SUPERLINEAR CONVERGENCE OF A MINIMAX METHOD

by

S. P. Han

TR78-336

Department of Computer Science
Cornell University
Ithaca, NY 14853

* Research supported by National Science Foundation under grant number ENG 77-02647.

SUPERLINEAR CONVERGENCE OF A MINIMAX METHOD*

by

S. P. Han

Department of Computer Science
Cornell University
Ithaca, NY 14853

Abstract

To solve a minimax problem Han [1977b] suggested the use of quadratic programs to find search directions. If the matrices in the quadratic programs are positive definite, the method can be shown convergent globally. In this paper we study that for efficiency the matrices should also be good approximations to a certain convex combination of Hessians on some subspace. Therefore, we suggest Powell's scheme [Powell 1977] for updating these matrices. By doing so, we can avoid computing Hessians. Meanwhile, the matrices maintain positive definiteness and Han's global convergence theorems can apply. Besides, the convergence of the resulting method is superlinear, indeed.

---

* Research supported by National Science Foundation under grant number ENG 77-02647.

1. Introduction

Consider the minimax problem

(1.1) $$\min_{x \in R^n} \max_{i=1,\dots,m} \{f_i(x)\}$$

where $f_1, f_2, \dots, f_m$ are real-valued functions on R^n which have continuous second derivatives. The minimization of a smooth function may be viewed as a special case of (1.1) with $m = 1$, for which variable metric methods are very effective. Many attempts have been made for the extension of variable metric methods to the more general problem (1.1). This work is a continuing effort on the approach adopted by Han [7], who makes use of nonlinear programming techniques.

The equivalence of a nonlinear programming problem to a minimax problem is apparent and has been extensively exploited. However, the attention has seemingly been focused on applying minimax techniques to nonlinear programming [see 1,2,3,11, for instance]. This is perhaps because a completely satisfactory nonlinear programming method was lacking and the minimax problem was deemed better understood. The advent of new nonlinear programming techniques seems able to reverse the trend. To tackle the minimax problem Han [7] suggests the recently developed and very successful variable metric methods [5,6,8,9] to solve its equivalent nonlinear programming problem

(1.2) $$\min_{(p,\delta)} \quad \delta$$

$$\text{s.t.} \quad f_i(x) \leq \delta \quad i = 1,2,\dots,m.$$

More specifically, to generate search directions we iteratively solve the quadratic programming problem

$$\min_{(d,\delta)} \quad \delta + \tfrac{1}{2}d^T B_k d$$
$$\text{s.t.} \quad f_i(x_k) + f_i'(x_k)^T d \leq \delta \qquad i = 1,\ldots,m.$$

If the matrices $\{B_k\}$ are positive definite and a suitable line-search is applied, the method converges globally [7]. We study in this paper that, for efficiency, the matrix B_k should also be a good approximation to a certain convex combination of the Hessians $\{f_i''(x_k)\}$ on some subspace. In this context Powell's scheme [9] is particularly suitable for updating these matrices. Firstly, as using other variable metric updating schemes, we can avoid computing the Hessians. Secondly, the updated matrices are positive definite and Han's [7] global convergence theorems can apply. In this paper we analyze the convergence rate of this method. It is superlinear, indeed.

It may be worthwhile pointing out another important advantage of this approach. The problem with constraints will not cause any extra difficulties. To handle the constrained minimax problem

$$\min_{x \in S} \max_{i=1,\ldots,m} \{f_i(x)\}$$

where

$$S = \{x : g_i(x) \leq 0,\ j = 1,\ldots,r;\ h_\ell(x) = 0,\ \ell = 1,\ldots,q\},$$

we may iteratively solve the following quadratic programming problem

$$\min_{(d,\delta)} \quad \delta + \frac{1}{2}d^T B_k d$$

$$\begin{aligned}
\text{s.t.} \quad & f_i(x_k) + f_i'(x_k)^T d \leq \delta \qquad i = 1,\ldots,m, \\
& g_j(x_k) + g_j'(x_k)^T d \leq 0 \qquad j = 1,\ldots,r, \\
& h_\ell(x_k) + h_\ell'(x_k)^T d = 0 \qquad \ell = 1,\ldots,q.
\end{aligned}$$

All our analysis and results can be carried over to this case routinely. Restricting ourselves only to the unconstrained problem is merely for reason of convenience.

The organization of this paper is as follows: in the next section we study the optimality conditions for this problem. In Section 3 we state the method. Finally, in Section 4, the analysis of convergence rate of the method is given.

Some nonstandard notation used in this paper is described below. We use $f_I'(x)$ to denote the matrix whose columns are the gradients $f_i'(x)$ with $i \in I$, where I is a given index set. We use the symbol e to denote the column vector of ones, whose dimension will be clear from the context. We will also use $L'(x,v)$ and $L''(x,v)$ to denote the gradient and the Hessian, respectively, of the function $L(x,v)$ with respect to x only. The 2-norm is used throughout this paper.

2. Optimality Conditions

The problem to be considered can be stated as

$$\min_{x \in R^n} \psi(x) \tag{2.1}$$

where

$$\psi(x) = \max_{i=1,\ldots,m} \{f_i(x)\}.$$

In this section we are concerned with conditions for a point to be a local solution to Problem (2.1). To facilitate our discussion we introduce for any point x in R^n an associated index set

$$I(x) = \{i : f_i(x) = \psi(x)\}$$

and an associated convex set Conv(x) which is the convex hull of the gradients $f_i'(x)$ with $i \in I(x)$.

Some necessary conditions for a solution to Problem (2.1) are known in the literature. We state them in the following theorem.

<u>Theorem 2.1</u> If a point x^* is a local solution to Problem (2.1) then

(1) $\min_{||d||=1} \max_{i \in I(x^*)} \{f_i'(x^*)^T d\} \geq 0;$

(2) $0 \in \mathrm{Conv}(x^*);$

(3) there exists some v^* in R^m such that

(2.2)

(a) $\sum_{i=1}^{m} v_i^* = 1,$

(b) $\sum_{i=1}^{m} v_i^* f_i'(x^*) = 0,$

(c) $v^* \geq 0,$

(d) $v_i^*(f_i(x^*) - \psi(x^*)) = 0 \quad i = 1,\ldots,m.$

Moreover, Conditions (1), (2) and (3) are equivalent. ■

The necessity of Condition (1) becomes apparent when we note that the directional derivative of the function ψ at a point x and in a direction d is given by

$$\psi'(x;d) = \max_{i \in I(x)} \{f_i'(x)^T d\}. \tag{2.3}$$

The equivalence of Conditions (1) and (2) can be established by using a separation theorem. We refer its proof to Demjanov [3]. Conditions (2) and (3) are obviously the same statement.

In the sequel we will call x* a stationary point if it satisfies these necessary conditions and will call (x*,v*) a stationary pair. Note also that Condition (3) above is just the Kuhn-Tucker condition of the nonlinear programming problem (1.2). Therefore, we may define for Problem (2.1) a Lagrangian function by

$$L(x,v) = \sum_{i=1}^{m} v_i f_i(x)$$

and may call v* a Lagrange multiplier associated with the stationary point x*. For a Lagrange multiplier v* there is also an associated index set defined by

$$J(v^*) = \{i : v_i^* > 0\}.$$

Some first order sufficiency conditions also exist for Problem (2.1). A local minimum point can sometimes be completely detected through only first order informations. This is uncommon to the smooth case, where, without an additional condition like convexity, the first derivatives can only inform us of stationary points.

<u>Theorem 2.2</u> A point x* is a strict local minimum point of Problem (2.1) if any one of the following conditions holds:

(1) $\min\limits_{||d||=1} \max\limits_{i\in I(x^*)} \{f_i'(x^*)^T d\} > 0;$

(2) the null vector is an interior point of the set Conv(x*);

(3) there exists some v* in R^m such that (x*,v*) is a stationary pair and the matrix

$$A(x^*,v^*) = \begin{bmatrix} f_J'(x^*) \\ e^T \end{bmatrix}$$

has full row rank, where J = J(v*).

Moreover, Conditions (1), (2) and (3) are equivalent.

<u>Proof</u>: The sufficiency of Condition (1) is obvious. For the proof of the equivalence of Conditions (1) and (2) we also refer to Demjanov [3]. We only need to show that (3) is equivalent to (2). Consider the following system

$$\begin{bmatrix} f_I'(x^*) \\ e^T \end{bmatrix} u = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad u \geqq 0,$$

where I = I(x*). Condition (3) is necessary and sufficient for the solvability of the following perturbed system

$$\begin{bmatrix} f_I'(x^*) \\ e^T \end{bmatrix} u = \begin{bmatrix} \varepsilon \\ 1 \end{bmatrix}, \quad u \geqq 0,$$

where ε is a sufficiently small but arbitrary vector [see 10]. This is equivalent to the statement that the null vector is an interior point of Conv(x*). The proof is then completed. ■

In the rest of this section we will study second order sufficiency conditions, which are relevant to the analysis of our method. We first define a set at a point x by

$$T(x) = \{d : \max_{i \in I(x)} \{f_i'(x)^T d\} = 0\}.$$

This may be called the tangent cone of the function ψ at x because (2.3).

<u>Theorem 2.3</u> A sufficient condition for a point x* to be a strict local minimum point of the function ψ is that x* is a stationary point and has a Lagrange multiplier v* such that

$$d^T L''(x^*, v^*) d > 0$$

for any nonzero vector d in the tangent cone $T(x^*)$.

<u>Proof</u>: We prove this theorem by contradiction. If x* is not a strict local minimum point then there is a sequence $\{x_k\}$ converging to x* such that

$$\psi(x_k) \leqq \psi(x^*).$$

Let

$$d_k = \frac{x_k - x^*}{||x_k - x^*||}, \quad \alpha_k = ||x_k - x^*||.$$

then for any i in $I(x^*)$ and any k there is a vector $\xi_{k,i}$ in the line segment between the two points x_k and x* such that

$$(2.4) \qquad \alpha_k f_i'(x^*)^T d_k + \frac{1}{2}\alpha_k^2 d_k^T f_i''(\xi_{k,i}) d_k = f_i(x_k) - f_i(x^*) \leqq \psi(x_k) - \psi(x^*) \leqq 0.$$

Passing to a subsequence, if necessary, we have $d_k \to \bar{d}$ for some $\bar{d}$

with $||\bar{d}|| = 1$. It is clear from (2.4) that for any i in $I(x^*)$

$$f_i'(x^*)^T\bar{d} \leq 0.$$

Since x^* is stationary, it follows from Theorem 2.1 that

$$\max_{i \in I(x^*)} \{f_i'(x^*)^T\bar{d}\} \geq 0.$$

Therefore, $\bar{d}$ is in the tangent cone $T(x^*)$. Taking the equality $\sum_{i=1}^{m} v_i^* f_i'(x^*) = 0$ into account, we then have from (2.4) that

$$\bar{d}^T L''(x^*,v^*)\bar{d} \leq 0.$$

This contradicts our assumption and hence the proof is completed. ■

It may be interesting to point out that, if we have $T(x^*) = \{0\}$ at a stationary point x^*, then the assumption of Theorem 2.3 are trivially satisfied. Note that the condition $T(x^*) = \{0\}$ is just Condition (1) in Theorem 2.2, which is a first order sufficiency condition.

The central assumption in Theorem 2.3 is that the Hessian $L''(x^*,v^*)$ is positive definite on the tangent cone $T(x^*)$. The set $T(x^*)$ is independent of the particular Lagrange multiplier v^* under consideration. A stronger assumption may be that the Hessian $L''(x^*,v^*)$ is positive definite on the subspace

(2.5) $\Omega(x^*,v^*) = \{d : f_i'(x^*)^T d = 0 \text{ for all } i \in J(v^*)\}.$

<u>Lemma 2.4</u> If (x^*,v^*) is a stationary pair then

$$T(x^*) \subset \Omega(x^*,v^*).$$

If, in addition, the strict complementarity condition $I(x^*) = J(v^*)$ holds, then

$$T(x^*) = \Omega(x^*,v^*)$$

<u>Proof:</u> To prove the first part we first note that if a vector d is in $T(x^*)$ then for any $i \in J(x^*)$ we have

$$f_i'(x^*)^T d \leq \max_{i \in I(x^*)} \{f_i'(x^*)^T d\} = 0.$$

Note also that

$$\sum_{i \in J(v^*)} v_i^* f_i'(x^*)^T d = 0 \text{ and } v^* \geq 0.$$

Thus, we have that $f_i'(x^*)^T d = 0$ for all i in $J(v^*)$ and d is in $\Omega(x^*,v^*)$. This proves the first part.

The second part follows trivially from the fact that if $I(x^*) = J(v^*)$ then

$$\max_{i \in I(x^*)} \{f_i'(x^*)^T d\} = \max_{i \in J(v^*)} \{f_i'(x^*)^T d\}.$$

■

Without the strict complementarity condition the tangent cone $T(x^*)$ may be a proper subset of $\Omega(x^*,v^*)$. This can be seen from the following simple example. Let $\psi(x) = \max\{x,x^2\}$. Then a stationary point is at $x^* = 0$ and it has a Lagrange multiplier $v^* = (0,1)^T$. In this case the set $\Omega(x^*,v^*)$ contains all the real numbers but $T(x^*)$ contains only non-positive numbers.

<u>Theorem 2.5</u> A sufficient condition for a point x^* to be a strict local minimum point of the function ψ is that x^* is a stationary point and has a Lagrange multiplier v^* such that

$$d^T L''(x^*,v^*)d > 0$$

for all nonzero vector d in $\Omega(x^*,v^*)$. ■

A different second order sufficiency condition has been studied by Demjanov and Malzemov [4]. Their condition is independent of Lagrange multipliers but takes into account an ε-tangent cone defined by

$$T_\varepsilon(x^*) = \{d : 0 \leq \max_{i\in I(x^*)} \{f'(x^*)^T d\} \leq \varepsilon||d||\}.$$

For comparison we state this condition below.

<u>Theorem 2.6</u> A point x^* is a strict local minimum point for $\psi(x)$ if x^* is a stationary point and there is a positive number ε such that

$$\max_{i\in K(x^*,d)} \{d^T f_i''(x^*)d\} > 0$$

for any nonzero vector d in $T_\varepsilon(x^*)$, where

$$K(x^*,d) = \{i : i \in I(x^*) \text{ and } f_i'(x^*)^T d = \psi'(x^*;d)\}.$$ ■

3. The Method

In this section we describe the method. We are content with finding a stationary pair, say (x^*,v^*). Assume that, at the k-th iteration, we have an estimate (x_k,v_k) of this pair and also have a matrix B_k which is an estimate to the Hessian $L''(x_k,v_k)$. To produce a new estimate (x_{k+1},v_{k+1}) we solve the quadratic programming problem

$$\text{(3.1)} \qquad \min_{(d,\delta)} \quad \delta + \tfrac{1}{2}d^T B_k d$$

$$\text{s.t.} \quad f_i(x_k) + f_i'(x_k)^T d \leq \delta \quad i = 1,\ldots,m.$$

Let (d_k, δ_k) be a Kuhn-Tucker point of (3.1), then we set $x_{k+1} = x_k + \alpha_k d_k$, where α_k is a suitable stepsize satisfying

$$(3.2) \qquad \psi(x_k + \alpha_k d_k) - \psi(x_k) \leqq - \omega_k \alpha_k d_k^T B_k d_k$$

for some $0 < \omega < \frac{1}{2}$. For the details of this line-search procedure and the validity of Condition (3.2) we refer to Han [7]. For the Lagrange multiplier v^* we take a Lagrange multiplier of the quadratic programming problem (3.1) as its new estimate v_{k+1}.

To find d_k and v_{k+1} we may also equivalently solve the dual problem of (3.1)

$$(3.3) \qquad \begin{aligned} \min_{v \in R^m} \quad & \tfrac{1}{2} v^T f'(x_k)^T B_k^{-1} f'(x_k) v - v^T f(x_k) \\ \text{s.t.} \quad & \sum_{i=1}^{m} v_i = 1, \\ & v \geqq 0. \end{aligned}$$

Let v_{k+1} be a solution to (3.3) then the vector d_k can be rediscovered by

$$d_k = -B_k^{-1} f'(x_k) v_{k+1}.$$

The matrix B_k in the quadratic programs (3.1) and (3.2) is preferrably positive definite. The reason is twofold. First, Han [7] has proven that the positive definiteness is very essential for the global convergence. Secondly, when B_k is positive definite, the quadratic programs can be much more effectively solved. Powell's scheme [9] is, therefore, particularly suitable for updating B_{k+1}. To simplify the description of this scheme we drop the subscript "k" and use the symbol "$\bar{\ }$" to replace the subscript "k+1". Let

$$s = \bar{x} - x \text{ and } y = L'(\bar{x},\bar{v}) - L'(x,\bar{v})$$

and define a vector z by

$$z = \theta y + (1 - \theta)Bs$$

where the member θ is given the value

$$\theta = \begin{cases} 1 & , s^T y \geq 0.2 s^T Bs \\ \dfrac{0.8 s^T Bs}{s^T Bs - s^T y} & , s^T y < 0.2 s^T Bs \end{cases} .$$

Then we construct the matrix $\bar{B}$ by

$$\bar{B} = B - \frac{Bss^T B}{s^T Bs} + \frac{zz^T}{s^T z} .$$

This is just the BFGS scheme when $\theta = 1$. It should also be pointed out that the matrix $\bar{B}$ is positive definite as long as so is the matrix B. We refer to Powell [9] for the rationale of this scheme. The description of the method is then completed.

4. Rate of Convergence

The method has been shown convergent globally [7]. In this section we assume that a sequence $\{x_k\}$ generated from the method converges to a stationary point x^* of Problem (2.1) and the stepsize one is used eventually, and then we analyze the rate of convergence of this sequence.

In many practical problems it is very likely that a stationary point has more than one Lagrange multipliers. This, indeed, causes difficulties. Among many Lagrange multiplies some are particularly of our concern. We define them below.

Definition: A Lagrange multiplier v^* of a stationary point x^* is regular if the matrix

$$A(x^*,v^*) = \begin{bmatrix} f'_J(x^*) \\ e^T \end{bmatrix}$$

has full column rank, where $J = J(v^*)$.

For gaining some insights it may be worthwhile mentioning that the following three statements are equivalent:

(1) $A(x^*,v^*)$ has full column rank;

(2) for some j in $J(v^*)$ the set of vectors $\{f'_i(x^*) - f'_j(x^*)\}$ with $i \in J(v^*)\backslash\{j\}$ are linearly independent;

(3) for any j in $J(v^*)$ the set of vectors $\{f'_i(x^*) - f'_j(x^*)\}$ with $i \in J(v^*)\backslash\{j\}$ are linearly independent.

Here, we adopt the convention that an empty set of vectors are linearly independent.

Theorem 4.1 Any stationary point has at least one regular Lagrange multiplier.

Proof: Let x^* be a stationary point and v^* be an associated Lagrange multiplier. It suffices to show that if v^* is not regular then we can construct another Lagrange multiplier, say $\bar{v}$, such that $J(\bar{v})$ is a proper subset of $J(v^*)$.

Without loss of generality we may assume that $J(v^*)$ contains only the first r indices. Because v^* is not regular there are numbers $\alpha_1,\alpha_2,\ldots,\alpha_r$, not all equal to zero, such that

$$\sum_{i=1}^{r} \alpha_i f'_i(x^*) = 0$$

and

$$\sum_{i=1}^{r} \alpha_i = 0.$$

Clearly, there is at least one α_i which is negative. Therefore, we may define a positive number β by

$$\beta = \min\{-v_i^*/\alpha_i : \alpha_i < 0, 1 \leqq i \leqq r\}.$$

We may also in turn define an m-vector $\bar{v}$ by

$$\bar{v}_i = \begin{cases} v_i^* + \beta\alpha_i & i \in J(v^*), \\ 0 & i \notin J(v^*). \end{cases}$$

From the choice of β we have $\bar{v} \geqq 0$. We also have

$$\begin{aligned} \sum_{i=1}^{m} \bar{v}_i f_i'(x^*) &= \sum_{i=1}^{m} v_i^* f_i'(x^*) + \beta \sum_{i=1}^{r} \alpha_i f_i'(x^*) \\ &= 0. \end{aligned}$$

and

$$\sum_{i=1}^{m} \bar{v}_i = \sum_{i=1}^{m} v_i^* + \beta \sum_{i=1}^{r} \alpha_i = 1.$$

Thus, the vector $\bar{v}$ is a Lagrange multiplier of the stationary point x^*.

Let j be an index such that $\beta = -v_j^*/\alpha_j$, then $j \notin J(\bar{v})$ but $j \in J(v^*)$. Therefore, $J(\bar{v})$ is a proper subset of $J(v^*)$. If $\bar{v}$ is also not regular then we may repeat this process until, after a finite number of steps, a regular Lagrange multiplier is obtained. The proof is then completed. ∎

For our analysis we also require that the stationary pair (x^*,v^*) satisfies the second order sufficiency condition of Theorem 2.5; that is, for any nonzero vector d in $\Omega(x^*,v^*)$ we have

$$d^T L''(x^*,v^*)d > 0,$$

where $\Omega(x^*,v^*)$ is defined in (2.5). We also need the following lemma.

Lemma 4.2 If (x^*,v^*) is a stationary pair then for any j in $J(v^*)$

$$\Omega(x^*,v^*) = \{d : (f_i'(x^*) - f_j'(x^*))^T d = 0, \forall i \in J(v^*)\backslash\{j\}\}.$$

Proof: It suffices to show that if for some real number η and for all $i \in J(v^*)$

$$f_i'(x^*)^T d = \eta,$$

then $\eta = 0$. This is obvious because

$$\begin{aligned} \eta &= \sum_{i=1}^{m} v_i^* \eta \\ &= \sum_{i=1}^{m} v_i^* f_i'(x^*))^T d \\ &= (\sum_{i=1}^{m} v_i^* f_i'(x^*))^T d = 0. \end{aligned}$$

■

Our main result is as follows.

Theorem 4.3 If a sequence of pairs $\{(x_k,v_k)\}$ generated from the method converges to a stationary pair (x^*,v^*) which satisfies the second order sufficiency condition of Theorem 2.5, and if the

Lagrange multiplier v^* is regular then the convergence of the sequence $\{x_k\}$ to x^* is R-superlinear.

Proof: Our proof is mainly based on Powell's proof in [9]. His proof is very long. Fortunately, we only need to show how his proof can apply to our case.

Note that, when k is sufficiently large, we have $J(v_k) = J(v^*)$. Without loss of generality we may assume that $J(v^*)$ contains the first r indices. We also have

$$\sum_{i=1}^{m} v_{k,i} = 1,$$

where $v_{k,i}$ is the i-th component of v_k. Therefore, the Lagrangian function $L(x_k,v_k)$ becomes

$$L(x_k,v_k) = f_r(x_k) + \sum_{i=1}^{r-1} v_{k,i}(f_i(x_k) - f_r(x_k)).$$

We may define

$$F(x) = f_r(x)$$

and

$$c_i(x) = f_r(x) - f_i(x) \quad i = 1,\dots,r-1,$$

and also may define for each k an $n \times (r-1)$ matrix N_k by

$$N_k = [c_1'(x_k), c_2'(x_k), \dots, c_{r-1}'(x_k)].$$

Because v^* is regular, the matrix N_k has full column rank for sufficiently large k. Hence, by Lemma 4.2, the projection matrix P_k on the subspace $\Omega(x_k,v_k)$ is given by

(4.1) $$P_k = I - N_k(N_k^T N_k)^{-1} N_k^T.$$

Lemma 4.2 and the second order sufficiency condition in Theorem 2.5 also imply that there exists an $\eta > 0$ such that

$$d^T L''(x_k, v_k) d \geq \eta ||d||^2$$

for all vectors d that satisfy the condition

$$N_k^T d = 0.$$

Therefore, all the conditions that are required in Powell's proof are satisfied. Then the superlinear convergence result follows. ∎

Another important result of Powell can also apply to our case similarly. This result is concerning a general method in which the matrices $\{B_k\}$ may not necessarily be updated from Powell's scheme. We state this result below.

<u>Theorem 4.4</u> Let the assumption of Theorem 4.3 holds. If the sequence $\{(x_k, v_k)\}$ are generated from solving iteratively the quadratic program (3.1) or (3.3) and the matrices $\{B_k\}$ satisfy the condition

$$\lim_{k\to\infty} \frac{||P_k(B_k - L''(x_k, v_k))P_k(x_{k+1} - x_k)||}{||x_{k+1} - x_k||} = 0$$

where P_k is the projection matrix defined in (4.1), then

$$\lim_{k\to\infty} \frac{||x_{k+1} - x^*||}{||x_{k-1} - x^*||} = 0.$$ ∎

The assumption in Theorem 4.3 that the sequence of multipliers

v_k converges to v^* may be relaxed if the strict complementarity condition holds.

Lemma 4.5 Let (x^*,v^*) be a stationary pair which satisfies the strict complementarity condition $I(x^*) = J(v^*)$ and let $\{(x_k,v_k)\}$ be a sequence generated from the method. If $x_k \to x^*$ and v^* is regular then $v_k \to v^*$.

Proof: We first observe that if $x_k \to x^*$ then $\delta_k \to \psi(x^*)$. From the quadratic program (3.1) we have

$$f_i(x_k) + f_i'(x_k)^T d_k \leq \delta_k, \quad i = 1,\ldots,m.$$

Therefore, it follows that for sufficiently large k and for any $i \notin I(x^*)$,

$$v_{k,i} = v_i^* = 0.$$

Thus, it suffices to show that $v_{k,I} \to v_I^*$ where $I = I(x^*)$. To do this we first note that

$$\begin{bmatrix} f_I'(x_k) \\ e^T \end{bmatrix} v_{k,I} = \begin{bmatrix} -B_k d_k \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} f_I'(x^*) \\ e^T \end{bmatrix} v_I^* = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Because $x_k \to x^*$, $d_k \to 0$ and the matrices $\{B_k\}$ are bounded, we have

$$\lim_{k\to\infty} \begin{bmatrix} f_I'(x^*) \\ e^T \end{bmatrix} (v_I^* - v_{k,I}) = 0. \tag{4.2}$$

Because v* is regular and the strict complementarity condition holds, the matrix in (4.2) has linearly independent columns. This implies $v_{k,I} \to v^*$ and, thus, the proof is completed. ■

Theorem 4.6 Let (x^*,v^*) be a stationary pair which satisfies the second order sufficiency condition in Theorem 2.3; that is, $d^T L''(x^*,v^*)d > 0$ for any nonzero d in the tangent cone $T(x^*)$. Let $\{x_k\}$ be a sequence generated from the method. If the sequence $\{x_k\}$ converges to x* and v* is regular and if the strict complementarity condition holds, then $\{x_k\}$ converges to x* at an R-superlinear rate.

Proof: The theorem follows directly from Theorem 4.4, Lemma 4.3 and Lemma 2.4. ■

It is not clear at this stage of research if the assumption on the regularity of the multiplier v* can be removed from the above theorems. However, from Theorem 4.1 we know that any stationary point has at least one regular Lagrange multiplier. Therefore, it may be worthwhile incorporating into the method a procedure to select Lagrange multipliers such that they can converge to such a regular multiplier.

We would conclude this paper with the following remark. Nonlinear programming and minimax are problems of one kind. A method is effective for one of them may also be effective for the other. In this paper we have shown that the effective nonlinear programming variable metric methods are, indeed, applicable to many minimax problems.

References

1. Beale, E.M.L. (1954). "On minimizing a convex function subject to linear inequalities", J. Roy. Stat. Soc., Ser. B, Vol. 17, pp. 173-177.

2. Cheney, E.W. and Goldstein, A.A. (1959). "Newton's method for convex programming and Tchebycheff approximation", Numerische Mathematik, Vol. 1, pp. 253-268.

3. Demjanov, V.F. (1968). "Algorithms for some minimax problems", J. of Computer and System Sciences, Vol. 2, pp. 342-380.

4. Demjanov, V.F. and Malozemov, V.N. (1974). Introduction to Minimax, Halsted Press, New York.

5. Han, S.P. (1976). "Superlinearly convergent variable metric methods for general nonlinear programming", Mathematical Programming, Vol. 11, pp. 263-282.

6. Han, S.P. (1977a). "Dual variable metric methods for constrained optimization", SIAM J. Control & Optimization, Vol. 15, pp.

7. Han, S.P. (1977b). "Variable metric methods for minimizing a class of nondifferentiable functions", TR77-322, Computer Science Dept., Cornell University.

8. Powell, M.J.D. (1977a). "A fast algorithm for nonlinearly constrained optimization calculations", presented at the 1977 Dundee Conference on Numerical Analysis.

9. Powell, M.J.D. (1977b). "The convergence of variable metric methods for nonlinearly constrained optimization calculations", presented at Nonlinear Programming Symposium 3, Madison, Wisconsin.

10. Robinson, S.M. (1975). "Stability theory for systems of inequalities, Part I: Linear systems", SIAM J. Numer. Anal., Vol. 12, pp. 754-769.

11. Zangwill, W.I., (1967). "An algorithm for the Chebyshev problem--with an application to concave programming", Management Science, Vol. 14, pp. 58-78.