An Outer Approximation Algorithm for Solving General Convex Programs

MASAO FUKUSHIMA
Kyoto University, Kyoto, Japan

(Received August 1980; accepted February 1982)

A new outer approximation algorithm is proposed for solving general convex programs. A remarkable advantage of the algorithm over existing outer approximation methods is that the approximation of the constraint set is not cumulative. That is, the algorithm solves at each iteration a quadratic program whose constraints depend only on the current estimate of an optimal solution. Convergence of the algorithm is proved and possible applications are discussed.

RECENTLY outer approximation methods have drawn much attention in the context of nonsmooth optimization. The first outer approximation method was the cutting plane method independently introduced by Kelley [1960], and Cheney and Goldstein [1959]. Although those methods can solve nonsmooth convex programs, they possess the drawback that the number of constraints in the subproblem rapidly grows as the iteration proceeds. To overcome this difficulty, Eaves and Zangwill [1971] have proposed several schemes for dropping old constraints systematically instead of retaining all constraints previously generated (see also Topkis [1970a, b]. Furthermore, they have suggested that such constraint dropping strategies might be applied to other outer approximation methods (e.g., Veinott [1967], Dantzig [1963]). Since then, a variety of outer approximation methods incorporating constraint dropping strategies have appeared, e.g., Elzinga and Moore [1975], Gonzaga and Polak [1979], and Hogan [1973a].

In this paper, we propose a new outer approximation algorithm for solving general convex programs. The advantage of the algorithm is that the approximation of the constraint set is not cumulative. That is, the algorithm solves at each iteration a quadratic program whose constraints depend only on the current estimate of an optimal solution. In other words, it does not require any information from the previous iterations, while other outer approximation algorithms must retain old information at least until judged to be useless by some constraint dropping criterion.

Subject classification: 642, 644 outer approximation algorithm.

Operations Research
Vol. 31, No. 1, January–February 1983

0030-364X/83/3101-0101 $01.25

The subproblem to be solved at each iteration contains as many inequality constraints as the original problem. Furthermore, it is often the case in various interesting applications that the subproblem involves just a single inequality constraint. An example of such applications is the dual approach to mathematical programs, which will be discussed in Section 4.

1. PROBLEM AND ASSUMPTIONS

We shall consider the following convex program:

$$(P \qquad \text{Maximize} \quad cx$$
$$\text{subject to} \quad g_i(x) \leqq 0, \qquad i = 1, \cdots, m, \qquad x \in X,$$

where c and x are n-vectors, $g_i : R^n \to R$ is a convex function for each $i = 1, \cdots, m$, and X is a nonempty convex compact set in R^n.

As is well known, the linearity assumption on the objective function does not affect the generality of the problem, because any convex program

$$\text{Maximize} \quad \theta(x)$$
$$\text{subject to} \quad g_i(x) \leqq 0, \qquad i = 1, \cdots, m, \qquad x \in X,$$

where $\theta : R^n \to R$ is concave and g_i and X are as in (P), can be rewritten as the following convex program with a linear objective function by introducing the variable x_{n+1}:

$$\text{Maximize} \quad x_{n+1}$$
$$\text{subject to} \quad g_i(x) \leqq 0, \qquad i = 1, \cdots, m, \qquad x_{n+1} - \theta(x) \leqq 0,$$
$$x \in X, \qquad x_{n+1} \in [f_1, f_2],$$

where $f_1 = \min\{\theta(x) \mid x \in X\}$ and $f_2 = \max\{\theta(x) \mid x \in X\}$.

Theoretically, the set X can be any convex compact subset of R^n. From a computational viewpoint, however, X should preferably possess a simple structure such as simple upper and lower bounds on the variables.

Throughout this paper, we assume that problem (P) satisfies Slater's condition, namely, there exists a point $x \in X$ such that

$$g_i(x) < 0, \qquad i = 1, \cdots, m.$$

Clearly, problem (P) has a finite optimal solution since the feasible region of (P) is nonempty and compact.

We also assume that at every point $x \in X$, we can calculate at least one subgradient of each function g_i, $i = 1, \cdots, m$. (A vector $\gamma \in R^n$ is called a subgradient of g_i at x if it satisfies the inequality

$$g_i(x') \geqq g_i(x) + \gamma(x' - x), \qquad \forall x' \in R^n \tag{1}$$

and the set of all subgradients of g_i at x is denoted by $\partial g_i(x)$.) Various properties of subgradients may be found in Rockafellar [1970].

Remark. Under our assumptions on g_i, problem (P) could be transformed into an equivalent convex program with a single inequality constraint, i.e.,

$$(P') \qquad \text{Maximize} \quad cx$$
$$\text{subject to} \quad g_0(x) \leqq 0, \qquad x \in X,$$

where $g_0 : R^n \to R$ is a convex function defined by

$$g_0(x) = \max_{1 \leqq i \leqq m} g_i(x).$$

In fact, many important classes of problems can be formulated in the form of (P') (cf. Section 4). However, we consider the problem in the more general form of (P).

2. ALGORITHM

Let $\{t_k\}$ be a sequence of positive numbers such that

$$\lim_{k\to\infty} t_k = 0, \qquad \sum_{k=1}^{\infty} t_k = \infty, \tag{2}$$

for example, $t_k = a/k$ where a is a positive constant. Then the algorithm can be stated as follows:

Step 1. Choose an initial point $x^1 \in X$ and arbitrary

$$\gamma_i^1 \in \partial g_i(x^1), \qquad i = 1, \cdots, m. \qquad \text{Let } k = 1.$$

Step 2. Solve the subproblem

$$(SP_k) \qquad \text{Maximize} \quad t_k cx - (1/2)\|x - x^k\|^2$$
$$\text{subject to} \quad g_i(x^k) + \gamma_i^k(x - x^k) \leqq 0, \qquad i = 1, \cdots, m,$$
$$x \in X,$$

where $\|\cdot\|$ denotes the Euclidean norm in R^n. Let the optimal solution of (SP^k) be x^{k+1}.

Step 3. If $x^{k+1} = x^k$, then stop. Otherwise, choose arbitrary $\gamma_i^{k+1} \in \partial g_i(x^{k+1})$, increase k by one and go to Step 2.

Let S and T_k denote the feasible regions of problems (P) and (SP_k), respectively, i.e.,

$$S = \{x \in X;\ g_i(x) \leqq 0,\ i = 1, \cdots, m\}, \tag{3}$$

$$T_k = \{x \in X;\ g_i(x^k) + \gamma_i^k(x - x^k) \leqq 0,\ i = 1, \cdots, m\}. \tag{4}$$

Then, for any choice of subgradients $\gamma_i^k \in \partial g_i(x^i)$, the inclusion $S \subset T_k$

$\subset X$ must hold, since by (1) $g_i(x) \leqq 0$ implies $g_i(x^k) + \gamma_i^k(x - x^k) \leqq 0$ for $i = 1, \cdots, m$. This fact together with the strong concavity of the function $t_k cx - (1/2)\|x - x^k\|^2$ and the convexity and the compactness of the set T_k assures the existence of a unique optimal solution in problem (SP_k).

Notice that, when X is a convex polyhedron, problem (SP_k) is a quadratic program and methods for solving it are readily available (Panne [1975]).

Remark. The above algorithm is by all means applicable to the equivalent problem (P') in place of problem (P). In this case, the subproblem to be solved at each iteration would have just a single inequality constraint

$$g_0(x^k) + \gamma^k(x - x^k) \leqq 0, \qquad x \in X \tag{5}$$

where $\gamma^k \in \partial g_0(x^k) = co\{\partial g_i(x^k);\ i \in I(x^k)\}$ with $I(x^k) = \{i;\ g_0(x^k) = g_i(x^k)\}$ (Clarke [1975], Theorem 2.1) and co stands for convex hull. This subproblem appears attractive because its constraint is extremely simple. However, keep in mind that this simplicity is gained because the constraint (5) might have discarded useful information on S. Specifically, (5) does not utilize any information from the functions g_i, $i \notin I(x^k)$.

3. CONVERGENCE

We first show that if the algorithm terminates at Step 3, then the current point is actually an optimal solution to problem (P).

LEMMA 1. *If x^k is optimal to problem (SP_k) for some k, i.e., $x^{k+1} = x^k$, then x^k is also optimal to problem (P).*

Proof. First observe that problem (SP_k) satisfies Slater's condition. Then, the necessary and sufficient condition for x to be optimal to (SP_k) is that there exist scalars λ_i, $i = 1, \cdots, m$, such that

$$\begin{cases} [t_k c - (x - x^k) - \sum_{i=1}^m \lambda_i \gamma_i^k](z - x) \leqq 0, & \forall z \in X \\ \sum_{i=1}^m \lambda_i [g_i(x^k) + \gamma_i^k(x - x^k)] = 0, & x \in X \\ \lambda_i \geqq 0, \quad g_i(x^k) + \gamma_i^k(x - x^k) \leqq 0, & i = 1, \cdots, m, \end{cases} \tag{6}$$

(see Mangasarian [1969]). If x^k is optimal to (SP_k), we may replace x by x^k in (6) to obtain

$$[t_k c - \textstyle\sum_{i=1}^m \lambda_i \gamma_i^k)(z - x^k) \leqq 0, \qquad \forall z \in X$$

$$\textstyle\sum_{i=1}^m \lambda_i g_i(x^k) = 0, \qquad x^k \in X$$

$$\lambda_i \geqq 0, \qquad g_i(x^k) \leqq 0, \qquad i = 1, \cdots, m,$$

which in turn imply that x^k is an optimal solution of the problem $\max\{t_k cx \mid x \in S\}$. Since $t_k > 0$, the lemma follows immediately.

Next, we consider the case in which the algorithm generates an infinite sequence $\{x^k\}$. Let the distance from a point x to the set S be defined by

$$\mathrm{dist}[x, S] = \min_{z\in S}\|x - z\|.$$

LEMMA 2. *Let T be a compact convex set such that $S \subset T \subset X$ and let y be the point in T closest to a point $x \in X$. Then we have*

$$\|y - z\|^2 \leqq \|x - z\|^2 - \|y - x\|^2, \qquad \forall z \in S. \tag{7}$$

Furthermore, it holds that

$$\mathrm{dist}[y, S]^2 \leqq \mathrm{dist}[x, S]^2 - \|y - x\|^2. \tag{8}$$

Proof. As y is the optimal solution of the problem $\min\{\|z - x\|^2 | z \in T\}$, y must satisfy the inequality (Luenberger [1969], p. 178)

$$(y - x)(z - y) \geqq 0, \qquad \forall z \in T. \tag{9}$$

But, since $\|x - z\|^2 - \|y - x\|^2 - \|y - z\|^2 = 2(y - x)(z - y)$ and $S \subset T$, it is easily seen that (9) implies (7). To prove (8), let z' be the closest point to x in S. Then, taking account of (7), we obtain

$$\mathrm{dist}[y, S]^2 \leqq \|y - z'\|^2 \leqq \|x - z'\|^2 - \|y - x\|^2 = \mathrm{dist}[x, S]^2 - \|y - x\|^2.$$

LEMMA 3. *For any k,* $\mathrm{dist}[x^{k+1}, S] \leqq \mathrm{dist}[x^k, S] + t_k\|c\|$.

Proof. First note that problem (SP_k) is equivalent to the problem $\min\{\|x - (x^k + t_kc)\|^2 | x \in T_k\}$, where T_k is the feasible region of (SP_k) defined by (4). In other words, x^{k+1} is the closest point to $x^k + t_kc$ in T_k. Since $T_k \supset S$, it then follows from Lemma 2 that

$$\begin{aligned}\mathrm{dist}[x^{k+1}, S] &\leqq (\mathrm{dist}[x^k + t_kc, S]^2 - \|x^{k+1} - (x^k + t_kc)\|^2)^{1/2} \\ &\leqq \mathrm{dist}[x^k + t_kc, S].\end{aligned}$$

On the other hand, if we let z' be the closest point to x^k in S, we have

$$\begin{aligned}\mathrm{dist}[x^k + t_kc, S] &\leqq \|x^k + t_kc - z'\| \leqq \|x^k - z'\| + t_k\|c\| \\ &= \mathrm{dist}[x^k, S] + t_k\|c\|.\end{aligned}$$

Combining above inequalities, we obtain the desired result.

It is convenient to introduce the following notation: Given a point $x \in X$, vectors $\gamma_i \in R^n$, $i = 1, \cdots, m$, and a scalar $t > 0$, the optimal solution of the problem

$$\max\{tcz - (1/2)\|z - x\|^2 | g_i(x) + \gamma_i(z - x) \leqq 0, i = 1, \cdots, m, z \in X\}$$

is denoted by $\tilde{x}(x, \Gamma, t)$, where $\Gamma = (\gamma_1, \cdots, \gamma_m) \in R^{mn}$. Using this notation, we may write the optimal solution of problem (SP_k) as $x^{k+1} = \tilde{x}(x^k, \Gamma_k, t_k)$, where $\Gamma_k = (\gamma_1{}^k, \cdots, \gamma_m{}^k)$.

LEMMA 4. *Let* $X \sim S = \{x \in X; x \notin S\}$. *Then there exists a scalar* $\bar{\alpha} \in [0, 1)$ *such that*

$$\operatorname{dist}[\tilde{x}(x, \Gamma, 0), S] \leqq \bar{\alpha} \operatorname{dist}[x, S]$$

for any $x \in X \sim S$ *and any* $\Gamma = (\gamma_1, \cdots, \gamma_m)$ *with* $\gamma_i \in \partial g_i(x)$, $i = 1, \cdots, m$.

Proof. Since $\tilde{x}(x, \Gamma, 0)$ is by definition the closest point to x in the compact convex set $T = \{z \in X; g_i(x) + \gamma_i(z - x) \leqq 0, i = 1, \cdots, m\}$ and since $T \supset S$, we can apply Lemma 2 to obtain the inequality

$$\operatorname{dist}[\tilde{x}(x, \Gamma, 0), S]^2 \leqq \operatorname{dist}[x, S]^2 - \|\tilde{x}(x, \Gamma, 0) - x\|^2.$$

Note that, if $x \in X \sim S$, then $x \notin T$, so that $x \neq \tilde{x}(x, \Gamma, 0)$. On the other hand, since each g_i is Lipschitz continuous on the compact set X (see Rockafellar, Theorem 10.4) and since Slater's condition is satisfied, it follows from an argument similar to Robinson's [1976] (see p. 242) that there exists a constant κ ($\geqq 1$) such that

$$\operatorname{dist}[x, S] \leqq \kappa \|\tilde{x}(x, \Gamma, 0) - x\|$$

for any $x \in X \sim S$ and any $\gamma_i \in \partial g_i(x)$, $i = 1, \cdots, m$. Combining above two inequalities and putting $\bar{\alpha} = \sqrt{\kappa^2 - 1}/\kappa$, we obtain the inequality $\operatorname{dist}[x(\tilde{x}, \Gamma, 0), S] \leqq \bar{\alpha} \operatorname{dist}[x, S]$.

Remark. For $x \in X \sim S$ and $\gamma_i \in \partial g_i(x)$, $i = 1, \cdots, m$, the point $\tilde{x}(x, \Gamma, 0)$ is exactly the same as the solution of the subproblem in Robinson's algorithm for solving the system of convex inequalities.

LEMMA 5. *Let* $\delta > 0$. *Then there exist* $\alpha \in [0, 1)$ *and* $\bar{t} > 0$ *such that* $t \in [0, \bar{t}]$ *implies*

$$\operatorname{dist}[\tilde{x}(x, \Gamma, t), S] \leqq \alpha \operatorname{dist}[x, S] \tag{10}$$

for any $x \in X$ *such that* $\operatorname{dist}[x, S] \geqq \delta/2$ *and any* $\Gamma = (\gamma_1, \cdots, \gamma_m)$ *with* $\gamma_i \in \partial g_i(x)$, $i = 1, \cdots, m$.

Proof. Let us define $V_\delta = \{x \in X; \operatorname{dist}[x, S] \geqq \delta/2\}$ and

$$\varphi_\delta(t) = \max\{(\operatorname{dist}[\tilde{x}(x, \Gamma, t), S])/\operatorname{dist}[x, S] \mid \gamma_i \in \partial g_i(x), \quad i = 1, \cdots, m, x \in V_\delta\}.$$

Since $\operatorname{dist}[\tilde{x}(x, \Gamma, t), S]/\operatorname{dist}[x, S]$ is continuous with respect to (x, Γ, t) and the set $\{(x, \Gamma); \gamma_i \in \partial g_i(x), i = 1, \cdots, m, x \in V_\delta\}$ is compact, the function φ_δ is continuous with respect to t (Hogan [1973b], Theorem 7). Moreover, as $V_\delta \subset X \sim S$, we have $\varphi_\delta(0) \leqq \bar{\alpha}$ for some $\bar{\alpha} \in [0, 1)$ by Lemma 4. Therefore, for any $\epsilon > 0$, there exists a $\bar{t} > 0$ such that $t \in [0, \bar{t}]$ implies $\varphi_\delta(t) \leqq \bar{\alpha} + \epsilon$. Since $\bar{\alpha} < 1$, by choosing ϵ so as to satisfy $\bar{\alpha} + \epsilon < 1$, we finally obtain the inequality (10) with $\alpha = \bar{\alpha} + \epsilon$.

With this preparation, we can show that the sequence generated by the algorithm approaches a feasible set of problem (P).

LEMMA 6. *Suppose that the algorithm generates an infinite sequence* $\{x^k\}$. *Then* $\mathrm{dist}[x^k, S]$ *approaches zero as* k *tends to infinity.*

Proof. Let us choose $\delta > 0$ arbitrarily. Then, by (2), we may define $k' = \min\{k;\ t_l \leqq \delta/(2\|c\|),\ \forall l \geqq k\}$ and $k'' = \min\{k;\ t_l \leqq \bar{t},\ \forall l \geqq k\}$, where $\bar{t}$ is a positive number as specified in Lemma 5. Let $k_0 = \max\{k', k''\}$. We consider the following two cases: (i) $\mathrm{dist}[x^{k_0}, S] \geqq \delta/2$ and (ii) $\mathrm{dist}[x^{k_0}, S] < \delta/2$.

In case (i), as $t_k \leqq \bar{t}$ for all $k \geqq k_0$, it follows from (10) that $\mathrm{dist}[x^{k_0}, S] > \mathrm{dist}[x^{k_0+1}, S] > \cdots > \mathrm{dist}[x^{k_1-1}, S] \geqq \delta/2 > \mathrm{dist}[x^{k_1}, S]$ holds for some $k_1 > k_0$. (Recall that $x^{k+1} = \tilde{x}(x^k, \Gamma_k, t_k)$.) Now, suppose that $\mathrm{dist}[x^k, S] > \delta$ for some $k > k_1$ and let k_2 be the least such integer, i.e., $\mathrm{dist}[x^l, S] \leqq \delta$ for all $k_1 \leqq l \leqq k_2 - 1$. Then it cannot be that $\mathrm{dist}[x^{k_2-1}, S] \geqq \delta/2$, because, if so, by (10) one would have $\mathrm{dist}[x^{k_2}, S] < \mathrm{dist}[x^{k_2-1}, S] \leqq \delta$ and this contradicts the supposition $\mathrm{dist}[x^{k_2}, S] > \delta$. Moreover, it is also impossible that $\mathrm{dist}[x^{k_2-1}, S] < \delta/2$, because from Lemma 3 and the definition of k', we must have $\delta < \mathrm{dist}[x^{k_2}, S] \leqq \mathrm{dist}[x^{k_2-1}, S] + t_k\|c\| \leqq \mathrm{dist}[x^{k_2-1}, S] + \delta/2$, i.e., $\mathrm{dist}[x^{k_2-1}, S] \geqq \delta/2$. Consequently, $\mathrm{dist}[x^k, S] \leqq \delta$ must hold for all $k \geqq k_1$. In case (ii), by applying the preceding arguments with k_1 replaced by k_0, we can show that $\mathrm{dist}[x^k, S] \leqq \delta$ for all $k \geqq k_0$. Therefore, in any case, there exists an integer $\bar{k}$ such that $\mathrm{dist}[x^k, S] \leqq \delta$ for all $k \geqq \bar{k}$. Since δ was arbitrary, this implies that $\mathrm{dist}[x^k, S]$ must converge to zero as k tends to infinity.

Now we can state the main convergence theorem for the algorithm.

THEOREM. *The algorithm either terminates at an optimal solution or generates an infinite sequence with the property that an accumulation point of the sequence is an optimal solution.*

Proof. If the algorithm terminates at the kth iteration, the point x^k must be an optimal solution by Lemma 1. Now, we suppose that an infinite sequence $\{x^k\}$ is generated. Let $f^* = \max\{cx \mid x \in S\}$. Then, the latter half of the theorem is proved if we show the following inequality holds, since any accumulation point of $\{x^k\}$ belongs to S by Lemma 6:

$$\limsup_{k\to\infty} cx^k \geqq f^*.$$

In order to obtain contradiction, suppose that there exist an $\epsilon > 0$ and an integer k_1 such that

$$cx^k \leqq f^* - \epsilon, \qquad \forall k \geqq k_1. \tag{11}$$

If we choose any $x^* \in S$ such that $cx^* = f^*$, we have

$$c(x^* - \rho c) \geqq f^* - \epsilon \tag{12}$$

for any fixed $\rho \in [0, \epsilon/\|c\|^2]$. Since x^{k+1} is the closest point to $x^k + t_k c$ in T_k and since $x^* \in S \subset T_k$, it follows from Lemma 2 that

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &\leqq \|x^k + t_k c - x^*\|^2 - \|x^k + t_k c - x^{k+1}\|^2 \\ &\leqq \|x^k + t_k c - x^*\|^2 \\ &= \|x^k - x^*\|^2 + t_k(t_k - 2\rho)\|c\|^2 + 2t_k c(x^k - x^* + \rho c).\end{aligned} \tag{13}$$

Let $k_2 = \min\{k \mid t_l \leqq \rho, \forall l \geqq k\}$ and $k_0 = \max\{k_1, k_2\}$. Then, by (11), (12) and (13), we obtain

$$\|x^{k+1} - x^*\|^2 \leqq \|x^k - x^*\|^2 - \rho t_k\|c\|^2, \qquad \forall k \geqq k_0. \tag{14}$$

Moreover, by adding the inequalities (14) from $k = k_0$ to $k_0 + l$, we have

$$\|x^{k_0+l} - x^*\|^2 \leqq \|x^{k_0} - x^*\|^2 - \rho\|c\|^2 \textstyle\sum_{k=k_0}^{k_0+l} t_k.$$

However, since $\sum_{k=1}^{\infty} t_k = \infty$ (cf. Equation 2), the right hand side of this inequality becomes arbitrarily small as l increases. This is impossible because the left hand side is nonnegative. This contradiction implies that (11) is not true and there must exist a subsequence $\{x^{k_i}\}$ such that $\lim_{i\to\infty} cx^{k_i} \geqq f^*$.

4. APPLICATION

In this section, the algorithm is applied to a dual approach to mathematical programming problems. Let us consider the mathematical program defined by

$$\begin{aligned}&\text{minimize} && f_0(y) \\ &\text{subject to} && f_j(y) \leqq b_j, \qquad j = 1, \cdots, q_0, \\ &&& f_j(y) = b_j, \qquad j = q_0 + 1, \cdots, q, \\ &&& y \in Y,\end{aligned} \tag{15}$$

where Y is a compact set and f_j, $j = 0, 1, \cdots, q$, are continuous real valued functions on Y. Associated with problem (15) is the Lagrangean dual problem

$$\begin{aligned}&\text{maximize} && -bu + v(u) \\ &\text{subject to} && u_j \geqq 0, \qquad j = 1, \cdots, q_0, \\ &&& u \in R^q,\end{aligned} \tag{16}$$

where the function $v: R^q \to R$ is defined as the minimal value of the problem

$$\begin{aligned}&\text{minimize} && f_0(y) + \textstyle\sum_{j=1}^{q} u_j f_j(y) \\ &\text{subject to} && y \in Y.\end{aligned} \tag{17}$$

By the compactness of Y and the continuity of each f_j, $v(u)$ is finite for any u and the set of optimal solutions to problem (17) $Y(u) = \{y \in Y;\ v(u) = f_0(y) + \sum_{j=1}^{q} u_j f_j(y)\}$ is nonempty and compact. Moreover, it can be shown that the function v is concave and the set of subgradients of v is expressed as

$$\partial v(u) = co\{\gamma \in R^q;\ \gamma_j = f_j(y), j = 1, \cdots, q, y \in Y(u)\},$$

(for details, see Grinold [1970]). Thus, the dual problem (16) is a convex program which involves a nonsmooth function.

The use of dual problems has been proved quite effective in various problems such as traveling salesman problems (Held and Karp [1970, 1971]), scheduling problems (Fisher [1973]) and large scale problems (Shapiro [1979], Chapter 6; see also Fisher et al. [1975]).

In order to apply our algorithm to problem (16), we rewrite with no loss of generality problem (16) as

$$\begin{aligned} &\text{maximize} && -ub + \sigma \\ &\text{subject to} && \sigma - v(u) \leqq 0, \\ &&& (u, \sigma) \in X, \end{aligned} \tag{18}$$

where $X = \{(u, \sigma) \in R^{q+1};\ 0 \leqq u_j \leqq M, j = 1, \cdots, q_0, |u_j| \leqq M, j = q_0 + 1, \cdots, q, |\sigma| \leqq M\}$ and M is a sufficiently large positive number. Clearly, problem (18) is of the form of (P) with $m = 1$ and the algorithm in Section 2 can be used to solve this problem. Moreover, subproblem (SP_k) to be solved at each iteration takes the form

$$\begin{aligned} &\text{maximize} && t_k(\sigma - ub) - (½)\|u - u^k\|^2 - (½)(\sigma - \sigma_k)^2 \\ &\text{subject to} && \sigma - \gamma^k u \leqq v(u^k) - \gamma^k u^k, \\ &&& (u, \sigma) \in X, \end{aligned} \tag{19}$$

where $v(u^k)$ and $\gamma^k \in \partial v(u^k)$ may be obtained by solving problem (17) with $u = u^k$. Note that problem (17) is a simple quadratic program with a single inequality constraint.

Finally, we should point out that there is certain flexibility in formulating the dual problem (16). For example, consider a representation of Y such that $Y = \cup_{i=1}^{m} Y_i$, where each Y_i is a nonempty compact convex subset of Y. Then, it is easily seen that problem (16) is equivalent to the following convex program

$$\begin{aligned} &\text{maximize} && -ub + \sigma \\ &\text{subject to} && \sigma - v_i(u) \leqq 0, \qquad i = 1, \cdots, m, \\ &&& (u, \sigma) \in X, \end{aligned} \tag{20}$$

where each v_i is defined as the optimal value of the problem

$$\begin{aligned} &\text{minimize} \quad f_0(y) + \textstyle\sum_{j=1}^{q} u_j f_j(y) \\ &\text{subject to} \quad y \in Y_i. \end{aligned} \tag{21}$$

In this case, associated with problem (20) is the quadratic programming subproblem (19) with the constraint replaced by m inequalities $\sigma - \gamma_i^k u \leqq v_i(u^k) - \gamma_i^k u^k$, $i = 1, \cdots, m$, where $\gamma_i^k \in \partial v_i(u^k)$.

5. NUMERICAL EXAMPLE

To illustate the behavior of the algorithm, the following problem was solved:

$$\begin{aligned} &\text{Maximize} \quad x_5 \\ &\text{subject to} \quad x_5 + f_0(x) + 3\max\{0, f_1(x), f_2(x), f_3(x)\} \leqq 0, \\ &\qquad -100 \leqq x_i \leqq 100, \qquad i = 1, \cdots, 5, \\ &\qquad x \in R^5, \end{aligned}$$

where

$$\begin{aligned} f_0(x) &= x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4, \\ f_1(x) &= x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8, \\ f_2(x) &= x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10, \end{aligned}$$

and

$$f_3(x) = 2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5.$$

This problem is derived from the well-known Rosen-Suzuki problem via an exact penalty function, and has the optimal solution $\bar{x} = (0, 1, 2, -1, 44)$ with the optimal value 44.

The computational results are shown in Figure 1 for three starting points (i) $x^1 = (0, 0, 0, 0, 40)$, (ii) $x^1 = (5, 5, 5, 5, 40)$ and (iii) $x^1 = (10, 10, 10, 10, 40)$, where the sequence $\{t_k\}$ was taken as $t_k = 1/k$. From these results, it is seen that the algorithm could find a near-optimal solution in relatively few iterations but there was difficulty achieving the optimal solution precisely. This kind of behavior is shared with other nonsmooth optimization algorithms such as the cutting plane algorithm and subgradient algorithm (Polyak [1967]). The algorithm was also started at points which are far from the optimal solution. In this case, the algorithm failed to attain a near-optimal solution within 200 iterations, although the sequence was still approaching the optimum. In particular, starting with poor estimates of variable x_5 resulted in slow convergence. This is probably due to the choice of $\{t_k\}$ which becomes very small quickly. Generally speaking, for this kind of algorithm, the choice of parameters is the most critical factor in the speed of convergence.

6. CONCLUSION

We proposed a new type of outer approximation algorithm and proved its convergence. Despite the slow convergence for a numerical example of the previous section, the proposed algorithm offers significant insight in designing methods for dealing with nonsmooth optimization problems. Possible future work includes (i) accelerating the convergence by retaining old constraints to some extent, (ii) using deeper cuts such as those in Veinott, (iii) using approximate subgradients when evaluation of exact

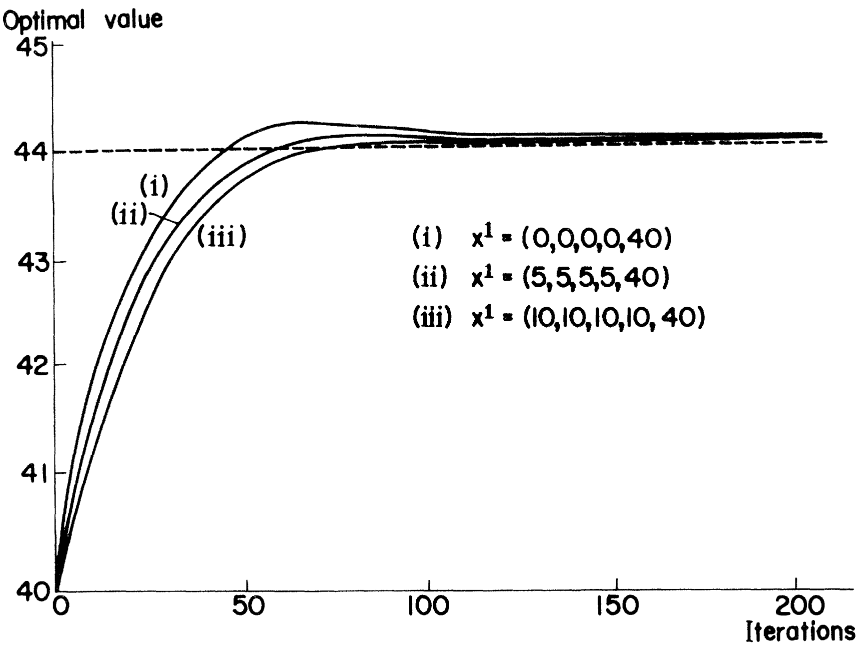

Figure 1. Behavior of the algorithm: number of iterations vs. optimal value of the subproblem.

subgradients is costly, and (iv) combining the algorithm with other methods such as the boxstep method (Marsten et al. [1975]).

ACKNOWLEDGMENTS

This work was performed at the University of Waterloo and supported in part by NSERC grant A7396. I am especially indebted to Professor S. Toida who gave me an opportunity to stay there. I am also grateful to Professor H. Mine of Kyoto University for his encouragement and helpful suggestions and to the referees for their valuable comments on an earlier version of this paper.

REFERENCES

CHENEY, E. W., AND A. A. GOLDSTEIN. 1959. Newton's Method for Convex Programming and Tchebycheff Approximation. *Numerische Math.* **1,** 253–268.

CLARKE, F. H. 1975. Generalized Gradients and Applications. *Trans. Am. Math. Soc.* **205,** 247–262.

DANTZIG, G. B. 1963. *Linear Programming and Extensions.* Princeton University Press, Princeton, N.J.

EAVES, B. C., AND W. I. ZANGWILL. 1971. Generalized Cutting Plane Algorithms. *SIAM J. Control* **9,** 529–542.

ELZINGA, J., AND T. G. MOORE. 1975. A Central Cutting Plane Algorithm for the Convex Programming Problem. *Math. Program.* **8,** 134–145.

FISHER, M. L. 1973. Optimal Solution of Scheduling Problems Using Lagrange Multipliers: Part I. *Opns. Res.* **21,** 1114–1127.

FISHER, M. L., W. D. NORTHUP AND J. F. SHAPIRO. 1975. Using Duality to Solve Discrete Optimization Problems: Theory and Computational Experiences. *Math. Program. Study* **3,** 56–94.

GONZAGA, C., AND E. POLAK. 1979. On Constraint Dropping Schemes and Optimality Functions for a Class of Outer Approximations Algorithms. *SIAM J. Control Opt.* **17,** 477–493.

GRINOLD, R. C. 1970. Lagrangean Subgradients. *Mgmt. Sci.* **17,** 185–188.

HELD, M., AND R. M. KARP. 1970. The Traveling Salesman Problem and Minimum Spanning Trees. *Opns. Res.* **18,** 1138–1162.

HELD, M., AND R. M. KARP. 1971. The Traveling Salesman Problem and Minimum Spanning Trees: Part II. *Math. Program.* **1,** 6–25.

HOGAN, W. W. 1973a. Applications of a General Convergence Theory for Outer Approximation Algorithms. *Math. Program.* **5,** 151–168.

HOGAN, W. W. 1973b. Point-to-Set Maps in Mathematical Programming. *SIAM Rev.* **15,** 591–603.

KELLEY, J. E., JR. 1960. The Cutting-Plane Method for Solving Convex Programs. *J. Soc. Indust. Appl. Math.* **8,** 703–712.

LUENBERGER, D. G. 1969. *Optimization by Vector Space Methods.* John Wiley & Sons, New York.

MANGASARIAN, O. L. 1969. *Nonlinear Programming.* McGraw-Hill, New York.

MARSTEN, R. E., W. W. HOGAN AND J. W. BLANKENSHIP. 1975. The Boxstep Method for Large-Scale Optimization. *Opns. Res.* **23,** 389–405.

PANNE, VAN DE, C. 1975. *Methods for Linear and Quadratic Programming.* North-Holland, Amsterdam.

POLYAK, B. T. 1967. A General Method of Solving Extremum Problems. *Soviet Math. Dokl.* **18**(3), 593–597.

ROBINSON, S. M. 1976. A Subgradient Algorithm for Solving K-Convex Inequalities. In *Optimization and Operations Research* (W. Oettli and K. Ritter, eds.), pp. 237–245. Springer-Verlag, Berlin.

ROCKAFELLAR, R. T. 1970. *Convex Analysis.* Princeton University Press, Princeton, N.J.

SHAPIRO, J. F. 1979. *Mathematical Programming: Structures and Algorithms.* John Wiley & Sons, New York.

TOPKIS, D. M. 1970a. Cutting-Plane Methods without Nested Constraint Sets. *Opns. Res.* **18,** 404–413.

TOPKIS, D. M. 1970b. A Note on Cutting-Plane Methods without Nested Constraint Sets. *Opns. Res.* **18,** 1216–1220.

VEINOTT, A. F., JR. 1967. The Supporting Hyperplane Methods for Unimodal Programming. *Opns. Res.* **15,** 147–152.