VARIABLE METRIC METHODS FOR MINIMIZING A CLASS OF NONDIFFERENTIABLE FUNCTIONS*

S. P. Han

TR77-322

Department of Computer Science
Cornell University
Ithaca, N.Y. 14853

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

VARIABLE METRIC METHODS FOR MINIMIZING A CLASS OF NONDIFFERENTIABLE FUNCTIONS*

S. P. Han

Department of Computer Science
Cornell University
Ithaca, NY 14853

Abstract

We develop a class of methods for minimizing a nondifferentiable function which is the maximum of a finite number of smooth functions. The methods proceed by solving iteratively quadratic programming problems to generate search directions. For efficiency the matrices in the quadratic programming problems are suggested to be updated in a variable metric way. By doing so, the methods possess many attractive features of variable metric methods and can be viewed as their natural extension to the nondifferentiable case. To avoid the difficulties of an exact line search, a practical stepsize procedure is also introduced. Under mild assumptions the resulting method converge globally.

---

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

1. Introduction

We are concerned with the problem of minimizing a nondifferentiable function γ of the following form

$$\gamma(x) = \max_{i=1,\ldots,m} \{f_i(x)\}, \tag{1.1}$$

where f_i $(i=1,\ldots,m)$ are real-valued functions defined on the n-dimensional Euclidean space R^n and have continuous second order derivatives. The problem is also known as a minimax problem in the literature and abounds with applications.

Variable metric method, which are also usually called quasi-Newton methods, are effective for minimizing smooth functions. For minimizing the nondifferentiable function γ we present a class of methods which possess many attractive features of variable metric methods and can be viewed as their natural extention to the nondifferentiable case. Like many other methods for minimizing nondifferentiable functions, the method proceeds by solving quadratic programming problems iteratively to generate search directions. The matrices in the quadratic programming problems are preferrably updated according to the rule used in a variable metric method. A practical stepsize procedure is also introduced to avoid the difficulties arising from an exact line-search. With this stepsize procedure the method can be shown convergent globally under reasonable assumptions. Specific updating schemes for the matrices and rates of convergence have also been analyzed but they will be published somewhere else in order not to make the paper unduly long.

The symbol $||\cdot||$ denotes the ℓ_2 norm. All vectors are column vectors and a row vector is denoted by the superscript T. However, for convenience a column vector in R^{n+m} is sometimes denoted by (x,v) even though x and v are also column vectors in R^n and R^m respectively.

2. The Method

The problem to be considered is usually stated in the form

$$\min_{x\in R^n} \max_{i=1,\ldots,m} \{f_i(x)\}. \tag{2.1}$$

For any point x in R^n there is a corresponding index set $I(x) = \{i : f_i(x) = \gamma(x)\}$ where γ is defined as in (1.1). It is well known [see 2, for instance] that a necessary condition for a point, x^* say, to be a solution of Problem (2.1) is that the convex hull $\mathrm{conv}(x^*)$ of all the gradients $f_i'(x^*)$ with $i \in I(x^*)$ contains the null vector; that is,

$$0 \in \mathrm{conv}(x^*). \tag{2.2}$$

When γ is convex this condition is also sufficient. If we consider the minimization of a differentiable function as a special case of Problem (2.1) with $m = 1$, then condition (2.2) is just that the gradient of the objective function vanishes at a stationary point.

Condition (2.2) is equivalent to the statement that there exists an m-vector v^* such that

$$\text{(2.3)}\quad \text{(a)}\quad \sum_{i=1}^{m} v_i^* f_i'(x^*) = 0,$$

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

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

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

This condition is very similar to the Kuhn-Tucker condition of a nonlinear programming problem. For this reason we may define a Lagrangian function L(x,v) for Problem (2.1) by

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

and may call the m-vector v* a Lagrange multiplier associated with the stationary point x*.

The analogy between condition (2.3) and a Kuhn-Tucker condition is by no means a coincidence. When we put Problem (2.1) into the following equivalent nonlinear programming form,

(2.4)
$$\begin{aligned} &\underset{(x,\delta)\in R^{n+1}}{\text{Min}} && \delta \\ &\text{s.t.} && f_i(x) \leq \delta \quad i = 1,\ldots,m, \end{aligned}$$

then condition (2.3) is just the Kuhn-Tucker condition of this problem. Actually, the proposed method is essentially the recently developed variable metric method for nonlinear programming [4,5,11] applied to Problem (2.4) with its special structure taken into account.

We now describe the method as follows. Having an estimate x^k of a solution to Problem (2.1) and a positive definite symmetric matrix B_k at the k-th iteration, we solve the following quadratic programming problem

$$(2.5) \qquad \min_{(p,\delta)\in R^{n+1}} \ \tfrac{1}{2}p^T B_k p + \delta$$

$$\text{s.t.} \quad f_i(x^k) + f_i'(x^k)p \le \delta \qquad i = 1,\ldots,m.$$

From a solution (p^k,δ) of this quadratic programming problem we take the n-vector p^k as a search direction for x^k. Instead of (2.5) we may also equivalently solve its dual problem

$$(2.6) \qquad \min_{v\in R^m} \ \tfrac{1}{2}v^T J_k H_k J_k^T v - v^T f(x^k)$$

$$\text{s.t.} \quad \sum_{i=1}^{m} v_i = 1,$$

$$v \ge 0,$$

where $H_k = B_k^{-1}$ and J_k is the m x n Jacobian matrix of the vector function $f(x) = (f_1(x),\ldots,f_m(x))^T$ at the point x^k. If an m-vector v^k is a solution to Problem (2.6), then the vector p^k can be rediscovered by the equality

$$p^k = -H_k\Big(\sum_{i=1}^{m} v_i^k f_i(x^k)\Big).$$

A property of Problem (2.1) which is not shared by a general nonlinear programming problem is that there is a natural function γ which can be used to assess improvements and determine stepsizes. To accomplish these in the general nonlinear programming case we usually resort to a penalty function or an augmented Lagrangian [6,7,11].

Once the search direction p^k is obtained we determine a stepsize α_k and set $x^{k+1} = x^k + \alpha_k p^k$. The stepsize α_k is chosen as the first number in a generated sequence $\{\beta_0^{(k)},\beta_1^{(k)},\ldots\}$ satisfying

$$\gamma(x^k + \beta p^k) \leq \gamma(x^k) + \omega\beta\lambda_k \tag{2.7}$$

where $\lambda_k = -(p^k)^T B_k p^k$ and $0 < \omega < \frac{1}{2}$. We now describe how to generate the sequence $\{\beta_0^{(k)}, \beta_1^{(k)}, \ldots\}$. For convenience we omit the superscript k. We try the value one first and set $\beta_0 = 1$ because this value works well in the smooth case [see 3, for instance]. If β_j does not satisfy (2.7) then construct a quadratic function $\theta_j(\beta)$ which interpolates the function $\theta(\beta) = \gamma(x^k + \beta p^k)$ at $\beta = 0$ and $\beta = \beta_j$, and also satisfies the end condition

$$\theta_j'(0) = \lambda_k.$$

We let β_{j+1} be the greater of $0.1\beta_j$ and the value of β that minimizes $\theta_j(\beta)$. By a direct calculation we have

$$\beta_{j+1} = \max\{0.1\beta_j,\ \lambda_k\beta_j^2/2(\lambda_k\beta_j + \theta(0) - \theta(\beta_j))\}. \tag{2.8}$$

Stepsize procedures of this type are used frequently in the smooth case. Our procedure is also very similar to the one suggested by Powell in his work on a general nonlinear programming method [11]. One thing which is quite uncommon is that we set $\lambda_k = -(p^k)^T$ rather than the usual way $\lambda_k = \theta'(0)$. Because of the special structure of Problem (2.1), our choice is an appropriate one and will be justified in the next section.

The simplest way to update the matrix B_k in (2.5) is to use the identity matrix. But, for efficiency it is preferrable to use B_k that estimates the Hessian $\nabla_{xx}L(x^k, v^{k-1})$. Meanwhile, it is also very important for the matrices $\{B_k\}$ to be uniformly positive definite and bounded; that is, for some

positive number ρ and η,

$$\eta x^T x \leq x^T B_k x \leq \rho x^T x \tag{2.9}$$

for all x in R^n and for each k. Once condition (2.9) is maintained, specific updating formulas for B_k are not essential for our analysis of global convergence. We concentrate in this paper on the global convergence analysis of the method and will consider only the general case that the matrices $\{B_k\}$ satisfy condition (2.9). A detail discussion of specific updating schemes and rates of convergence will appear somewhere else [8].

For a better understanding of the method we compare it with a steepest descent method of Dem'yanov [2]. We recall that a directional derivative $h'(x;p)$ of a real-valued function h at a point x in a direction p is the quantity defined by

$$h'(x;p) = \lim_{t \to 0^+} \frac{h(x + tp) - h(x)}{t}.$$

It can be shown [1,2] that, though the function γ is nondifferentiable, its directional derivative $\gamma'(x;p)$ exists at any point and in any direction p and can be computed by

$$\gamma'(x;p) = \max_{i \in I} \{f_i'(x)p\}.$$

We may call a nonzero vector s a direction of steepest descent for the function γ at a point x if

$$\gamma'(x; \frac{s}{||s||}) \leq \min_{||p||=1} \gamma'(x;p).$$

It is also known that the negative of the smallest vector, measured by the ℓ_2 norm, in the convex hull conv(x) is a direction of steepest descent. Therefore, a practical way to get such a direction is to set $s = -\bar{p}$ with the vector $\bar{p}$ solving the constrained problem

$$\min_{p \in R^n} \; ||p||^2$$

$$\text{s.t.} \quad p \in \text{conv}(x),$$

or equivalently to set $s = -\sum_{i=1}^{m} \bar{v}_i f_i'(x)$ with $\bar{v}$ solving the quadratic programming problem.

$$\min_{v \in R^n} \; \frac{1}{2} v^T J(x) J(x)^T v \tag{2.10}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} v_i = 1,$$

$$v \geq 0,$$

$$v_i = 0 \quad \text{for} \quad i \notin I(x),$$

where J(x) is the Jacobian matrix of f at x. Clearly, (2.6) is closely related to (2.10). A difference between (2.6) and (2.10) is the appearance of a positive definite and symmetric matrix B_k in (2.6). This is because that our method uses a metric which may change iteratively and may not be the one induced from the ℓ_2 norm. Thus, the method is characterized by a property of a variable metric method. Another important difference is the way selecting gradients $f_i'(x)$ to form a search direction. By setting $v_i = 0$ for all $i \notin I(x)$ the steepest

descent method takes only those gradients with $i \in I(x)$. The drawback of not considering other gradients is so severe that the generated points may jam to a wrong point. An example illustrating this phenomenon is given in [2, pp. 74]. A procedure to circumvent this difficulty is to replace the condition that $v_i = 0$ for all $i \notin I(x)$ in (2.10) by that $v_i = 0$ for all $i \in I(x;\varepsilon) = \{i : f_i(x) - \gamma(x) \leq \varepsilon\}$ for a suitably small positive number ε. But, besides that the convergence of this method is slow, to determine a suitable ε in each iterative is not a trivial work in practice. However, in our proposed method, the term $-v^T f(x^k)$ appearing in the objective function of the quadratic programming problem (2.6) seems to be an appropriate device for automatically selecting adequate gradients $f_i(x^k)$ to form a search direction for x^k. Indeed, it will be shown that the direction p^k generated by solving (2.5) or (2.6) not only can avoid a jam but also allows the more practical inexact line search described before.

Some other related methods exist in the literature, notably Wolfe's method [12] and Lamarechal's method [9]. But they consider the minimization of a nondifferentiable function which is convex but not necessarily of form (1.1).

3. Convergence

In this section global convergence properties of the method will be analyzed. We will content ourselves with finding a stationary point of Problem (2.1), by which it is meant that a point, x^* say, satisfies condition (2.3) for some m-vector v^*. Note

that, at the k-th iteration of the method, we solve quadratic programming Problem (2.5) or (2.6) to find an (n + 1)-vector (p^k, δ_k) and an m-vector v^k satisfying the Kuhn-Tucker condition:

$$
\begin{aligned}
(3.1)\quad &(a)\quad B_k p^k + \sum_{i=1}^{m} v_i^k f_i'(x^k) = 0, \\
&(b)\quad \sum_{i=1}^{m} v_i^k = 1, \\
&(c)\quad v_i^k (f_i(x^k) + f_i'(x^k) p^k - \delta_k) = 0 \qquad i = 1,\ldots,m, \\
&(d)\quad f_i(x^k) + f_i'(x^k) p^k \leqq \delta_k \qquad i = 1,\ldots,m, \\
&(e)\quad v^k \geqq 0.
\end{aligned}
$$

Comparing condition (2.3) and condition (3.1) we immediately obtain the following result.

Theorem 3.1: If $p^k = 0$ then $\gamma(x^k) = \delta_k$ and the point x^k is a stationary point of Problem (2.1) with the vector v^k as its associated Lagrange multiplier. ∎

When $p^k \neq 0$, it is valid to use the vector p^k as a search direction if it is descent for γ at the point x^k. Therefore, it is desirable that the directional derivative $\gamma'(x^k; p^k)$ be negative. We establish this result below.

Lemma 3.2: $-(p^k)^T B_k p^k \geqq \delta_k - \gamma(x^k)$.

Proof: Let $\bar{I} = \{i : v_i^k > 0\}$. Then we have from (3.1.c) that for every $i \in \bar{I}$

$$f_i'(x^k) p^k = \delta_k - f_i(x^k).$$

Thus, it follows that

$$(3.2)\qquad \sum_{i=1}^{m} v_i^k f_i'(x^k)p^k = \sum_{i\in\bar{I}} v_i^k f_i'(x^k)p^k$$

$$= \sum_{i\in\bar{I}} v_i^k (\delta_k - f_i(x^k))$$

$$\geq \sum_{i\in\bar{I}} v_i^k (\delta_k - \gamma(x^k))$$

$$= (\delta_k - \gamma(x^k)) \sum_{i=1}^{m} v_i^k$$

$$= \delta_k - \gamma(x^k).$$

On the other hand, we have from (3.1.a) that

$$(p^k)^T B_k p^k = -\sum_{i=1}^{m} v_i^k f_i'(x^k)p^k,$$

which in conjunction with (3.2) implies the desired result. ■

Lemma 3.3: If $\beta \in (0,1]$, then

$$\frac{\gamma(x^k + \beta p^k) - \gamma(x^k)}{\beta} \leq -(p^k)^T B_k p^k + \frac{\beta}{2} M ||p^k||^2$$

where

$$M = \max_{\substack{\eta\in(0,\beta)\\ i=1,\ldots,m}} ||f_i''(x^k + \eta p^k)||.$$

Proof: Given a β in $(0,1]$, let j be any index in the set $I(x^k + \beta p^k)$. Then we have that for some $\eta \in (0,\beta)$

$$\gamma(x^k + \beta p^k) = f_j(x^k + \beta p^k)$$

$$= f_j(x^k) + \beta f_j'(x^k)p^k + \frac{\beta^2}{2}(p^k)^T f_j''(x^k + \eta p^k)\, p^k.$$

Meanwhile, we have from (3.1.d) that

$$\begin{aligned} f_j(x^k) + \beta f_j'(x^k)p^k &\leq f_j(x^k) + \beta(\delta_k - f_j(x^k)) \\ &= (1 - \beta)f_j(x^k) + \beta\delta_k \\ &\leq (1 - \beta)\gamma(x^k) + \beta\delta_k \\ &= \gamma(x^k) + \beta(\delta_k - \gamma(x^k)). \end{aligned}$$

Therefore, we have

$$\gamma(x^k + \beta p^k) \leq \gamma(x^k) + \beta(\delta_k - \gamma(x^k)) + \frac{\beta^2}{2} M||p^k||^2.$$

The lemma follows directly from the above inequality and Lemma 3.2. ■

Theorem 3.4: $\gamma'(x^k;p^k) \leq -(p^k)^T B_k p^k.$

Proof: We first note that directional derivatives of the function γ exist everywhere and in any direction. Hence the theorem follows from Lemma 3.3 immediately. ■

The above theorem is an important one because it shows that the vector p^k is qualified to be a search direction at the point x^k as long as the matrix B_k is positive definite. We can further deduce from the theorem the following result easily.

Corollary 3.5: If the point x^k is a local solution to Problem (2.1) and the matrix B_k is positive definite, then $p^k = 0$. ■

We now, in turn, give a justification for our stepsize procedure.

Theorem 3.6: If $p^k \neq 0$ and the matrix B_k is positive definite, then the stepsize α_k has a positive value and $\alpha_k = \beta_j^{(k)}$ for some finite j.

Proof: For simplicity we omit the superscript k from the sequence $\{\beta_0^{(k)}, \beta_1^{(k)}, \ldots\}$. We first observe that every element of this sequence is positive. This is because that $\beta_0 = 1$ and $\beta_{j+1} \geq 0.1\beta_j$ for all j. Therefore, it suffices to show that there exists some j such that β_j satisfies inequality (2.7). Suppose that it were not the case. Then we have that for every j

$$\theta(\beta_j) > \theta(0) + \omega\beta_j\lambda_k$$

which in turn implies that

$$\beta_{j+1} \leq \max\{0.1, 1/2(1 - \omega)\}\beta_j.$$

Hence it follows from $\omega < \frac{1}{2}$ that

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

But, from Lemma 3.3 and the assumption that the matrix B_k is positive definite, we have that there exists $\bar{\beta} > 0$ such that inequality (2.7) holds for all $\beta \in [0,\bar{\beta}]$. This is a contradiction and hence there must be some β_j satisfying inequality (2.7). The theorem is then completed. ■

From theorem 3.1 and Corollary 3.5 it seems very reasonable to expect that a point x^k with a small p^k should satisfy the necessary condition (2.3) of a solution approximately. This expectation is confirmed in the following theorem.

Theorem 3.7: If the gradients $f_i'(x)$ $(i = 1,\ldots,m)$ are bounded on the set $\Omega = \{x: \gamma(x) \leq \gamma(x^0)\}$ the matrices $\{B_k\}$ are also bounded, then for any $\varepsilon > 0$ there exists a $\delta > 0$ such that if $||p^k|| \leq \delta$ then

$$(3.3) \qquad \text{(a)} \quad \Big|\Big| \sum_{i=1}^{m} v_i^k f_i'(x^k) \Big|\Big| \leq \varepsilon,$$

$$\text{(b)} \quad 0 \leq v_i^k(\gamma(x^k) - f_i(x^k)) \leq \varepsilon,$$

$$\text{(c)} \quad \sum_{i=1}^{m} v_i^k = 1,$$

$$\text{(d)} \quad v^k \geq 0.$$

Proof: It is obvious that conditions (3.3.c) and (3.3.d) hold. To show that (3.3.a) and (3.3.b) also hold, we first let

$$\max_{\substack{i=1,\ldots,m \\ x\in\Omega}} ||f_i'(x)|| \leq M_1$$

and let

$$||B_k|| \leq M_2 \qquad \text{for each } k.$$

Assume that an $\varepsilon > 0$ is given. Then we set

$$\delta = \min\{\varepsilon/M_2,\ \varepsilon/2M_2\}.$$

If $||p^k|| \leq \delta$ for some k then we have from (3.1.a) that

$$\Big|\Big| \sum_{i=1}^{m} v_i^k f_i'(x^k) \Big|\Big| \leq ||B_k p^k|| \leq M_2||p^k|| \leq \varepsilon.$$

Thus (3.3.a) holds. Meanwhile, condition (3.3.b) holds trivially

for the index i with $v_i^k = 0$. Now we consider the index i with $v_i^k > 0$. Let j be in the set $I(x^k)$. Then we have from (3.1.c) and (3.1.d) that

$$\begin{aligned} 0 \leq v_i^k(\gamma(x^k) - f_i(x^k)) &= v_i^k(\gamma(x^k) - \delta_k + f_i'(x^k)p^k) \\ &\leq \gamma(x^k) - \delta_k + f_i'(x^k)p^k \\ &= f_j(x^k) - \delta_k + f_i'(x^k)p^k \\ &\leq -f_j'(x^k)p^k + f_i'(x^k)p^k \\ &\leq 2M_1||p^k|| \leq \varepsilon. \end{aligned}$$

Hence the proof is completed. ■

Because of Theorem 3.7 it is meaningful for the method to produce a point x^k whose corresponding quantity $||p^k||$ is smaller than any given small positive tolerance. We give this result in the following theorem.

Theorem 3.8: If the function γ is bounded below and the Hessians $f_i''(x)$ $(i = 1,\ldots,m)$ are also bounded, and if there exists a $\eta > 0$ such that $x^T B_k x \geq \eta x^T x$ for all x in R^n and for every k, then either the sequence $\{x^k\}$ generated by the method terminates at a stationary point of Problem (2.1) or

$$\liminf_{k\to\infty} ||p^k|| = 0.$$

Proof: Suppose that the conclusion of the theorem were false. Then there exists an $\varepsilon > 0$ such that $||p^k|| \geq \varepsilon$ for every k. By the choice of the stepsize α_k we have that

$$\begin{aligned}\gamma(x^k) - \gamma(x^{k+1}) &\geq -\omega\lambda_k\alpha_k \\ &= \omega\alpha_k p^{k^T} B_k p^k \\ &\geq \omega\eta||p^k||^2.\end{aligned}$$

Taking the sums of both sides of the above inequality and taking into account the assumption that the function γ is bounded below, we obtain that

$$\infty > \sum_{k=0}^{\infty} \omega\eta\alpha_k||p^k||^2.$$

Since $||p^k|| \geq \varepsilon > 0$ for every k, we get

$$\lim_{k\to\infty} \alpha_k = 0. \tag{3.4}$$

On the other hand, Theorem 3.6 indicates that for each k there exists a $\beta_j^{(k)}$ in the sequence $\{\beta_0^{(k)}, \beta_1^{(k)}, \ldots\}$ such that $\alpha_k = \beta_j^{(k)}$. Because of (3.4) and $\beta_0^{(k)} = 1$, the integer j is not zero for sufficiently large k and we can define the number $\bar{\alpha}_k$ by

$$\bar{\alpha}_k = \beta_{j-1}^{(k)}.$$

Since $\beta_j^{(k)}$ is the first number in the sequence $\{\beta_0^{(k)}, \beta_1^{(k)}, \ldots\}$ satisfying inequality (2.7), we then have

$$\frac{\gamma(x^k + \bar{\alpha}_k p^k) - \gamma(x^k)}{\bar{\alpha}_k} > \omega\lambda_k = -\omega(p^k)^T B_k p^k. \tag{3.5}$$

Let

$$\max_{\substack{i=1,\ldots,m \\ x\in R^n}} ||f_i''(x)|| \leq M \quad \text{for some } M > 0.$$

It follows from Lemma 3.3 that if $0 < \alpha \leq \eta/M$, then

$$\begin{aligned}\frac{\gamma(x^k + \alpha p^k) - \gamma(x^k)}{\alpha} &\leq -(p^k)^T B_k p^k + \frac{\alpha M}{2}||p^k||^2 \\ &\leq -(p^k)^T B_k p^k + \frac{\eta}{2}||p^k||^2 \\ &\leq -(p^k)^T B_k p^k + \frac{1}{2}(p^k)^T B_k p^k \\ &= -\frac{1}{2}(p^k)^T B_k p^k.\end{aligned}$$

Because of $0.1\bar{\alpha}_k \leq \alpha_k$ and (3.4), it holds for sufficiently large k that $0 < \bar{\alpha}_k \leq \eta/M$ and

$$\frac{\gamma(x^k + \bar{\alpha}_k p^k) - \gamma(x^k)}{\bar{\alpha}_k} \leq -\frac{1}{2}(p^k)^T B_k p^k.$$

The above inequality contradicts (3.5) because $\omega < \frac{1}{2}$. Hence we have deduced a contradiction and the proof is completed. ■

From Theorems 3.7 and 3.8 we immediately have the following corollary.

Corollary 3.9: Let the assumptions of Theorems 3.7 and 3.8 hold. Then for any given positive tolerance ε the method will produce an n-vector x^k and an m-vector v^k satisfying condition 3.3. ■

A stronger result can be obtained if we assume that the generated points $\{x^k\}$ eventually move into a region where the function γ is convex.

Theorem 3.10: If for some i the level set $\Omega_i = \{x : \gamma(x) \leq \gamma(x^i)\}$ is bounded and γ is convex on Ω_i, and also if there exist posi-

tive members ρ and η such that $\eta x^T x \leq x^T B_k x \leq \rho x^T x$ for all x in R^n and for every $k \geq i$, then the method calculates a sequence of function values $\gamma(x^k)$ $(k = 0,1,2,\ldots)$, that either converges to or terminates at the least value of $\gamma(x)$.

Proof: Clearly the set Ω_i is compact. Hence all the assumptions of Theorems 3.7 and 3.8 hold. Note also that the sequence $\{\gamma(x^k)\}$ is monotonically decreasing. Thus, it suffices to show that it has a subsequence that converges to the least value of the function γ.

By the compactness of the set Ω_i and Theorem 3.8 we have that there exists a subsequence of triples $\{(x^{k_j}, v^{k_j}, p^{k_j})\}$ such that for some $\bar{x}$ in Ω_i and some $\bar{v}$ in R^m we have $x^{k_j} \to \bar{x}$, $v^{k_j} \to \bar{v}$ and $p^{k_j} \to 0$. It follows from Theorem 3.7 that the pair $(\bar{x},\bar{v})$ satisfies condition (2.3) and hence $\bar{x}$ is a stationary point of Problem (2.1) with $\bar{v}$ as its Lagrange multiplier. Moreover, the point $\bar{x}$ is actually a solution to Problem (2.1) because the function γ is convex in Ω_i. Therefore, the subsequence $\{\gamma(x^{k_j})\}$ converges to $\gamma(\bar{x})$ that is the least value of γ. Hence we complete the proof. ∎

We conclude this section with a corollary which is concerning the convergence of the generated points $\{x^k\}$ rather than function values. The proof is omitted because it is straightforward and very similar to the one given in [10, Theorem 14.1.4].

Corollary 3.11: Let the assumptions of Theorem 3.10 hold and

let the solution set S be defined by

$$S = \{x : \gamma(x) = \min_{y \in R^n} \gamma(y)\}.$$

Then

$$\lim_{k \to \infty} [\inf_{y \in S} ||x^k - y||] = 0.$$

In particular, if the set S consists of a single point x^*, then $\lim_{k \to \infty} x^k = x^*$. ■

4. Conclusions

The paper is an attempt to extend the efficient variable metric methods to a class of nondifferentiable minimization problems. Our work is mainly motivated by a natural connection of the problem to a general nonlinear programming problem. This approach seems advantageous. Firstly, we can gain much more insights into the problem through our knowledge and techniques of nonlinear programming. Secondly, the difference between the nondifferentiable minimization problem with constraints and without constraints becomes nonexistent and they can be treated in an unified way. More specifically, to solve the constrained problem

$$\min_{x \in R^n} \gamma(x)$$

$$\text{s.t. } g_j(x) \leq 0 \qquad j = 1, \ldots, 9,$$

our method becomes to solve iteratively, instead of (2.5), the following subproblem

$$\min_{(p,\delta)\in R^{n+1}} \quad \tfrac{1}{2}p^T B_k p + \delta$$

$$f_i(x^k) + f_i'(x^k)p \leqq \delta \qquad i = 1,\ldots,m,$$

$$g_j(x^k) + g_j'(x^k)p \leqq 0 \qquad j = 1,\ldots,9.$$

The theory behind this extension can be readily developed through the results of this paper and the results in the literature of nonlinear programming.

The methods have not yet been seriously implemented. But they are expected to be very promising because not only that they have strong theoretical support but also that a similar approach has achieved a great success in nonlinear programming [4, 5, 6, 11].

References

[1] J.M. Danskin, "The theory of max-min, with applications", SIAM J. Appl. Math. 14(1966) pp. 641-664.

[2] V.F. Dem'yanov and V.N. Malozemov, Introduction to minimax (John Wiley & Sons, New York, 1974).

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

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

[5] S.P. Han, "Dual variable metric methods for constrained optimization", SIAM J. Control & Optimization, 15 (1977).

[6] S.P. Han, "A global convergent method for nonlinear programming", J. Optimization Theory Appl., 22 (1977) pp. 297-

309.

[7] S.P. Han, "A hybrid method for nonlinear programming", presented at Nonlinear Programming Symposium 3, Madison, Wisconsin, (1977).

[8] S.P. Han, "A superlinearly convergent method for a class of nondifferentiable minimization problems", (in preparation).

[9] C. Lamarechal, "An extension of Davidon methods to non-differentiable problems", Mathematical Programming Study 3 (1975) pp. 95-109.

[10] J.M. Ortega and W.C. Rheinboldt, Iterative solution of nonlinear equations in several variables (Academic Press, New York, 1970).

[11] M.J.D. Powell, "A fast algorithm for nonlinear constrained optimization calculations", presented at the 1977 Dundee Conference on Numerical Analysis.

[12] P. Wolfe, "A method of conjugate subgradients for minimizing nondifferentiable functions", Mathematical Programming Study 3 (1975) pp. 145-173.