2 .$$

Then

(5.31) $$\hat{\lambda} = \left(\frac{G(x^{k-1})}{||\nabla f(x^{k-1})||^2 + \Delta||\nabla f(x^{k-1})||}\right) > 0$$

and

$$h_o(\hat{\lambda}) = h(\hat{\lambda}) = \left(\frac{G(x^{k-1})||\nabla f(x^{k-1})||^2}{||\nabla f(x^{k-1})||^2 + \Delta||\nabla f(x^{k-1})||}\right) \cdot \tag{5.32}$$

$$\cdot\left[1 - \left(\frac{\mu}{2}\right)\left(\frac{G(x^{k-1})}{||\nabla f(x^{k-1})||^2 + \Delta||\nabla f(x^{k-1})||}\right)\right].$$

Now let $\tilde{\lambda} = \min\{\bar{\lambda},\hat{\lambda}\}$. If $\bar{\lambda} \leq \hat{\lambda}$ then $\tilde{\lambda} = \bar{\lambda}$ and by (5.30)

$$h(\tilde{\lambda}) \geq h_o(\tilde{\lambda}) = \left(\frac{1}{2}\right)\left(\frac{1}{\mu}\right)||\nabla f(x^{k-1})||^2 . \tag{5.33}$$

If $\bar{\lambda} > \hat{\lambda}$ then $\tilde{\lambda} = \hat{\lambda}$ and by (5.29) and (5.31)

$$\left(\frac{1}{\mu}\right) > \left(\frac{G(x^{k-1})}{||\nabla f(x^{k-1})||^2 + \Delta||\nabla f(x^{k-1})||}\right) \tag{5.34}$$

and by (5.32)

$$h(\tilde{\lambda}) = h_o(\tilde{\lambda}) = \left(\frac{1}{2}\right)\left(\frac{G(x^{k-1})||\nabla f(x^{k-1})||^2}{||\nabla f(x^{k-1})||^2 + \Delta||\nabla f(x^{k-1})||}\right) \cdot \tag{5.35}$$

$$\cdot\left[2 - \left(\frac{\mu G(x^{k-1})}{||\nabla f(x^{k-1})||^2 + \Delta||\nabla f(x^{k-1})||}\right)\right].$$

Then combining (5.34) and (5.35) yields

$$h(\tilde{\lambda}) = h_o(\tilde{\lambda}) > \left(\frac{1}{2}\right)\left(\frac{G(x^{k-1})||\nabla f(x^{k-1})||^2}{||\nabla f(x^{k-1})||^2 + \Delta||\nabla f(x^{k-1})||}\right)$$

which implies by the definition Δ_o

$$h(\tilde{\lambda}) = h_o(\tilde{\lambda}) > \left(\frac{1}{2}\right)\left(\frac{1}{\Delta_o^2 + \Delta\Delta_o}\right)G(x^{k-1})||\nabla f(x^{k-1})||^2 . \tag{5.36}$$

Combining (5.33) and (5.36) yields for either case

$$(5.37) \qquad h(\tilde\lambda) \geq h_o(\tilde\lambda) - \left(\frac{\epsilon^2}{n}\right) \min\,[G(x^{k-1}),1] > 0$$

where n is defined by (5.12) and ϵ is defined by (5.9). Then (5.37) together with (5.26) and (5.21) implies (5.19) and together with (5.28) implies for each $i \in \{1,2,\ldots,m\}$

$$g_i(x(\tilde\lambda)) \geq \left(\frac{\epsilon}{\ }\right)\left(\frac{g_i(x^{k-1})}{G(x^{k-1})}\right) \min\,[G(x^{k-1}),1] \geq \left(\frac{\epsilon^2}{n}\right) \min\left[g_i(x^{k-1}), \left(\frac{g_i(x^{k-1})}{G(x^{k-1})}\right)\right]$$

which along with (5.21) implies (5.20) since $G(x^{k-1}) \geq g_i(x^{k-1})$. ||

Before proceeding further with the first steepest ascent step two preliminary lemmas, one dealing with iteration $k = 1$ and the other with iterations $k \geq 2$, will be established. The results of these lemmas have different forms due to the fact that x^0, the starting point for iteration $k = 1$, is in general not an approximate center for some previous iteration.

<u>Lemma 5.2</u>:

$$(5.38) \qquad \frac{(f(x) - f^0)}{\min\,[G(x^0),1]} \leq \left(\frac{f^* - f^0}{G(x^0)}\right) \max\,[G(x^0),1] \qquad \text{for all } x \in \hat S^1$$

and for each $i \in \{1,2,\ldots,m\}$

$$(5.39) \qquad \frac{g_i(x)}{\min\left[g_i(x^0),1\right]} \leq \left(\frac{\tilde g}{G(x^0)}\right) \max\,[G(x^0),1] \qquad \text{for all } x \in \hat S^1 .$$

<u>Proof</u>:

By the definition of G for each $i \in \{1,2,\ldots,m\}$

$$(5.40)\qquad \frac{1}{\min\left[g_1(x^o),1\right]} \leqq \frac{1}{\min\left[G(x^o),1\right]} = \left(\frac{1}{G(x^o)}\right)\max\left[G(x^o),1\right] .$$

Then (5.38) follows from the equality relation in (5.40) since $f(x) \leqq f^*$ for all $x \in S^1$ and (5.39) follows from (5.40) since $g(x) \geqq \hat{g}$ for all $x \in S^1$.||

Lemma 5.3:

For each integer $k \geqq 2$

$$(5.41)\qquad \frac{(f(x) - f^{k-1})}{\min\left[G(x^{k-1}),1\right]} \leqq \left(\frac{f^* - f^o}{G(x^o)}\right)\left(\frac{e(f,\varepsilon)}{\beta}\right) \qquad \text{for all } x \in S^1$$

and for each $i \in \{1,2, \ldots, m\}$

$$(5.42)\qquad \frac{g_i(x)}{\min\left[g_i(x^{k-1}),1\right]} \leqq \left(\frac{e(\beta,\varepsilon)}{\beta}\right) \qquad \text{for all } x \in S^k .$$

Proof:

By Lemma 3.2

$$(5.43)\qquad f^* - f^{k-1} \leqq (f^{k-1} - f^{k-2})(\beta m + \varepsilon\gamma) \qquad \text{for } k = 2,3, \ldots$$

and by Lemma 3.5 and the definition of G

$$(5.44)\qquad G(x^{k-1}) \geqq \beta(f^{k-1} - f^{k-2})\left(\frac{G(x^o)}{f^* - f^o}\right)\left(\frac{1}{1 + \beta m + \varepsilon\gamma}\right) \qquad \text{for } k = 2,3, \ldots .$$

The definition of f^* , (5.43) and (5.44) imply for each integer $k \geqq 2$ that

$$(5.46)\quad \frac{(f(x) - f^{k-1})}{\min\,[G(x^{k-1}),1]} \leq \left(\frac{f^* - f^0}{G(x^0)}\right) \max\left[G(x^0), \left(m + \left(\frac{\varepsilon}{\delta}\right)\gamma\right)(1 + \varepsilon m + \varepsilon\gamma)\right]$$

for all $x \in S^1$

since for all $k \geq 1$

$$f(x) - f^{k-1} \leq f^* - f^0 = \left(\frac{f^* - f^0}{G(x^0)}\right)G(x^0) \qquad \text{for all } x \in S^1 .$$

Now since $G(x^0) \geq \delta$, (5.41) follows from (5.46) and (5.14). Also for each integer $k \geq 2$ by Lemma 3.1 and the definition of γ

$$(5.47)\quad f(x) - f^{k-1} \geq (f^{k-1} - f^{k-2})\left[m - \sum_{i=1}^{m} \frac{g_i(x)}{g_i(x^{k-1})} + \varepsilon\gamma\right] \qquad \text{for all } x \in S^1 .$$

Since for all $k \geq 1$, $\hat{S}^k \subset S^1$ and $f(x) \geq f^{k-1}$ for all $x \in \hat{S}^k$, (5.47) implies

$$0 \leq m - \sum_{i=1}^{m} \frac{g_i(x)}{g_i(x^{k-1})} + \left(\frac{\varepsilon}{\delta}\right)\gamma, \qquad \text{for all } x \in \hat{S}^k .$$

Therefore for each integer $k \geq 2$ and each $i \in \{1,2, \ldots, m\}$

$$\frac{g_i(x)}{g_i(x^{k-1})} \leq m + \left(\frac{\varepsilon}{\delta}\right)\gamma \qquad \text{for all } x \in \hat{S}^k$$

which implies

$$(5.48)\qquad \frac{g_i(x)}{\min\left[g_i(x^{k-1}),1\right]} \leq \max\left[\bar{g}, \left(m + \left(\frac{\varepsilon}{\delta}\right)\gamma\right)\right] \qquad \text{for all } x \in \hat{S}^k$$

since $g_i(x) \leq \bar{g}$ for all $x \in S^1$. Then (5.42) follows from (5.48) since

$1 < 1 + \beta m + \varepsilon\gamma$ implies

$$\max\left[\tilde{g}, m + \left(\frac{\varepsilon}{\beta}\right)\gamma\right] \leqq \left(\frac{e(\beta,\varepsilon)}{\beta}\right) . ||$$

The next lemma employs the results of Lemmas 5.1, 5.2 and 5.3 to obtain an upper bound on $d^k(x^k) - d^k\left(z_1^k\right)$ for each $k \geqq 1$.

<u>Lemma 5.4</u>:

$$(5.49) \quad d^1(x^1) - d^1\left(z_1^1\right) \leqq (1 + \beta m) \ln [\theta \max \{G(x^o),1\}]$$

and for $k = 2,3, \ldots$

$$(5.50) \qquad d^k(x^k) - d^k\left(z_1^k\right) \leqq (1 + \beta m) \ln \left[\theta\left(\frac{e(\beta,\varepsilon)}{\beta}\right)\right] .$$

<u>Proof</u>:

For some $k \geqq 1$ let $x(\lambda) = x^{k-1} + \lambda\nabla f(x^{k-1})$ for $\lambda \geqq 0$ such that $x(\lambda) \in \hat{S}^k$ and let λ^* be such that

$$d^k(x(\lambda^*)) = \max \{d^k(x(\lambda)) \mid \lambda \geqq 0 \text{ and } x(\lambda) \in \hat{S}^k\} .$$

Such a λ^* exists by Lemma 5.1 and the continuity of d^k on the bounded set $\hat{S}^k$. Then $z_1^k = x(\lambda^*)$ and with $\tilde{\lambda}$ as in Lemma 5.1

$$d^k\left(z_1^k\right) \geqq d^k(x^k + \tilde{\lambda}\nabla f(x^k))$$

and by (5.19) and (5.20)

$$(5.51) \quad d^k\left(z_1^k\right) \geqq \ln\left[\left(\frac{\sigma^2}{n}\right) \min \{G(x^{k-1}),1\}\right] + \beta \sum_{i=1}^{m} \ln\left[\left(\frac{\sigma^2}{n}\right) \min \left|g_i(x^{k-1}),1\right|\right] .$$

Then (5.51) and the definition of $d^k(x^k)$ imply for each integer $k \geqq 1$

$$d^k(x^k) - d^k(z_1^k) \leqq \ln\left[\left(\frac{n}{\sigma^2}\right)\left(\frac{(f(x^k) - f^{k-1})}{\min\{G(x^{k-1}),1\}}\right)\right] + \tag{5.52}$$

$$+ \beta \sum_{i=1}^{m} \ln\left[\left(\frac{n}{\sigma^2}\right)\left(\frac{g_i(x^k)}{\min\{g_i(x^{k-1}),1\}}\right)\right]$$

Then (5.49) follows from Lemma 5.2 and (5.52) with $k = 1$ and (5.50) follows from Lemma 5.3 and (5.52) with $k \geqq 2$ since $x^k \in \hat{S}^k$ for all $k \geqq 1$ and by (5.13)

$$\sigma \geqq \left(\frac{1}{n^2}\right)\left(\frac{f^* - f^0}{G(x^0)}\right)$$

and

$$\sigma \geqq \left(\frac{1}{n^2}\right)\left(\frac{\tilde{g}}{G(x^0)}\right) \geqq \left(\frac{1}{n^2}\right) \quad ||$$

By employing arguments similar to those used in proving the previous lemma, Lemmas 5.1, 5.2 and 5.3 may be combined to provide lower bounds on $f(x) - f^{k-1}$ and $g_i(x)$ for $i = 1,2, \ldots, m$ for all $x \in T^k$ where by (5.18) T^k contains the points z_j^k for $j \geqq 1$ generated by the modified steepest ascent algorithm.

Lemma 5.5

$$f(x) - f^0 \geqq \tilde{g}(b_1(\beta))^{-1} \qquad \text{for all } x \in T^1 \tag{5.53}$$

and for each $i \in \{1,2, \ldots, m\}$

$$g_i(x) \geqq \tilde{g}(b_1(\beta))^{-1/\beta} \qquad \text{for all } x \in T^1 \tag{5.54}$$

and for $k = 2,3, \ldots$

$$(5.55)\qquad f(x) - f^{k-1} \geqq (b(\sigma,\varepsilon))^{-1}\left(\frac{\sigma(\sigma,\varepsilon)}{b}\right) \min\,[G(x^{k-1}),1] \qquad \text{for all } x \in T^k$$

and for each $i \in \{1,2, \ldots, m\}$

$$(5.56)\qquad g_i(x) \geqq (b(\ ,\ ))^{-1/\beta}\left(\frac{\varepsilon(\ ,\varepsilon)}{\ }\right) \min\,[G(x^{k-1}),1] \qquad \text{for all } x \in T^k .$$

Proof

For each integer $k \geqq 1$

$$(5.57)\qquad d^k(x) = \ln\,(f(x) - f^{k-1}) + \beta \sum_{i=1}^{m} \ln g_i(x) \geqq d^k\left(z_1^k\right) \qquad \text{for all } x \in T^k .$$

Combining (5.57) with (5.51) yields

$$(5.58)\qquad \begin{aligned} \ln\,(f(x) - f^{k-1}) + \beta \sum_{i=1}^{m} \ln g_i(x) &\geqq \ln\left[\left(\frac{\sigma^2}{n}\right) \min\,[G(x^{k-1}),1]\right] + \\ &+ \beta \sum_{i=1}^{m} \ln\left[\left(\frac{\sigma}{n}\right) \min\,\left[g_i(x^{k-1}),1\right]\right] \qquad \text{for all } x \in T^k \end{aligned}$$

which implies

$$(5.59)\qquad \begin{aligned} \ln\,(f(x) - f^{k-1}) &\geqq \ln\left[\left(\frac{\sigma^2}{n}\right) \min\,[G(x^{k-1}),1]\right] - \\ &- \beta \sum_{i=1}^{m} \ln\left[\left(\frac{n}{\sigma}\right)\left(\frac{g_i(x)}{\min\,\left[g_i(x^{k-1}),1\right]}\right)\right] \qquad \text{for all } x \in T^k . \end{aligned}$$

Then by (5.39) of Lemma 5.2 and (5.59) with $k = 1$

$$\begin{aligned} \ln\,(f(x) - f^{0}) &\geqq -\ln\left[\left(\frac{n}{\sigma^2}\right)\left(\frac{\tilde{g}}{G(x^{0})}\right)\left(\frac{1}{b}\right) \max\,[G(x^{0}),1]\right] - \\ &- \beta m \ln\left[\left(\frac{n}{\sigma^2}\right)\left(\frac{\tilde{g}}{G(x^{0})}\right) \max\,[G(x^{0}),1]\right] \qquad \text{for all } x \in T^1 \end{aligned}$$

or since $\theta \geq \left(\frac{\eta}{\sigma^2}\right)\left(\frac{\tilde{f}}{G(x^0)}\right)$

$$\ln (f(x) - f^0) \geq -\ln \left\{\left(\frac{1}{\tilde{g}}\right)[\theta \max [G(x^0),1]]^{\delta m+1}\right\} \qquad \text{for all} \quad x \in T^1$$

which by (5.15) is equivalent to the desired result (5.53). In a similar manner (5.55) follows from (5.42) of Lemma 5.3, (5.59) with $k \geq 2$ and (5.16) since $\theta \geq \left(\frac{\eta}{\sigma^2}\right)$. Relation (5.58) also implies for each $i \in \{1,2, \ldots, m\}$

$$\begin{aligned} \ln g_i(x) \geq & -\left(\frac{1}{\omega}\right) \ln \left[\left(\frac{\eta}{\sigma^2}\right)\left(\frac{f(x) - f^{k-1}}{\min [G(x^{k-1}),1]}\right)\right] - \\ & - \sum_{\substack{j=1 \\ j \neq i}}^{m} \ln \left[\left(\frac{\eta}{\sigma^2}\right)\left(\frac{g_j(x)}{\min \left[g_j(x^{k-1}),1\right]}\right)\right] + \\ & + \ln \left[\left(\frac{\eta}{\sigma^2}\right)^{-1} \min \left[g_i(x^{k-1}),1\right]\right] \qquad \text{for all} \quad x \in T^k . \end{aligned} \tag{5.60}$$

Then for each $i \in \{1,2, \ldots, m\}$ by Lemma 5.2 and (5.60) with $k = 1$

$$\begin{aligned} \ln g_i(x) \geq & -\left(\frac{1}{\omega}\right) \ln \left[\left(\frac{\eta}{\sigma}\right)\left(\frac{f^* - f^0}{G(x^0)}\right) \max [G(x^0),1]\right] - \\ & - (m - 1) \ln \left[\left(\frac{\eta}{\sigma^2}\right)\left(\frac{\tilde{g}}{G(x^0)}\right) \max [G(x^0),1]\right] - \\ & - \ln \left[\left(\frac{\eta}{\sigma^2}\right)\left(\frac{\tilde{g}}{G(x^0)}\right)\left(\frac{1}{\tilde{g}}\right) \max [G(x^0),1]\right] \qquad \text{for all} \quad x \in T^1 \end{aligned}$$

or by the definition of θ

$$\ln g_i(x) \geq -\ln \left\{\left(\frac{1}{\tilde{g}}\right)[\theta \max [G(x^0),1]]^{m+1/\delta}\right\} \qquad \text{for all} \quad x \in T^1$$

which by (5.15) is equivalent to the desired result (5.54). In a similar manner (5.56) follows from Lemma 5.3, (5.60) with $k \geq 2$ and (5.16) since

$g_i(x^{k-1}) \geqq G(x^{k-1})$ for $i = 1, 2, \ldots, m$. ||

For $k = 1, 2, \ldots$ let $H^k(x)$ be the matrix of second partial derivatives of $d^k(x)$ for $x \in \hat{S}^k$. The results of Lemma 5.5 may be used to bound the norm of $H^k(x)$ for all $x \in T^k$.

Lemma 5.6.

For all $y \in E^n$

$$\sup_{x \in T^1} y[-H^1(x)]y \leqq n\left(\frac{1}{2\bar{g}}\right)\Big\{b_1(\beta) + \beta m(b_1(\beta))^{1/\beta} + \tag{5.61}$$

$$+ \left(\frac{1}{2\bar{g}}\right)\left[(b_1(\beta))^2 + \beta m(b_1(\beta))^{2/\beta}\right]\Big\}||y||^2$$

and for $k = 2, 3, \ldots$

$$\sup_{x \in T^k} y[-H^k(x)]y \leqq n\left(\frac{1}{2G(x^o)}\right)\Big\{[b(\beta,\varepsilon) + \beta m(b(\beta,\varepsilon))^{1/\beta}] \cdot$$

$$\cdot\left(\frac{f^* - f^o}{f^* - f^{k-1}}\right) + \left(\frac{1}{2G(x^o)}\right)[(b(\beta,\varepsilon))^2 + \beta m(b(\beta,\varepsilon))^{2/\beta}] \cdot \tag{5.62}$$

$$\cdot\left(\frac{f^* - f^o}{f^* - f^{k-1}}\right)^2\Big\}||y||^2 .$$

Proof:

For each integer $k \geqq 1$

$$\nabla d^k(x) = \frac{\nabla f(x)}{(f(x) - f^k)} + \beta \sum_{i=1}^{m} \frac{\nabla g_i(x)}{g_i(x)} \qquad \text{for } x \in \hat{S}^k$$

which implies by Assumption (5.1)

$$H^k(x) = \frac{H_o(x)}{(f(x) - f^{k-1})} - \frac{[\nabla f(x)\nabla f(x)^T]}{(f(x) - f^{k-1})^2} + \beta \sum_{i=1}^{m} \left\{\frac{H_i(x)}{g_i(x)} - \frac{[\nabla g_i(x)\nabla g_i(x)^T]}{g_i(x)^2}\right\}$$

where $H_i(x)$ for $i = 0,1,\ldots,m$ are defined by (5.4) and (5.5) and $[\nabla f(x)\nabla f(x)^T]$, for example, is an $n \times n$ symmetric matrix whose ij^{th} element is $\left(\frac{\partial f(x)}{\partial x_i}\right)\left(\frac{\partial f(x)}{\partial x_j}\right)$. Then for any $y \in E^n$

$$y[-H^k(x)]y = \frac{y[-H_0(x)]y}{(f(x) - f^{k-1})} + \frac{(\nabla f(x)\cdot y)^2}{(f(x) - f^{k-1})^2} + \beta \sum_{i=1}^{m} \left\{\frac{y[-H_i(x)]y}{g_i(x)} + \frac{(\nabla g_i(x)\cdot y)^2}{g_i(x)^2}\right\}$$

for $x \in S^k$

and by the definition of Λ and Relation (5.8) and the definitions of Δ_0 and Δ and the Cauchy-Schwarz inequality

(5.63)
$$y[-H^k(x)]y \le \left[\frac{\Lambda}{(f(x) - f^{k-1})} + \frac{\Delta_0^2}{(f(x) - f^{k-1})^2} + \beta \sum_{i=1}^{m} \left(\frac{\Lambda}{g_i(x)} + \frac{\Delta^2}{g_i(x)^2}\right)\right] \|y\|^2 \qquad \text{for all } x \in \hat{S}^k .$$

Then since $\eta = 2\max\left\{\Lambda, 2\Delta_0^2, 2\Delta^2\right\}$, (5.63) implies for any $y \in E^n$ that

(5.64)
$$y[-H^k(x)]y \le \left(\frac{\eta}{2}\right)\left[\frac{1}{(f(x) - f^{k-1})} + \frac{1}{2(f(x) - f^{k-1})^2} + \beta \sum_{i=1}^{m} \left(\frac{1}{g_i(x)} + \frac{1}{2g_i(x)^2}\right)\right] \|y\|^2 \qquad \text{for } x \in \hat{S}^k .$$

For the case when $k = 1$, Lemma 5.5 implies

(5.65)
$$\frac{1}{(f(x) - f^0)} \le \left(\frac{1}{\beta}\right) b_1(\beta) \qquad \text{for all } x \in T^1$$

and for each $i \in \{1,2,\ldots,m\}$

$$\frac{1}{g_1(x)} = \left(\frac{1}{b}\right)(h_1(x))^{1/\beta} \qquad \text{for all } x \in T^1 . \tag{5.66}$$

Then (5.61) follows from (5.65), (5.66) and (5.64) with $k = 1$. For the case when $k \geq 2$ (5.41) implies

$$\frac{(f^* - f^{k-1})}{\min\{G(x^{k-1}),1\}} \leq \left(\frac{f^* - f^o}{G(x^o)}\right)\left(\frac{c(\varepsilon,\beta)}{b}\right)$$

which when combined with Lemma 5.5 implies

$$\frac{1}{f(x) - f^{k-1}} \leq b(\beta,\varepsilon)\left(\frac{1}{G(x^o)}\right)\left(\frac{f^* - f^o}{f^* - f^{k-1}}\right) \qquad \text{for all } x \in T^k \tag{5.67}$$

and for each $i \in \{1,2, \ldots, m\}$

$$\frac{1}{g_i(x)} \leq (b(\varepsilon,\beta))^{1/\beta}\left(\frac{1}{G(x^o)}\right)\left(\frac{f^* - f^o}{f^* - f^{k-1}}\right) \qquad \text{for all } x \in T^k . \tag{5.68}$$

Then (5.62) follows from (5.67), (5.68) and (5.64) with $k \geq 2$.||

The next lemma provides an upper bound on $\ell(k)$, the number of steps required by the modified steepest ascent algorithm to find $x^k = z^k_{\ell(k)}$ starting from $x^{k-1} = z^k_o$ for each $k \geq 1$. Clearly $\ell(k) = 1$ and $x^k = z^k_1$ if $||\nabla d^k\left(z^k_1\right)|| \leq \varepsilon$. Otherwise $\ell(k) > 1$ and the remaining steepest ascent steps are carried out on T^k . For $k = 1,2, \ldots$ let

$$\mu_k = \sup_{x \in T^k} ||H^k(x)|| . \tag{5.69}$$

Lemma 5.6 implies the existence of μ_k for $k = 1,2, \ldots$ since for a negative semidefinite symmetric matrix such as $H^k(x)$

$$||H^k(x)|| = \sup \quad y[-H^k(x)]y . \tag{5.70}$$

$$\cdot(k) \cdot \binom{2\cdot}{\cdot}\left(d^{k}(\cdot) - d^{k}\left(\cdot_j\right)\right) + 1 .$$

... (5.2) ... the definitions of $d^k(\cdot)$ and I^k it is ... to see that I^k is a convex set. Assumption (5.1) implies ... second order Taylor's theorem that for $j \geq 1$

$$d^k\left(z_j^k + \lambda \nabla d^k\left(z_j^k\right)\right) = d^k\left(z_j^k\right) + \lambda \nabla d^k\left(z_j^k\right)\cdot\nabla d^k\left(z_j^k\right) +$$

(5.71)

$$+ \frac{1}{2}\lambda^2 \nabla d^k\left(z_j^k\right) H^k(\cdot)\nabla d^k\left(z_j^k\right) \qquad \text{for all } \lambda \geq 0 \text{ such that } z_j^k + \lambda \nabla d\left(z_j^k\right) \in I^k$$

where $\cdot$ lies on the line segment connecting z_j^k and $z_j^k + \lambda \nabla d^k\left(z_j^k\right)$. From (5.69) and (5.70)

(5.72)

$$\sup_{x \in I^k} y[-H^k(x)]y \leq \mu_k ||y||^2 \qquad \text{for all } y \in E^n .$$

Then (5.71) and (5.72) imply that for $j \geq 1$

$$d^k\left(z_j^k + \lambda \nabla d^k\left(z_j^k\right)\right) \geq d^k\left(z_j^k\right) + \lambda||\nabla d^k\left(z_j^k\right)||^2 - \frac{1}{2}\lambda^2\mu_k||\nabla d^k\left(z_j^k\right)||^2$$

(5.73)

$$\text{for all } \lambda \geq 0 \text{ such that } z_j^k + \lambda\nabla d^k\left(z_j^k\right) \in I^k .$$

Let λ^* maximize over nonnegative real numbers the function of λ on the right hand side of (5.73) which is a concave function of λ. Then $\frac{1}{\mu_k}$ ... 0 and since this function increases to its maximum value

$d^k\left(z_j^k\right) + \left(\frac{1}{2\mu_k}\right)||\nabla d^k\left(z_j^k\right)||^2$ as λ increases from 0 to λ^*, (5.73) implies $z_j^k + \lambda\nabla d^k\left(z_j^k\right) \in T^k$ for all $\lambda \in [0,\lambda^*]$. Then by the definition of z_{j+1}^k for $j \geqq 1$

$$d^k\left(z_{j+1}^k\right) \geqq d^k\left(z_j^k + \lambda^*\nabla d^k\left(z_j^k\right)\right) \geqq d^k\left(z_j^k\right) + \left(\frac{1}{2\mu_k}\right)||\nabla d^k\left(z_j^k\right)||^2$$

and if $j < \ell(k)$ then $||\nabla d^k\left(z_j^k\right)|| > \varepsilon$ which implies

$$d^k\left(z_{j+1}^k\right) \geqq d^k\left(z_j^k\right) + \left(\frac{1}{2\mu_k}\right)\varepsilon^2 .$$

Then by induction on j for $j = 1,2, \ldots, \ell(k) - 1$

$$d^k\left(z_{\ell(k)}^k\right) \geqq d^k\left(z_1^k\right) + (\ell(k) - 1)\left(\frac{\varepsilon^2}{2\mu_k}\right)$$

which is equivalent to the desired result since $x^k = z_{\ell(k)}^k$. ||

Now Lemmas 5.4, 5.6 and 5.7 and Corollary 3.14 may be combined with the definition of $\alpha(\beta,\varepsilon)$ to give an exponentially increasing function of k which upper bounds $\ell(k)$ for all $k \geqq 1$.

Theorem 5.8:

(5.74) $$\ell(1) \leqq a_1(\beta,\varepsilon) + 1$$

and for $k = 2,3, \ldots$

(5.75) $$\ell(k) \leqq a_2(\beta,\varepsilon)(\alpha(\beta,\varepsilon))^{k-1} + a_3(\beta,\varepsilon)(\alpha(\beta,\varepsilon))^{2(k-1)} + 1$$

where

$$a_1(\beta,\varepsilon) = \left(\frac{1}{\varepsilon}\right)\left(\frac{n}{g}\right)\left\{b_1(\beta) + \beta m(b_1(\beta))^{1/\beta} + \left(\frac{1}{2g}\right)\left[(b_1(\beta))^2 + \beta m(b_1(\beta))^{2/\beta}\right]\right\} \ln\,(b_1(\beta)) \tag{5.76}$$

$$a_2(\beta,\varepsilon) = \left(\frac{1}{\varepsilon^2}\right)\left(\frac{n}{G(x^0)}\right)[b(\beta,\varepsilon) + \beta m(b(\beta,\varepsilon))^{1/\beta}] \ln\,(b(\beta,\varepsilon)) \tag{5.77}$$

and

$$a_3(\beta,\varepsilon) = \left(\frac{1}{\varepsilon^2}\right)\left(\frac{n}{2}\right)\left(\frac{1}{G(x^0)}\right)^2[(b(\beta,\varepsilon))^2 + \beta m(b(\beta,\varepsilon))^{2/\beta}] \ln\,(b(\beta,\varepsilon)) \; . \tag{5.78}$$

<u>Proof</u>

From Lemma 5.5 and the definitions of $b_1(\beta)$ and $b(\beta,\varepsilon)$

$$d^1(x^1) - d^1\left(x_1^1\right) \leqq \ln\,(b_1(\beta)) \tag{5.79}$$

and for $k = 2,3,\ \ldots$

$$d^k(x^k) - d^k\left(x_1^k\right) \leqq \ln\,(b(\beta,\varepsilon)) \; . \tag{5.80}$$

Lemma 5.6, (5.69) and (5.70) imply

$$\mu_1 \leqq \left(\frac{1}{2}\right)\left(\frac{n}{g}\right)\left\{b_1(\beta) + \beta m(b_1(\beta))^{1/\beta} + \left(\frac{1}{2g}\right)\left[(b_1(\beta))^2 + \beta m(b_1(\beta))^{2/\beta}\right]\right\} \tag{5.81}$$

and for $k = 2,3,\ \ldots$

$$\mu_k \leqq \left(\frac{1}{2}\right)\left(\frac{n}{G(x^0)}\right)\left\{[b(\beta,\varepsilon) + \beta m(b(\beta,\varepsilon))^{1/\beta}]\left(\frac{f^* - f^0}{f^* - f^{k-1}}\right) + \left(\frac{1}{2G(x^0)}\right)[(b(\beta,\varepsilon))^2 + \beta m(b(\beta,\varepsilon))^{2/\beta}]\left(\frac{f^* - f^0}{f^* - f^{k-1}}\right)^2\right\} \; . \tag{5.82}$$

From Corollary 3.14 and (5.17)

$$(5.83) \qquad \left(\frac{f^* - f^o}{f^* - f^{k-1}}\right) \leq (\alpha(\beta,\varepsilon))^{k-1} \qquad \text{for } k = 1,2, \ldots$$

Then (5.74) follows from (5.79), (5.81) and Lemma 5.7 with $k = 1$ where $a_1(\beta,\varepsilon)$ is defined by (5.76) and (5.75) follows from (5.80), (5.82), (5.83) and Lemma 5.7 with $k \geq 2$ where $a_2(\beta,\varepsilon)$ and $a_3(\beta,\varepsilon)$ are defined by (5.77) and (5.78) respectively.||

It should be noted that it is possible to find an upper bound in terms of the definitions of this section for the factor $\alpha(\beta,\varepsilon) = \dfrac{1 + \beta\tilde{p}}{\beta\tilde{p}}$ appearing in Theorem 5.8. From (3.19) and the definition of G

$$(5.84) \qquad \tilde{u}_i \leq (1 + \beta m + \varepsilon\gamma)\left(\frac{f^* - f^o}{G(x^o)}\right) \qquad \text{for } i = 1,2, \ldots, m$$

and from (3.23)

$$(5.85) \qquad \tilde{p} \geq \sum_{i=1}^{m} \frac{u_i^*}{\tilde{u}_i} \qquad \text{for all } u^* = \left(u_1^*, u_2^*, \ldots, u_m^*\right) \varepsilon U^* .$$

For $x^* \varepsilon X^*$ and $u^* \varepsilon U^*$

$$\nabla f(x^*) = -\sum_{i=1}^{m} u_i^* \nabla g_i(x^*)$$

and by the definitions of σ and Δ and the triangle inequality

$$(5.86) \qquad \sigma \leq ||\nabla f(x^*)|| \leq \sum_{i=1}^{m} u_i^* ||\nabla g_i(x^*)|| \leq \Delta \sum_{i=1}^{m} u_i^* .$$

Combining (5.84), (5.85) and (5.86) yields

$$\hat{p} = \left(\frac{\omega(x^{0})}{f^{*} - f^{0}}\right)\left(\frac{\cdot}{\cdot}\right)\left(\frac{1}{1 + m + \epsilon_{1}}\right)$$

which implies

$$\alpha(\beta,\epsilon) = \frac{1 + \hat{p}}{\hat{p}} \leq \frac{\left(1 + \cdot\left[1 + (\cdot)\left(\frac{\omega(x^{0})}{f^{*} - f^{0}}\right)\right] + \cdot\right)}{\cdot\left(\frac{\beta}{\Delta}\right)\left(\frac{\omega(x^{0})}{f^{*} - f^{0}}\right)} .$$

This bound could also be used in conjunction with Corollaries 3.9, 3.14 and 3.15 and Theorem 3.16 to obtain corresponding bounds which replace the dependence on f^{*} with dependence on ω and Δ.

By combining the results of Theorem 5.8 and Corollary 3.12 an upper bounding function of t may be found for the total number of steepest ascent steps required to find an x^k starting from x^0 such that $f^{*} - f(x^k) \leq t$ where t is a termination parameter for the algorithm.

<u>Theorem 5.9</u>

Let $a_1(\beta,\epsilon)$, $a_2(\beta,\epsilon)$ and $a_3(\beta,\epsilon)$ be as defined in Theorem 5.8 and let $n(t)$ be the total number of steepest ascent steps required to find a point $x^k \in \hat{S}^k$ starting from $x^o \in \hat{S}$ such that $f^{*} - f(x^k) \leq t$ where $t < f^{*} - f(x^o)$. Then

$$n(t) \leq k(t,\beta,\epsilon) + 1 + a_1(\beta,\epsilon) + a_2(\beta,\epsilon)\left(\frac{\alpha(\beta,\epsilon)}{\alpha(\beta,\epsilon) - 1}\right)\cdot$$

$$\cdot\left[(\alpha(\beta,\epsilon))^{k(t,\beta,\epsilon)} - 1\right] + a_3(\beta,\epsilon)\left[\frac{(\alpha(\beta,\epsilon))^{2}}{(\alpha(\beta,\epsilon))^{2} - 1}\right]\cdot \tag{5.87}$$

$$\cdot\left[(\alpha(\beta,\epsilon))^{2k(t,\beta,\epsilon)} - 1\right]$$

where

$k(t,\beta,\varepsilon)$ is the greatest integer less than $\dfrac{\ln\left(\frac{f^* - f^o}{t}\right)}{\ln\left(\frac{1 + \beta m + \varepsilon\gamma}{\beta m + \varepsilon\gamma}\right)}$.

Proof:

From Corollary 3.12 if $k \geqq \dfrac{\ln\left(\frac{f^* - f^o}{t}\right)}{\ln\left(\frac{1 + \beta m + \varepsilon\gamma}{\beta m + \varepsilon\gamma}\right)}$ then $f^* - f^k \leqq t$ which implies $f^* - f^{k(t,\beta,\varepsilon)+1} \leqq t$ and

$$n(t) \leqq \sum_{k=1}^{k(t,\beta,\varepsilon)+1} \ell(k) \tag{5.88}$$

where $\ell(k)$ is number of steepest ascent steps required to solve subproblem k . For the case when $k(t,\beta,\varepsilon) = 0$, (5.87) follows immediately from (5.88) and (5.74). For the case when $k(t,\beta,\varepsilon) \geqq 1$, (5.88) and Theorem 5.6 imply

$$n(t) \leqq k(t,\beta,\varepsilon) + 1 + a_1(\beta,\varepsilon) + a_2(\beta,\varepsilon)\alpha(\beta,\varepsilon) \sum_{k=2}^{k(t,\beta,\varepsilon)+1} (\alpha(\beta,\varepsilon))^{k-2} +$$

$$+ a_3(\beta,\varepsilon)(\alpha(\beta,\varepsilon))^2 \sum_{k=2}^{k(t,\beta,\varepsilon)+1} \alpha(\beta,\varepsilon)^{2(k-2)}$$

which is equivalent to (5.87).||

REFERENCES

[1] Bui-Trong-Lieu and P. Huard, "La Méthode des Centres dans un Espace Topologique," Numer. Math., Vol. 8, No. 1, pp. 56-67, (1966).

[2] Carroll, C. W., "The Created Response Surface Technique for Optimizing Nonlinear Restrained Systems," Opns. Res., Vol. 9, pp. 169-184, (1961).

[3] Cauchy, A., "Méthode Générale pour la Résolution des Systèmes d'Équations Simultanées," C. R. Acad. Sci. Paris, Vol. 25, pp. 536-538, (1847).

[4] Curry, H., "The Method of Steepest Descent for Non-linear Minimization Problems," Quart. Appl. Math., Vol. 2, pp. 258-261, (1944).

[5] Faure, P., "Note Sur la Rapidité de la Convergence de la Méthode des Centres," Note E.D.F. No. HR 750 /5

[6] Faure, P. and P. Huard, "Résultats Nouveaux Relatifs à la Méthode des Centres," Paper presented at Fourth International Conference on Operations Research, Boston, (1966).

[7] Fiacco, A. V., "Sequential Unconstrained Minimization Methods for Nonlinear Programming," Thesis, Northwestern University, Evanston, Illinois, (1967)

[8] Fiacco, A. V. and G. P. McCormick, "The Sequential Unconstrained Minimization Technique for Nonlinear Programming, A Primal-Dual Method," Management Sci., Vol. 10, pp. 360-366, (1964).

[9] Fiacco, A. V. and G. P. McCormick, "Computational Algorithm for the Sequential Unconstrained Minimization Technique for Nonlinear Programming," Management Sci., Vol. 10, pp. 601-617, (1964).

[10] Fiacco, A. V. and G. P. McCormick, "The Sequential Unconstrained Minimization Technique without Parameters," Opns. Res., Vol. 15, pp. 820-827, (1967).

[11] Fiacco, A. V. and G. P. McCormick, NONLINEAR PROGRAMMING: SEQUENTIAL UNCONSTRAINED MINIMIZATION TECHNIQUES, John Wiley and Sons, Inc., New York, (1968).

[12] Frisch, K. R., "Principles of Linear Programming with Particular Reference to the Double Gradient Form of the Logarithmic Potential Method," Memorandum of October 18, 1954, University Institute of Economics, Oslo.

[13] Frisch, K. R., "The Logarithmic Potential Method of Convex Programming," Memorandum of May 13, 1955, University Institute of Economics, Oslo.

[14] Goldstein, A. A., CONSTRUCTIVE REAL ANALYSIS, Harper and Row, New York, (1967).

[15] Huard, P., "Résolution des P.M. à Constraintes Non-linéaires par la Méthode des Centres," Note E.D.F. No. HR 5, 690 du 6 Mai 1964; also, "Resolution of Mathematical Programming with Nonlinear Constraints by the Method of Centers," in J. Abadie (ed.), NONLINEAR PROGRAMMING, North-Holland Publishing Co., Amsterdam, pp. 207-219, (1967).

[16] Huard, P., "Programmation Mathématique Convexe," R.I.R.O., Vol. 7, pp. 43-59, (1968).

[17] John F., "Extremum Problems with Inequalities as Subsidiary Conditions," in Studies and Essays, Courant Anniversary Volume, Interscience, New York, pp. 187-204, (1948).

[18] Kleibohm, K., "Äquivalence eines Optimierungsproblems mit Restriktionen und einer Folge von Optimierungsproblemen orne Restriktionen," Unternehmensforschung, Vol. 11, pp. 111-118, (1967).

[19] Kuhn, H. W. and A. W. Tucker, "Nonlinear Programming," in J. Neyman (ed.), Proceedings of the Second Berkeley Symposium in Mathematical Statistics and Probability, pp. 481-492, University of California Press, Berkeley, California, (1951).

[20] Levitan, E. S. and B. T. Polyak, "Constrained Minimization Methods," Zh. vychisl. Mat. mat. Fiz. (in Russian), Vol. 6, pp. 787-823, (1966), also, U.S.S.R. Computational Mathematics and Mathematical Physics (in English), pp. 1-50, (1968).

[21] Lootsma, F. A., "Logarithmic Programming: A Method of Solving Nonlinear-programming Problems," Philips Res. Repts., Vol. 22, pp. 329-344, (1967).

[22] Lootsma, F. A., "Extrapolation in Logarithmic Programming," Philips Res. Repts., Vol. 23, pp. 108-116, (1968).

[23] Lootsma, F. A., "Constrained Optimization via Penalty Functions," Philips Res. Repts., Vol. 23, pp. 408-423, (1968).

[24] Lootsma, F. A., "Constrained Optimization via Parameter-free Penalty Functions," Philips Res. Repts., Vol. 23, pp. 424-437, (1968).

[25] Mangasarian, O. L., "Pseudo-convex Functions," SIAM J. Control, Vol. 3, pp. 281-290, (1965).

[26] Mangasarian, O. L., NONLINEAR PROGRAMMING, McGraw-Hill Book Co., New York, (1969).

[27] Parisot, G. R., "Résolution Numerique Approchée du Problème de Programmation Linéaire par Application de la Programmation Logarithmique," Rev. Fr. Recherche Operationelle, Vol. 20, pp. 227-259, (1961).

[28] Pironneau, O. and E. Polak, "On the Rate of Convergence of Certain Methods of Centers," Memorandum No. ERL-M296, Electronics Research Laboratory, University of California, Berkeley, (1971).

[29] Polak, E. COMPUTATIONAL METHODS IN OPTIMIZATION, Academic Press, New York, (1971).

[30] Rockafellar, R. T., CONVEX ANALYSIS, Princeton University Press, Princeton, New Jersey, (1970).

[31] Rudin, W., PRINCIPLES OF MATHEMATICAL ANALYSIS, McGraw-Hill Book Co., New York, (1964).

[32] Topkis, D. M., "Cutting Plane Methods without Nested Constraint Sets," Opns. Res., Vol. 18, pp. 404-413, (1970).

[33] Topkis, D. M. and A. F. Veinott, "On the Convergence of Some Feasible Direction Algorithms for Nonlinear Programming," SIAM J. Control, Vol. 5, pp. 268-279, (1967).

[34] Tremolières, R., "La Méthode des Centres à Troncature Variable," These, Paris, (1968).

[35] Zangwill, W. I., NONLINEAR PROGRAMMING: A UNIFIED APPROACH, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, (1969).