Journal of Computational and Applied Mathematics 10 (1984) 147–156
North-Holland

On the convergence of a class of outer approximation algorithms for convex programs

Masao FUKUSHIMA
Department of Applied Mathematics and Physics, Faculty of Engineering, Kyoto University, Kyoto, Japan 606

Received 8 September 1982
Revised 17 October 1983

Abstract: This paper presents a new class of outer approximation methods for solving general convex programs. The methods solve at each iteration a subproblem whose constraints contain the feasible set of the original problem. Moreover, the methods employ quadratic objective functions in the subproblems by adding a simple quadratic term to the objective function of the original problem, while other outer approximation methods usually use the original objective function itself throughout the iterations. By this modification, convergence of the methods can be proved under mild conditions. Furthermore, it is shown that generalized versions of the cut construction schemes in Kelley–Cheney–Goldstein's cutting plane method and Veinott's supporting hyperplane method can be incorporated with the present methods and a cut generated at each iteration need not be retained in the succeeding iterations.

Keywords: Outer approximation algorithm, cutting plane, convex program, subgradient.

1. Introduction

The class of outer approximation methods is one of the fundamental tools for solving general, i.e. nonsmooth, convex optimization problems. The basic idea of earlier methods such as Kelley–Cheney–Goldstein's cutting plane method [1,4] and Veinott's supporting hyperplane method [11] is quite simple. Specifically, one successively generates a halfspace called a cut containing the constraint set of the problem and solves linear programming subproblems whose constraints are defined as the intersection of the previously generated halfspaces. This procedure is very attractive especially to the practitioner since it is easy to understand and to code a computer program.

From a computational viewpoint, however, there is a serious drawback with these methods. That is, the size of the subproblems becomes too large very quickly, because cuts generated at each iteration must be cumulatively retained in the constraints of the subproblems. In [2], Eaves and Zangwill develop some procedures which permit deleting old cuts from the subproblem without spoiling the convergence property. Their work is not only interesting theoretically but also helpful in improving practicability of the earlier methods. However, the difficulty does not seem to be completely resolved, since it does not give any bound on the size of the subproblems and hence the number of retained cuts could still be large.

The purpose of this paper is to present a new class of outer approximation methods and to prove its convergence. These methods differ from others mainly in constructing objective functions of the subproblems. To put it concretely, the present methods employ quadratic objective functions obtained by adding a simple quadratic term to the objective function of the original problem, while other outer approximation methods use the original objective function itself throughout the iterations. By this modification, it can be shown that a cut generated at each iteration need not be retained in the succeeding iterations. This means that, theoretically, one only needs to solve subproblems with a single inequality constraint to obtain convergence, though it might be meritorious to retain some old cuts in order to improve the speed of convergence.

In the next section, we state the method in a somewhat general setting and prove its convergence under certain conditions on constraints of subproblems. In Section 3, we show that such conditions are met by commonly used outer approximation schemes like in [1,4,11], and hence the present method has enough flexibility in designing computational algorithms. In Section 4, computational results are reported.

2. Basic algorithm and its convergence

In this section, we consider the problem

$$\begin{array}{ll} \text{maximize} & cx, \\ \text{subject to} & x \in S, \end{array} \qquad \text{(P)}$$

where cx denotes the inner product of the n-vectors c and x, and S is a nonempty closed convex set in $\mathbb{R}^n$. Throughout this paper, any vector is treated as either a row vector or a column vector, depending on the situation, but the distinction should always be clear from the context in which they are used.

As is well known [5, p. 306; 12, p. 302], any problem of maximizing a concave function over a convex set can be transformed without loss of generality into that of maximizing a linear function over a convex set by introducing an additional variable. Therefore, we can consider any convex program as a problem with a linear objective function. In fact, almost all of the existing outer approximation algorithms are designed to solve problems of this type. Hereafter, we assume the compactness of the feasible set S. This assumption will be needed to ensure the existence of an optimal solution of (P) and the boundedness of sequences generated by the algorithm.

The basic outer approximation algorithm for (P) is stated as follows:

Basic algorithm.

Step 0. Let $\{t_k\}$ be a sequence of positive numbers. Select a starting point x^1 and set $k = 1$.

Step 1. Construct a closed convex set T_k such that $S \subset T_k$.

Step 2. Solve the subproblem

$$\begin{array}{ll} \text{maximize} & cx - \dfrac{1}{2t_k}\|x - x^k\|^2, \\ \text{subject to} & x \in T_k, \end{array} \qquad (\mathrm{SP}_k)$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$, and let the (unique) optimal solution of this subproblem be x^{k+1}.

Step 3. If $x^{k+1} = x^k$, terminate; otherwise, increase k by one and return to Step 1.

It should be noted that problem (SP_k) is equivalent to the problem

$$\begin{array}{ll} \text{minimize} & \|x - (x^k + t_k c)\|^2 \\ \text{subject to} & x \in T_k. \end{array}$$

Namely, x^{k+1} is geometrically the point which is closest to the point $x^k + t_k c$ in the set T_k. Obviously, such a point always exists and is unique by closedness and convexity of T_k.

Let us define the distance from a point x to a nonempty closed set $A \subset \mathbb{R}^n$ by

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

We shall impose the following conditions on the choice of parameters t_k and the construction of the sets T_k:

Condition C1.

$$\lim_{k \to \infty} t_k = 0 \quad \text{and} \quad \sum_{k=1}^{\infty} t_k = +\infty.$$

Condition C2. There exists a positive number κ $(\geqslant 1)$ such that

$$\mathrm{dist}[x^k, S] \leqslant \kappa\, \mathrm{dist}[x^k, T_k]$$

for all k.

Condition C1 is satisfied, for example, by $t_k = 1/j$ for $k_j \leqslant k < k_{j+1}$, where $\{k_j\}$ is an infinite sequence of integers such that $k_1 = 1$ and $k_{j+1} > k_j$. This type of parameter selection rule is also used in the subgradient algorithm [8] for nonsmooth optimization problems. Condition C2 is slightly stronger than that used in [2] in the definition of a cut map in terms of a separator. However, it should be mentioned that this condition is not very restrictive as we shall see in the next section.

Lemma 2.1. *Let S and T be convex sets such that $S \subset T$, and x be an arbitrary point in $\mathbb{R}^n$. If y is the point in T closest to x, then the following inequalities hold:*

$$\|y - z\|^2 \leqslant \|x - z\|^2 - \|y - x\|^2 \quad \forall z \in S$$

and

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

The proof of this lemma is found in [3] (see also [9]).

The main result of this section is now stated as follows:

Theorem 2.2. *Let the Conditions* C1 *and* C2 *be satisfied. Then the basic algorithm either terminates at an optimal solution of* (P) *or generates an infinite sequence $\{x^k\}$ any of whose limit points is an optimal solution of* (P).

Proof. The algorithm terminates only if $x^{k+1} = x^k$ for some k. Since the latter implies $x^k \in T_k$, x^k must belong to S by Condition C2. Moreover, since x^k is the optimal solution of (SP_k), it should also satisfy the following optimality condition for (SP_k):

$$\left[c-(1/t_k)(x-x^k)\right](z-x) \leqslant 0 \quad \forall z \in T_k.$$

Substituting $x = x^k$ in the above inequality and taking account of the fact that $S \subset T_k$, we obtain

$$c(z-x^k) \leqslant 0 \quad \forall z \in S.$$

But this also implies that x^k is an optimal solution to problem (P).

We suppose now that an infinite sequence $\{x^k\}$ is generated. First, we prove that any limit point of this sequence belongs to S by showing that for any $\delta > 0$ there exists an integer $\bar{k}$ such that

$$\text{dist}[x^k, S] \leqslant \delta \quad \forall k \geqslant \bar{k}. \tag{2.1}$$

For each k, let $\tilde{x}^k$ denote the closest point to x^k in T_k. Then, by Lemma 2.1 and Condition C2, for any k,

$$\begin{aligned}\text{dist}[\tilde{x}^k, S]^2 &\leqslant \text{dist}[x^k, S]^2 - \|\tilde{x}^k - x^k\|^2 \\ &= \text{dist}[x^k, S]^2 - \text{dist}\left[x^k, T_k\right]^2 \\ &\leqslant (1-\kappa^{-2})\,\text{dist}[x^k, S]^2,\end{aligned}$$

namely,

$$\text{dist}[\tilde{x}^k, S] \leqslant \left(\sqrt{\kappa^2-1}\,/\kappa\right)\text{dist}[x^k, S]. \tag{2.2}$$

On the other hand, since x^{k+1} is the closest point to $x^k + t_k c$ in T_k, it follows from the nonexpansiveness of projection operators that

$$\begin{aligned}\|\tilde{x}^k - x^{k+1}\| &\leqslant \|x^k - \left(x^k + t_k c\right)\| \\ &= t_k\|c\|.\end{aligned} \tag{2.3}$$

It is also easily verified that

$$\text{dist}[x^{k+1}, S] \leqslant \text{dist}[\tilde{x}^k, S] + \|\tilde{x}^k - x^{k+1}\|. \tag{2.4}$$

Thus, by (2.2), (2.3) and (2.4), we obtain for all k

$$\begin{aligned}\text{dist}[x^{k+1}, S] \leqslant{}& \left(\sqrt{\kappa^2-1}\,/\kappa\right)\text{dist}[x^k, S] \\ &+ t_k\|c\|.\end{aligned} \tag{2.5}$$

Let us define a positive integer k_0 by

$$k_0 = \min\{k \mid t_l\|c\| \leqslant \epsilon\delta/2,\ \forall l \geqslant k\}, \tag{2.6}$$

where ϵ is a positive number small enough to satisfy

$$\alpha \triangleq \sqrt{\kappa^2-1}\,/\kappa + \epsilon < 1. \tag{2.7}$$

Note that k_0 is well defined by Condition C1.

Since (2.1) is obviously satisfied if dist$[x^k, S] \leqslant \delta/2$ for all $k \geqslant k_0$, we assume that dist$[x^{k_1}, S] \geqslant \delta/2$ for some $k_1 \geqslant k_0$. Then there must exist an integer $k_2 > k_1$ such that dist$[x^{k_2}, S] \leqslant \delta/2$, because (2.5) and (2.6) imply

$$\text{dist}[x^{k+1}, S] \leqslant \alpha\,\text{dist}[x^k, S] \tag{2.8}$$

for any k such that dist$[x^k, S] \geqslant \delta/2$ and $k \geqslant k_0$, where α is defined by (2.7). Now it is not difficult to see that

$$\text{dist}[x^k, S] \leqslant \delta \quad \forall k \geqslant k_2.$$

Indeed, if we suppose dist$[x^{k_3}, S] \leqslant \delta <$ dist$[x^{k_3+1}, S]$ for some $k_3 \geqslant k_2$, then by (2.8) we must have dist$[x^{k_3}, S] \leqslant \delta/2$. However, since

$$\begin{aligned}\text{dist}[x^{k_3+1}, S] &\leqslant \text{dist}\left[x^{k_3} + t_{k_3}c, S\right] \\ &\leqslant \text{dist}[x^{k_3}, S] + t_{k_3}\|c\| \\ &\leqslant \delta/2 + \epsilon\delta/2 = (1+\epsilon)\delta/2 < \delta,\end{aligned}$$

we have a contradiction. Thus, (2.1) is proved.

Next, we show that there exists a subsequence $\{x^{k_i}\}$ whose limit point is optimal to (P). Suppose to the contrary that, for some k_4 and $\epsilon > 0$, the following inequalities hold:

$$cx^k \leqslant f^* - \epsilon \quad \forall k \geqslant k_4, \tag{2.9}$$

where $f^* \triangleq \min\{cx \mid x \in S\}$. Let us choose any optimum x^*, i.e. $cx^* = f^*$ and $x^* \in S$. Since x^{k+1} is the closest point to $x^k + t_k c$ in T_k and $x^* \in S \subset T_k$, it follows from Lemma 2.1 that

$$\begin{aligned}&\|x^{k+1} - x^*\|^2 \leqslant \\ &\quad \leqslant \|x^k + t_k c - x^*\|^2 - \|x^k + t_k c - x^{k+1}\|^2 \\ &\quad = \|x^k - x^*\|^2 - \|x^k - x^{k+1}\|^2 + 2t_k c(x^{k+1} - x^*) \\ &\quad \leqslant \|x^k - x^*\|^2 - 2t_k\epsilon.\end{aligned} \tag{2.10}$$

By adding the above inequalities from $k = k_4$ to $k_4 + l$, we have

$$\|x^{k_4+l} - x^*\|^2 \leqslant \|x^{k_4} - x^*\|^2 - 2\epsilon \sum_{k=k_4}^{k_4+l} t_k.$$

However, by Condition C1, the right-hand side of the last inequality becomes arbitrarily small as l increases. This is impossible because the left-hand side is bounded from below by zero. This contradiction implies that (2.9) is not true and that there must exist a subsequence $\{x^{k_i}\}$ such that $\lim_{i\to\infty} cx^{k_i} = f^*$. Moreover, since any limit point of $\{x^k\}$ belongs to S as shown earlier, it also follows that there exists a subsequence of $\{x^k\}$ which converges to an optimum of (P).

Finally, we prove that the limit of any convergent subsequence is an optimum. In order to obtain contradiction, suppose that there exists a subsequence which converges to a nonoptimal point, say x'. Then $cx' < f^*$, because $x' \in S$. Furthermore, since there exists another subsequence converging to an optimum and since it is easily seen that $\lim_{k\to\infty}\|x^{k+1} - x^k\| = 0$, there must be an index j such that

$$cx^{j+1} < f^* \tag{2.11a}$$

and

$$\mathrm{dist}[x^{j+1}, S^*] > \mathrm{dist}[x^j, S^*], \tag{2.11b}$$

where S^* is the set of optimal solutions to (P). However, by Lemma 2.1,

$$\begin{aligned}\|x^{j+1} - z\|^2 &\leqslant \\ &\leqslant \|x^j + t_j c - z\|^2 - \|x^j + t_j c - x^{j+1}\|^2 \\ &= \|x^j - z\|^2 - \|x^{j+1} - x^j\|^2 + 2t_j c(x^{j+1} - z)\end{aligned}$$

for any $z \in S$. Now let z^j be the closest point to x^j in the set S^* and substitute $z = z^j$ in the last inequality. Then we get

$$\mathrm{dist}[x^{j+1}, S^*]^2 \leqslant \mathrm{dist}[x^j, S^*]^2 - \|x^{j+1} - x^j\|^2 + 2t_j c(x^{j+1} - z^j).$$

But, since $cz^j = f^*$, the last inequality contradicts (2.11). Therefore, the limit of any convergent subsequence of $\{x^k\}$ must be optimal to (P). This completes the proof. □

3. Procedures for convex inequality constraints

In this section, we focus our attention on the problem

$$\begin{aligned}&\text{maximize} && cx, \\ &\text{subject to} && g(x) \leqslant 0 \quad \text{and} \quad x \in X,\end{aligned} \tag{$\bar{\mathrm{P}}$}$$

where $g: \mathbb{R}^n \to \mathbb{R}$ is a convex function and X is a compact convex subset of $\mathbb{R}^n$. It is emphasized that the function g is not necessarily differentiable everywhere. Therefore, there is no loss of generality in assuming that the problem contains only one inequality constraint, because any constraint of the form

$$g_i(x) \leqslant 0, \quad i \in I, \qquad x \in X,$$

where each g_i is convex, can be converted into

$$g(x) \leqslant 0 \quad \text{and} \quad x \in X$$

by defining g by $g(x) = \sup\{g_i(x) \mid i \in I\}$. It is also noted that problem ($\bar{\mathrm{P}}$) is usually formulated in such a way that the set X possesses a simple structure like $X = \{x \mid a \leqslant x \leqslant b\}$ or $X = \{x \mid Ax \leqslant b\}$.

In what follows, we denote the feasible set of ($\bar{\mathrm{P}}$) by

$$\bar{S} = \{x \in X \mid g(x) \leqslant 0\}$$

and assume that problem ($\bar{\mathrm{P}}$) satisfies the Slater's condition:

$$\exists a \in X \text{ such that } g(a) < 0. \tag{3.1}$$

Recall that, for any $\epsilon \geqslant 0$, a vector $\gamma \in \mathbb{R}^n$ is called an ϵ-subgradient of g at x [10, p. 219] if

$$g(z) \geqslant g(x) + \gamma(z - x) - \epsilon \quad \forall z \in \mathbb{R}^n.$$

The set of all ϵ-subgradients of g at x is denoted by $\partial_\epsilon g(x)$. Clearly, $\partial_{\epsilon_1} g(x) \subset \partial_{\epsilon_2} g(x)$ if $0 \leqslant \epsilon_1 \leqslant \epsilon_2$, and $\partial_0 g(x)$ coincides with the set of subgradients at x and is simply denoted by $\partial g(x)$.

Lemma 3.1. *For any $x \in \mathbb{R}^n$ and any $\epsilon \geqslant 0$, $\partial_\epsilon g(x)$ is a bounded set. Furthermore, for any bounded sequence $\{\epsilon_k\}$ of nonnegative numbers and any sequence $\{x^k\}$ such that $x^k \in X$, there exists an $L > 0$ such that*

$$\|\gamma^k\| \leqslant L \quad \forall \gamma^k \in \partial_{\epsilon_k} g(x^k).$$

Proof. The set of ϵ-subgradients is characterized by the following relation [10, p. 220]:

$$\partial_\epsilon g(x) = \{\gamma \in \mathbb{R}^n \mid g^*(\gamma) - \gamma x \leqslant \epsilon - g(x)\},$$

where $g^*: \mathbb{R}^n \to (-\infty, +\infty]$ is the convex conjugate of g [10, p. 104]. Since the conjugate of the function h, defined by $h(\gamma) = g^*(\gamma) - \gamma x$, is given by $h^*(y) = g(y + x)$ and hence is finite everywhere, it follows from [10, Cor. 14.2.2] that the set $\partial_\epsilon g(x)$ is bounded.

In order to prove the last half of Lemma 3.1, let us suppose to the contrary that there exists a sequence $\{\gamma^k\}$ such that $\gamma^k \in \partial_{\epsilon_k} g(x^k)$ and $\limsup_{k\to\infty} \|\gamma^k\| = +\infty$. By the compactness of X, we can assume without loss of generality that $\{x^k\}$ converges to a point $x \in X$ and, taking a subsequence if necessary, we can suppose that $\lim_{k\to\infty}\|\gamma^k\| = +\infty$. Let $\epsilon = \sup\{\epsilon_k \mid k = 1, 2, \ldots\}$, which is finite by our assumption. Then we have

$\gamma^k \in \partial_\epsilon g(x^k)$ for all k. However, since $\partial_\epsilon g(\cdot)$ is a continuous point-to-set mapping [6], $\partial_\epsilon g(x)$ must be an unbounded set. This is a contradiction and the proof is complete. □

We now present our first algorithm for solving problem $(\bar{\mathrm{P}})$. This algorithm is regarded as a generalization of the one proposed in [3], though the latter deals with problems having multiple inequality constraints.

Algorithm 3.2.

Step 0. Let β be any number such that $0 \leqslant \beta < 1$, $\{t_k\}$ be a sequence of positive numbers satisfying Condition C1, and $\{s_k\}$ be an arbitrary sequence of nonnegative integers. Select a starting point $x^1 \in X$ and set $k = 1$.

Step 1. Find a $\gamma^k \in \partial_{\epsilon_k} g(x^k)$, where ϵ_k is any number such that

$$\begin{cases} 0 \leqslant \epsilon_k \leqslant \beta g(x^k) & \text{if } g(x^k) > 0, \\ 0 \leqslant \epsilon_k & \text{otherwise.} \end{cases}$$

Step 2. Solve the subproblem

$$\begin{aligned} &\text{maximize} && cx - \frac{1}{2t_k}\|x - x^k\|^2, && (\mathrm{SP}_k^1) \\ &\text{subject to} && g(x^i) + \gamma^i(x - x^i) \leqslant \epsilon_i, \\ &&& i = k, k-1, \ldots, k - s_k, \\ &&& x \in X \end{aligned}$$

and let the optimal solution be x^{k+1}.

Step 3. If $x^{k+1} = x^k$, stop; otherwise, increase k by one and return to Step 1.

Each subproblem is a quadratic program whose objective function contains a very simple quadratic term, provided the set X is a convex polyhedron expressed in terms of a system of linear equalities and inequalities. It should also be noted that the procedure of generating the constraints is a generalization of that used in the Kelley–Cheney–Goldstein's cutting plane algorithm [1,4], where $\epsilon_k = 0$, i.e. $\gamma^k \in \partial g(x^k)$, for each k. The use of such approximate cutting planes is also suggested by Parikh [7] for problems of minimizing a convex function over a convex set.

However, the most remarkable difference between the cutting plane algorithm and the present one lies, of course, in the size of the subproblem to be solved at each iteration. More specifically, the number of constraints of subproblems (SP_k^1) can be bounded in the present algorithm by choosing a bounded sequence $\{s_k\}$, while this is not necessarily the case for the ordinary cutting plane algorithms. In fact, this noncumulative nature is the most attractive feature of our algorithm.

Lemma 3.3. *There exists an* $M > 0$ *such that*

$$\operatorname{dist}[x, \bar{S}] \leqslant Mg(x) \quad \forall x \notin \bar{S},\ x \in X.$$

Proof. Fix a point $a \in X$ satisfying Slater's condition (3.1) and let $x \in X$ be an arbitrary point such that $x \notin \bar{S}$. Define the point y by

$$y = (g(x)a - g(a)x)/(g(x) - g(a)).$$

(Note that $g(a) < 0$ and $g(x) > 0$.) Then by the convexity of g and X, we have $g(y) \leqslant 0$ and $y \in X$, i.e. $y \in \bar{S}$. Hence,

$$\begin{aligned} \operatorname{dist}[x, S] &\leqslant \|x - y\| \\ &= \frac{g(x)}{g(x) - g(a)}\|x - a\|. \end{aligned} \tag{3.2}$$

Since the compactness of X implies the existence of an $m > 0$ such that

$$\|x - a\| \leqslant m \quad \forall x \in X$$

and since

$$g(x) - g(a) > -g(a) > 0,$$

it follows from (3.2) that

$$\operatorname{dist}[x, \bar{S}] \leqslant Mg(x),$$

with $M = -m/g(a) > 0$. □

Theorem 3.4. *Algorithm* 3.2 *either terminates at an optimal solution of* $(\bar{\mathrm{P}})$ *or generates an infinite sequence any of whose limit points is an optimal solution of* $(\bar{\mathrm{P}})$.

Proof. According to the notation of Section 2, we denote the feasible set of (SP_k^1) by

$$\begin{aligned} T_k^1 &= \{x \in X \mid g(x^i) + \gamma^i(x - x^i) \leqslant \epsilon_i, \\ &\qquad i = k, k-1, \ldots, k - s_k\} \\ &= H_k^1 \cap H_{k-1}^1 \cap \cdots \cap H_{k-s_k}^1, \end{aligned}$$

where H_i^1 are the subsets of X defined by

$$H_i^1 = \{x \in X \mid g(x^i) + \gamma^i(x - x^i) \leqslant \epsilon_i\}.$$

Then it is easy to see from the definition of ϵ-subgradients that $\bar{S} \subset T_k^1$ for any k. Thus, in view of Theorem 2.2, we only need to show that there exists a κ such that

$$\operatorname{dist}[x^k, \bar{S}] \leqslant \kappa \operatorname{dist}[x^k, T_k^1] \tag{3.3}$$

for all k. However, since $T_k^1 \subset H_k^1$, we have

$$\operatorname{dist}[x^k, T_k^1] \geqslant \operatorname{dist}[x^k, H_k^1].$$

Therefore, in order to prove (3.3), it suffices to show that

$$\operatorname{dist}[x^k, \bar{S}] \leqslant \kappa \operatorname{dist}[x^k, H_k^1] \tag{3.4}$$

holds for each k such that $x^k \notin \bar{S}$. (Note that (3.3) trivially holds if $x^k \in \bar{S}$, since $S \subset H_k^1$.)

Suppose that $x^k \notin \bar{S}$, i.e. $g(x^k) > 0$. Then, since $\epsilon_k \leqslant \beta g(x^k)$ and $0 \leqslant \beta < 1$, it is not difficult to see that x^k does not belong to H_k^1. This implies that the closest point to x^k in H_k^1, which we denote by y^k, satisfies the equality

$$g(x^k) + \gamma^k(y^k - x^k) = \epsilon_k.$$

Thus it follows from $\epsilon_k \leqslant \beta g(x^k)$ that

$$\begin{aligned}(1-\beta)g(x^k) &\leqslant -\gamma^k(y^k - x^k)\\ &\leqslant \|\gamma^k\| \|y^k - x^k\|\\ &= \|\gamma^k\| \operatorname{dist}[x^k, H_k^1].\end{aligned}$$

So, by Lemma 3.3,

$$\operatorname{dist}[x^k, \bar{S}] \leqslant \frac{M}{1-\beta} \|\gamma^k\| \operatorname{dist}[x^k, H_k^1] \tag{3.5}$$

for some $M > 0$. Since X is compact, the convex function g is bounded on X, so that the sequence $\{\epsilon_k\}$ is also bounded. Thus, from Lemma 3.1, there exists an $L > 0$ such that

$$\|\gamma^k\| \leqslant L \quad \forall k. \tag{3.6}$$

Combining (3.5) and (3.6), we finally obtain

$$\operatorname{dist}[x^k, \bar{S}] \leqslant \kappa \operatorname{dist}[x^k, H_k^1],$$

where $\epsilon = LM/(1-\beta)$. This shows that (3.4) holds and the proof is complete. □

It is worth noting that the convergence of Algorithm 3.2 is still valid if constraints of subproblem (SP_k^1) are replaced by

$$g(x^i) + \gamma^k(x - x^i) \leqslant \epsilon_i, \quad i \in I_k \cup \{k\} \text{ and } x \in X,$$

where I_k is an arbitrary subset of $\{1, 2, \ldots, k-1\}$. This follows from the observation that only the constraint $g(x^k) + \gamma^k(x - x^k) \leqslant \epsilon_k$ plays a crucial role and the presence of any other constraints is not essential in the proof of Theorem 3.4.

We next propose another algorithm which constructs deeper cuts by searching points closer to the constraint set $\bar{S}$.

Algorithm 3.5.

Step 0. Choose any numbers β and λ such that $0 \leqslant \beta < 1$ and $0 < \lambda < 1$, respectively. Let $a \in X$ be a point satisfying the Slater's condition (3.1), let $\{t_k\}$ be a sequence of positive numbers satisfying Condition C1 and let $\{s_k\}$ be an arbitrary sequence of nonnegative integers. Select a starting point $x^1 \in X$ and set $k = 1$.

Step 1. If $g(x^k) > 0$, then go to Step 2; otherwise, set $\hat{x}^k = x^k$ and go to Step 3.

Step 2. Let l_k be the smallest positive integer l such that

$$g(\lambda^l a + (1 - \lambda^l)x^k) > 0$$

and put $\lambda_k = \lambda^{l_k}$ and $\hat{x}^k = \lambda_k a + (1 - \lambda_k)x^k$.

Step 3. Find a $\hat{\gamma}^k \in \partial_{\epsilon_k} g(\hat{x}^k)$, where ϵ_k is any number such that

$$\begin{cases} 0 \leqslant \epsilon_k \leqslant g(\hat{x}^k) - \beta\lambda_k(1+\lambda_k)^{-1} g(a) \\ \qquad \text{if } g(x^k) > 0, \\ 0 \leqslant \epsilon_k \quad \text{otherwise.} \end{cases} \tag{3.7}$$

Step 4. Solve the subproblem

$$\begin{aligned} \text{maximize} \quad & cx - \frac{1}{2t_k}\|x - x^k\|^2, \qquad (\mathrm{SP}_k^2)\\ \text{subject to} \quad & g(\hat{x}^i) + \hat{\gamma}^i(x - \hat{x}^i) \leqslant \epsilon_i,\\ & i = k, k-1, \ldots, k - s_k,\\ & x \in X, \end{aligned}$$

and let the optimal solution be x^{k+1}.

Step 5. If $x^{k+1} = x^k$, stop; otherwise, increase k by one and return to Step 1.

The above constraint generation scheme is regarded as an implementable version of the one in Veinott's supporting hyperplane method [11]. In fact, Veinott's method generates a hyperplane at every iteration using information obtained from a point at which the line segment joining x^k and a intersects the boundary of the set $\bar{S}$. However, this procedure is ideal in the sense that it generally requires an infinite number of operations to find such a point. On the other hand, in the present algorithm, the point $\hat{x}^k$ can always be determined in a finite number of operations.

Lemma 3.6. *There exists a constant $\sigma > 0$ such that*

$$\hat{\gamma}^k(x^k - \hat{x}^k) \geqslant \sigma \|\hat{\gamma}^k\| \|x^k - \hat{x}^k\| \tag{3.8}$$

for all k.

Proof. First note that (3.8) trivially holds for any k such that $x^k \in \bar{S}$, because $x^k \in \bar{S}$ implies $x^k = \hat{x}^k$. So, in what follows, we assume $x^k \notin \bar{S}$, i.e. $g(x^k) > 0$.

Since $\hat{x}^k = \lambda_k a + (1 - \lambda_k)x^k$ and, by the definition of ϵ-subgradients, $g(a) \geqslant g(\hat{x}^k) + \hat{\gamma}^k(a - \hat{x}^k) - \epsilon_k$, we have

$$\hat{\gamma}^k(x^k - \hat{x}^k) \geqslant \lambda_k\{g(\hat{x}^k) - g(a) - \epsilon_k\}. \tag{3.9}$$

From (3.7), this inequality yields

$$\hat{\gamma}^k(x^k - \hat{x}^k) \geqslant$$
$$\geqslant -\lambda_k(1 + \lambda_k)^{-1}\{1 + (1 - \beta)\lambda_k\} g(a) > 0.$$

So, in order to prove Lemma 3.6, it suffices to show that

$$\liminf_{k \to \infty} \frac{\hat{\gamma}^k(x^k - \hat{x}^k)}{\|\hat{\gamma}^k\| \|x^k - \hat{x}^k\|} > 0. \tag{3.10}$$

Now suppose that there exists a subsequence $\{x^{k_i}\}$ such that

$$\lim_{i \to \infty} \frac{\hat{\gamma}^{k_i}(x^{k_i} - \hat{x}^{k_i})}{\|\hat{\gamma}^{k_i}\| \|x^{k_i} - \hat{x}^{k_i}\|} = 0.$$

Since X is compact and $0 < \lambda_k \leqslant \lambda$, the sequence $\{\epsilon_{k_i}\}$ is bounded by (3.7), and hence, by Lemma 3.1, $\{\hat{\gamma}^{k_i}\}$ is bounded also. So we can assume without loss of generality that the sequences $\{x^{k_i}\}$, $\{\hat{x}^{k_i}\}$, $\{\hat{\gamma}^{k_i}\}$, $\{\lambda_{k_i}\}$ and $\{\epsilon_{k_i}\}$ all converge to their respective limits x^0, $\hat{x}^0$, $\hat{\gamma}^0$, λ_0 and ϵ_0. Then it is easily seen that $\hat{\gamma}^0 \in \partial_{\epsilon_0} g(\hat{x}^0)$ and $\epsilon_0 \leqslant g(x^0) - \beta\lambda_0(1 + \lambda_0)^{-1}g(a)$. Thus by the definition of ϵ-subgradients, we have

$$g(a) \geqslant g(\hat{x}^0) + \hat{\gamma}^0(a - \hat{x}^0) - \epsilon_0$$
$$\geqslant \hat{\gamma}^0(a - \hat{x}^0) + \beta\lambda_0(1 + \lambda_0)^{-1}g(a),$$

namely,

$$\hat{\gamma}^0(\hat{x}^0 - a) \geqslant -\left(1 - \frac{\beta\lambda_0}{1 + \lambda_0}\right) g(a) > 0. \tag{3.11}$$

However, since $x^k - \hat{x}^k = \lambda_k(1 - \lambda_k)^{-1}(a - \hat{x}^k)$, (3.10) implies

$$\lim_{i \to \infty} \frac{\hat{\gamma}^{k_i}(x^{k_i} - \hat{x}^{k_i})}{\|\hat{\gamma}^{k_i}\| \|x^{k_i} - \hat{x}^{k_i}\|} = \lim_{i \to \infty} \frac{\hat{\gamma}^{k_i}(a - \hat{x}^{k_i})}{\|\hat{\gamma}^{k_i}\| \|a - \hat{x}^{k_i}\|} = 0,$$

which in turn implies

$$\lim_{i \to \infty} \gamma^{k_i}(a - \hat{x}^{k_i}) = \hat{\gamma}^0(a - \hat{x}^0) = 0$$

by the boundedness of $\{\hat{\gamma}^{k_i}\}$ and $\{\hat{x}^{k_i}\}$. This contradicts (3.11). □

Theorem 3.7. *Algorithm 3.5. either terminates at an optimal solution of* ($\bar{\text{P}}$) *or generates an infinite sequence any of whose limit points is an optimal solution of* ($\bar{\text{P}}$).

Proof. Let us denote the feasible set of (SP_k^2) by

$$T_k^2 = \{x \in X \mid g(\hat{x}^i) + \hat{\gamma}^i(x - \hat{x}^i) \leqslant \epsilon_i,$$
$$i = k, k-1, \ldots, k - s_k\}$$
$$= H_k^2 \cap H_{k-1}^2 \cap \cdots \cap H_{k-s_k}^2,$$

where H_i^2 are the subsets of X defined by

$$H_i^2 = \{x \in X \mid g(\hat{x}^i) + \hat{\gamma}^i(x - \hat{x}^i) \leqslant \epsilon_i\}.$$

Clearly, $\bar{S} \subset T_k^2 \subset H_k^2$ for all k. Therefore, just as in the proof of Theorem 3.4, the present theorem is proved if we show the existence of a constant $\kappa > 0$ satisfying the inequality

$$\text{dist}[x^k, \bar{S}] \leqslant \kappa \, \text{dist}[x^k, H_k^2] \tag{3.12}$$

for each k such that $x^k \notin \bar{S}$.

Suppose $x^k \notin \bar{S}$, i.e. $g(x^k) > 0$, and consider the problem

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}\|x - x^k\|^2, \\ &\text{subject to} && g(\hat{x}^k) + \hat{\gamma}^k(x - \hat{x}^k) \leqslant \epsilon_k. \end{aligned} \tag{3.13}$$

If we denote the optimal solution of (3.13) by z^k, then we have

$$\|z^k - x^k\| \leqslant \text{dist}[x^k, H_k^2], \tag{3.14}$$

because the feasible set of (3.13) is contained in the set H_k^2. Moreover, since (3.9) and (3.7) imply

$$g(\hat{x}^k) + \hat{\gamma}^k(x^k - \hat{x}^k) - \epsilon_k \geqslant$$
$$\geqslant (1 + \lambda_k)(g(\hat{x}^k) - \epsilon_k) - \lambda_k g(a)$$
$$\geqslant -(1 - \beta)\lambda_k g(a) > 0, \tag{3.15}$$

the inequality constraint in (3.13) is necessarily active at z^k. Thus, from the Kuhn–Tucker conditions for (3.13), z^k is explicitly expressed as

$$z^k = x^k - \|\hat{\gamma}^k\|^{-2}\{g(\hat{x}^k) + \hat{\gamma}^k(x^k - \hat{x}^k) - \epsilon_k\}\hat{\gamma}^k.$$

Substituting this into (3.14) and taking account of

(3.15), we obtain

$$\mathrm{dist}[x^k, H_k^2] \geqslant$$
$$\geqslant \|\hat{\gamma}^k\|^{-1}\{g(\hat{x}^k) + \hat{\gamma}^k(x^k - \hat{x}^k) - \epsilon_k\}. \quad (3.16)$$

Since it is not difficult to see from (3.7) and (3.9) that

$$g(\hat{x}^k) - \epsilon_k \geqslant -\beta\{1 + (1-\beta)\lambda_k\}^{-1}\hat{\gamma}^k(x^k - \hat{x}^k)$$
$$\geqslant -\beta\hat{\gamma}^k(x^k - \hat{x}^k),$$

we have the inequality

$$g(\hat{x}^k) + \hat{\gamma}^k(x^k - \hat{x}^k) - \epsilon_k \geqslant (1-\beta)\gamma^k(x^k - \hat{x}^k). \quad (3.17)$$

Hence, using (3.16) and (3.17) and applying Lemma 3.6, we get

$$\mathrm{dist}[x^k, H_k^2] \geqslant \|\hat{\gamma}^k\|^{-1}(1-\beta)\hat{\gamma}^k(x^k - \hat{x}^k)$$
$$\geqslant (1-\beta)\sigma\|x^k - \hat{x}^k\|. \quad (3.18)$$

Let $v^k = (\lambda_k/\lambda)a + (1 - (\lambda_k/\lambda))x^k$. Then, by the construction of λ_k, $g(v^k) \leqslant 0$, i.e. $v^k \in \bar{S}$, so that $\|x^k - v^k\| \geqslant \mathrm{dist}[x^k, \bar{S}]$. But since $v^k - x^k = \lambda^{-1}(\hat{x}^k - x^k)$, we get

$$\|\hat{x}^k - x^k\| \geqslant \lambda\, \mathrm{dist}[x^k, \bar{S}]. \quad (3.19)$$

Combining (3.18) and (3.19), we finally obtain the desired inequality (3.12) with $\kappa = [\sigma\lambda(1-\beta)]^{-1}$. This completes the proof. □

We have shown that the proposed algorithms do not require the approximation of the constraint set to be strictly cumulative. It is noted, however, that the present modification does not improve the rate of convergence of the original algorithms theoretically. This may be observed from the computational results, presented in the next section, which exhibit the diminishing efficiency of additional cuts in later stages of iterations.

4. Computational results

We have implemented Algorithms 3.2 and 3.5 described in the previous section to solve the following small test problem:

$$\text{maximize} \quad cx, \quad (4.1)$$

subject to

$$g(x) = \max\{g_1(x), g_2(x), g_3(x)\} \leqslant 0,$$
$$x \in X = \{x \in \mathbb{R}^5 \mid 0 \leqslant x_i \leqslant 5, i = 1, \ldots, 5\},$$

where

$$cx = 7x_1 + 7x_2 + 7x_3 + 6x_4 + 6x_5,$$
$$g_1(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 + x_1 - x_2 - x_4 + x_5 - 5,$$
$$g_2(x) = 2x_1^2 + 2x_2^2 + x_3^2 + 2x_5^2 + 2x_2 + x_3 + 5x_4 + x_5 - 16,$$
$$g_3(x) = 3x_1^2 + x_2^2 + 2x_4^2 + x_5^2 + x_1 - x_3 - x_4 - 8.$$

Table 1

Iteration	Algorithm 3.2		Algorithm 3.5			KCG algorithm	
	cx^k	$\|x^k - x^*\|$	cx^k	$\|x^k - x^*\|$	NFUN	cx^k	$\|x^k - x^*\|$
1	165.00000	8.94427	165.00000	8.94427	1	165.00000	8.94427
2	101.20377	5.90938	48.44000	4.25206	3	103.50000	7.01783
3	85.68508	4.49072	35.65141	1.24912	6	101.12613	6.43677
4	79.04270	3.70954	33.17567	1.04959	14	92.48786	6.28025
5	58.91937	2.34705	33.15836	0.55111	27	71.55762	4.99214
6	37.35069	1.62628	33.15035	0.34071	46	68.58803	4.04427
7	37.51909	1.03778	33.59944	0.18661	65	68.54307	2.81441
8	33.55543	0.84137	33.02578	0.08372	83	62.37854	2.65672
9	32.88895	0.56031	33.03109	0.05727	112	50.00979	3.50724
10	32.91948	0.42046	33.00301	0.03405	143	48.03216	2.71343
15	32.99353	0.08308	32.99998	0.00555	386	38.88152	1.70923
20	32.99888	0.03031	33.00000	0.00201	737	35.45111	0.96926
30	32.99992	0.00818	32.99975	0.01273	1033	34.11547	0.35309
40	32.99999	0.00330	32.99998	0.00429	1569	34.58001	0.98743
50	33.00068	0.01152	33.00000	0.00208	1891	33.31064	0.44212
60	32.99995	0.00668	33.00000	0.00115	1974	33.41396	1.32024

The optimal solution of this problem is $x^* = (1,1,1,1,1)$ with optimal value $f^* = 33$.

For both algorithms, the point (5,5,5,5,5) was chosen as the starting point, the parameters t_k were set equal to $1/k$ for all k, exact subgradients were used to generate constraints, i.e., $\epsilon_k = 0$ and $\beta = 0$, and the most recently generated five cuts were retained in the subproblems, i.e., $s_k = 5$. Furthermore, for Algorithm 3.5, the point (0,0,0,0,0) was selected as point a satisfying Slater's condition and the parameter λ was set at 0.8.

Computations were performed on a FACOM M-200 computer at Kyoto University. The results are summarized in Table 1, where NFUN in the column 'Algorithm 3.5' designates the total number of function calls to evaluate the value of g and its subgradient up to the indicated iteration. Of course, for Algorithm 3.2, NFUN coincides with the number of iterations, so is ommited from the table. It is observed from the table that near optimal solutions could be obtained in a relatively small number of iterations for both algorithms. After approaching the optimal solution to some degree, however, the convergence was slow as the generated points tended to fluctuate around the optimal solution. It is noted that generated sequences are not necessarily monotonic with respect to both the function value and the distance to the optimal solution. Moreover, it is possible that points are generated which are infeasible but have objective values less than f^*.

Comparing Algorithm 3.5 with Algorithm 3.2, the former converged faster than the latter especially during early iterations. As for the total number of function evaluations, however, the latter was apparently more efficient than the former. Thus, in practice, it may be worthwhile to use a modified version of Algorithm 3.2 which employs the cut generation scheme of Algorithm 3.5 only at some early iterations.

In order to evaluate the proposed algorithms further, the same test problem was solved by the Kelley–Cheney–Goldstein (KCG) cutting plane algorithm. As the original KCG algorithm was expected to encounter computational difficulties due to the rapid increase of constraints of the subproblems, we decided to retain only a fixed number of most recently generated constraints in each subproblem and to drop all other old constraints. By doing so, the monotonicity property of the algorithm is lost and the convergence is no longer guaranteed in theory. However, when the number of the retained constraints was ten, it appeared that the generated sequence was converging, although it was not monotonic. The behavior of the modified KCG algorithm is also shown in

Table 2

Iteration	Algorithm 3.2		Algorithm 3.5			KCG algorithm	
	cx^k	$\|x^k - x^*\|$	cx^k	$\|x^k - x^*\|$	NFUN	cx^k	$\|x^k - x^*\|$
1	243270.0	63.63961	243270.0	63.63961	1	243270.0	63.63961
2	678813.0	544.52906	539083.2	543.42477	3	720966.5	703.66098
3	490828.8	372.68174	174801.8	277.80036	5	703233.4	694.57910
4	409939.0	296.78978	94873.1	118.26225	7	682620.2	695.05448
5	348909.7	231.35430	85678.3	80.69773	9	670478.0	685.47566
6	276463.1	165.65203	76453.8	77.55629	11	664779.9	677.54278
7	248214.8	140.37221	53954.6	62.13399	13	663958.9	673.45669
8	207233.4	100.04105	43517.4	32.58802	15	663006.7	662.55112
9	193923.8	92.39707	41212.9	20.09855	18	662320.6	675.37713
10	169673.8	68.10812	40265.6	18.77762	21	660841.4	666.68603
20	118149.7	33.70981	60153.5	36.68110	52	619797.0	624.33628
30	57359.0	15.39333	46374.3	24.18243	93	599607.3	576.42765
40	38222.0	7.00633	31699.3	10.04953	146		
50	29678.0	4.85211	28440.8	4.55354	207		
100	26734.5	3.18294	26066.3	3.01205	651		
200	24883.4	1.54641	25853.4	1.63492	1896		
300	24537.0	0.81298	24606.3	0.95492	3386		
400	24464.0	0.58558	24462.1	0.61235	5095		
500	24431.3	0.50167	24454.5	0.61340	6976		

Table 1. Of course, up to the tenth iteration, the generated sequence should be identical with the one generated by the original KCG algorithm. It can be observed that the convergence of the KCG algorithm was considerably slower than that of Algorithms 3.2 and 3.5 even in earlier iterations. Computational times to perform 60 iterations were: about 2 seconds for the KCG algorithm, about 2 seconds for Algorithm 3.2, and about 3 seconds for Algorithm 3.5.

We have also solved a larger problem of a type similar to problem (4.1). The problem contains fifty variables and the function g is defined as the maximum of fifty convex quadratic functions. The optimal value of this problem is 24327. We have applied Algorithms 3.2 and 3.5, with $t_k = 1/k$ and $s_k = 10$ for all k, to solve this problem. The results are summarized in Table 2. This table also presents the initial thirty iterations of the KCG algorithm applied to the same problem. In view of Table 2, we may notice that the observations made for problem (4.1) hold for this larger problem as well.

Acknowledgment

A part of this work was performed at the University of Waterloo and supported by NSERC grants A7396 and A4315. I am sincerely indebted to Professor H. Mine of Kyoto University, Professor S. Toida of Old Dominion University and Professor T. Kameda of Simon Fraser University for their support and discussions. I also wish to thank M. Nakamura and Y. Tanaka for carrying out the computational experiments.

References

[1] E.W. Cheney and A.A. Goldstein, Newton's method for convex programming and Tchebycheff approximation, *Numer. Math.* **1** (1959) 253–268.

[2] B.C. Eaves and W.I. Zangwill, Generalized cutting plane algorithms, *SIAM J. Control* **9** (1971) 529–542.

[3] M. Fukushima, An outer approximation algorithm for solving general convex programs, *Oper. Res.* **31** (1983) 101–113.

[4] J.E. Kelley, Jr., The cutting-plane method for solving convex programs, *J. SIAM* **8** (1960) 703–712.

[5] D.G. Luenberger, *Introduction to Linear and Nonlinear Programming* (Addison-Wesley, Reading, MA, 1973).

[6] E.A. Nurminskii, Continuity of ϵ-subgradient mappings, *Cybernetics* **13** (1977) 790–791.

[7] S.C. Parikh, Approximate cutting planes in nonlinear programming, *Math. Programming* **11** (1976) 194–198.

[8] B.T. Polyak, A general method of solving extremum problems, *Soviet Math. Dokl.* **8** (1967) 593–597.

[9] S.M. Robinson, A subgradient algorithm for solving *K*-convex inequalities, in: W. Oettli and K. Ritter, Eds., *Optimization and Operations Research* (Springer, Berlin, 1976) 237–245.

[10] R.T. Rockafellar, *Convex Analysis* (Princeton Univ. Press, Princeton, NJ, 1970).

[11] A.F. Veinott, Jr., The supporting hyperplane methods for unimodal programming, *Oper. Res.* **15** (1967) 147–152.

[12] W.I. Zangwill, *Nonlinear Programming: A Unified Approach* (Prentice-Hall, Englewood Cliffs, NJ, 1969).