ACTA MATHEMATICA VIETNAMICA

Volume 8 Number 1 (1983)

MINIMIZING A CONCAVE FUNCTION OVER A COMPACT CONVEX SET

HOANG TUY AND NGUYEN QUANG THAI

Institute of Mathematics

Hanoi

1. INTRODUCTION

The problem of minimizing a concave function over a polytope has been extensively studied in recent years. In [3] (where a bibliography on the subject can be found), Thoai anh Tuy developed a method for solving this problem which is based upon a combination of the branch and bound procedure with the cutting plane technique elaborated earlier by Tuy [2].

In this paper we shall consider a more general problem, viz.

$$\text{Minimize } f(x), \text{ s. t. } x \in D, \tag{1}$$

where D is a compact convex set in R^n, f is a finite concave function defined throughout R^n. To our knowledge, this problem was first studied by Horst [1].

The method to be presented below for solving (1) proceeds along the lines of Thoai and Tuy in [3], with, however, an important improvement in the bounding procedure. This improvement on the one hand enables the method to extend to convex constraints, on the other hand dispenses us with the use of linear subprograms in the bound estimation. From a computational point of view, this may be an advantage, since computational experience with the algorithm in [3] has shown that most of the computation time was spent on solving the linear subprograms.

It should be noted that the problem (1) includes as a special case the conventional problem of minimizing a convex function $F(x)$ over a compact convex set D. Indeed, the latter problem can be formulated as that of minimizing x_{n+1}, subject to $F(x) \leqslant x_{n+1}$, $x \in D$. Therefore, our method will provide, as a by-product, an algorithm for solving convex and, in particular, linear programs, which is not related in any way to the simplex method.

2. DESCRIPTION OF THE METHOD

We shall assume that the constraint set D is defined by a system of inequalities of the form

$$g_i(x) \leqslant 0 \ (i = 1, \dots, m) \qquad (2)$$

where g_i are finite convex functions on R^n satisfying Slater's condition:

$$\text{There is one } x \text{ s.t. } g_i(x) < 0 \text{ for all } i \qquad (3)$$

So D has a nonempty interior and, by translating if necessary, we may assume $0 \in \text{int}\, D$, i.e. $g_i(0) < 0$ $(i = 1, \dots, m)$.

We shall engage in an iterative process in each step of which we shall have to examine a collection of cones vertexed at 0 and to perform three basic operations:

1. select a cone from this collection;
2. split this cone into two smaller cones;
3. assign to each newly generated cone M a number $\mu(M)$ which is a lower bound for $f(x)$ over $D \cap M$.

The precise rules for the last two operations will be specified later. Assuming for the moment that these rules are known, the method we propose can be described as follows.

Take an n-simplex $T = [s^1, \dots, s^{n+1}]$ in R^n containing 0 in its interior. For each $j = 1, \dots, n+1$ denote by $M_{o,j}$ the polyhedral convex cone generated by n halflines from 0 through s^i with $i \neq j$.

Initialization. Set $x^0 = \arg\min \{f(x) : x = \theta s^j,\ \theta > 0,\ \theta s^j \in D,\ j = 1, \dots, n+1\}$; $\mathcal{M}_o = \{M_{o,1}, \dots, M_{o,n+1}\}$. Compute $\mu(M_{o,j})$ for each $j = 1, \dots, n+1$.

Step $k = 0, 1, \dots$ Delete all cones $M \in \mathcal{M}_k$ with $\mu(M) \geqslant f(x^k)$. Let $\mathcal{K}_k$ be the set of remaining cones.

If $\mathcal{K}_k$ is empty, stop: x^k is a solution. Otherwise, select $M_k \in \mathcal{K}_k$ such that $\mu(M_k) = \min \{\mu(M) : M \in \mathcal{K}_k\}$, and split it into two subcones $M_{k,1}$, $M_{k,2}$. For each $h=1,2$ compute $\mu(M_{k,h})$. These operations generate some new points of D. Let x^{k+1} be the new current best feasible solution, i. e. the best among x^k and all newly generated points of D; let $\mathcal{M}_{k+1} = (\mathcal{K}_k \setminus \{M_k\}) \cup \{M_{k,1}, M_{k,2}\}$. Go to step $k+1$.

To complete the description of the method we still have to specify the rules for performing the two operations 2) and 3) mentioned above.

As in [3] we use the following *bisection rule* for splitting a cone M_k into two subcones.

Suppose that M_k is generated by n halflines emanating from O and cutting some facet of the simplex T at points $v^{k,1},\dots, v^{k,n}$ respectively. Take the longest side of the simplex $[v^{k,1},\dots, v^{k,n}]$ (if there are more than one candidates, choose any of them). Let it be $[v^{k,i_1}, v^{k,i_2}]$ and let u^k be its midpoint. For each $h=1,2$ denote by $M_{k,h}$ the cone whose set of edges obtains from that of M_k by replacing the edge passing through v^{k,i_h} with the halfline from 0 through u^k. Then it is straight forward that M_k is the union of $M_{k,1}$ and $M_{k,2}$.

Denote $S_k = [v^{k,1},\dots, v^{k,n}]$. In [3] it was estabilished that:

LEMMA 1. *If* M_{k_q}, $q = 1,2,\dots$, *is an infinite decreasing sequence of cones such that* $M_{k_{q+1}}$ *obtains from* M_{k_q} *by a bisection as described above, then* diam $S_{k_q} \to 0$ as $q \to \infty$. *In other words* $\bigcap_{q=1}^{\infty} M_{k_q}$ *is a halfline from* 0 *through* $z^* \in \bigcap_{q=1}^{\infty} S_{k_q}$.

Following [3] we say that the bounding operation (i. e. the operation of computing $\mu(M)$ for every given cone M) is *consistent* if for any infinite decreasing sequence of cones M_{k_q} tending to a halfline as described in Lemma 1, we have $f(x^{k_q}) - \mu(M_{k_q}) \to 0$ as $q \to \infty$.

(recall that M_k is the cone to be split in step k, x^k is the current best feasible solution in this step).

In [3] it was also established that:

LEMMA 2. *Let each* M_k *be split according to the above bisection rule. If the bounding operation is consistent then: either the procedure terminates at some step* k *with an optimal solution* x^k, *or it generates an infinite sequence* $\{x^k\}$, *every cluster point of which is an optimal solution. Each point* x^k *is an approximate optimal solution with an error not exceeding* $\gamma_k - \mu_k$, *where*

$$\gamma_k = f(x^k),\ \mu_k = \min\ \{\mu(M): M \in \mathcal{R}_k\} = \mu(M_k).$$

Thus, to obtain a convergent algorithm for solving our problem (1) according to the above scheme the only thing that remains to be done is to construct a consistent bounding operation. One such bounding method was provided in [3]. This bounding method, however, applies only to the case where D is a polytope; furthermore, it requires for each bound estimation the solution of an auxiliary linear program, which may be time-consuming. Therefore, we must seek a more suitable bounding method for the general case.

3. CONSISTENT BOUNDING OPERATION

Let M be a newly generated cone. We wish to compute a lower bound for $f(x)$ over the set $D \cap M$, using a simple enough procedure.

Denote by $v^1,\ldots, v^n$ the points where the n edges of M meet the boundary ∂T of $T = [s^1,\ldots, s^{n+1}]$ ($v^1,\ldots, v^n$ lies in one facet of T). For each $j = 1,\ldots, n$ let

$$\theta_j = \sup\ \{\theta : \theta v^j \in D\}. \qquad (4)$$

Clearly $0 < \theta_j < +\infty$, because D is bounded by hypothesis. Take any point $z = z(M) \in M \cap \partial D$ (∂D denotes the boundary of D). For instance, take $z = \theta_j v^j$ for some $j = 1,\ldots, n$. Since $z \in \partial D$, there is an index i such that $g_i(z) = 0$. Select any $t \in \partial g_i(z)$ (the subdifferential of g_i at point z)(1). Then $g_i(x) - g_i(z) \geqslant \langle t, x - z\rangle$ and hence,

$$g_i(x) \geqslant \langle t, x-z\rangle \text{ for all } x \in D. \qquad (5)$$

Noting that $g_i(0) < 0$, this yields $\langle t, z\rangle \geqslant -g_i(0) > 0$, which implies, in particular, that $t \neq 0$. Furthermore, it follows from (5) that D is contained in the halfspace

$$H = \{x \in R^n : \langle t, x-z\rangle \leqslant 0\}.$$

---

(1) Since g_i is a convex function finite throughout R^n, it is continuous and subdifferentiable everywhere.

Therefore, a lower bound for $f(x)$ over $D \cap M$ is furnished by the minimum of f over the set $H \cap M$ (whose structure is very simple). Specifically, let us distinguish two cases:

1. If $\langle t, v^j \rangle > 0$ for all $j = 1, \ldots, n$, then the hyperplane $\langle t, x - z \rangle = 0$ (which is a supporting hyperplane to D at z) cuts every j — th edge of M at point

$$z^j = S_j\, v^j, \text{ with } S_j = \langle t, z \rangle / \langle t, v^j \rangle$$

and, by the concavity of f, the minimum of f over the simplex $H \cap M$ equals

$$\alpha(M) = \min \{ f(0), f\langle z^1 \rangle, \ldots, f\langle z^n \rangle \}.$$

Consequently, in this case we set

$$\mu(M) = \begin{cases} \alpha(M) \text{ if } M \in \mathcal{M}_0 ; \\ \max \{ \alpha(M), \mu(M_{anc}) \} \text{ otherwise} \end{cases} \tag{6}$$

where M_{anc} denotes the immediate ancester of M.

2. If $\langle t, v^j \rangle \leqslant 0$ for at least one j, then the j — th edge of M lies entirely in H and the set $H \cap M$ is unbounded. In this case we set

$$\mu(M) = \begin{cases} -\infty \text{ if } M \in \mathcal{M}_0 ; \\ \mu(M_{anc}) \text{ otherwis.} \end{cases} \tag{7}$$

(of course, the new bound does not improve upon the old one, but, as will be seen in the proof Lemma 3 below, this unpleasant case cannot occur if the cone M is sufficiently «thin»)

Thus the rule for computing $\mu(M)$ is extremely simple.

LEMMA 3. *The above bounding operation is consistent.*

Proof. Consider an arbitrary infinite decreasing sequence of cones M_{k_q}, such that $\Gamma = \bigcap_{q=1}^{\infty} M_{k_q}$ is a halfline. Let $v_q = f(x^{k_q})$, $\mu_q = \mu(M_{k_q})$, we shall show that

$$v_q - \mu_q \to 0 \ (q \to \infty), f(z^*) = v^*, \tag{8}$$

where $v^* = \lim_{q \to \infty} v_q$ (this limit obviously exists) and z^* denotes the other endpoint of the line segment $D \cap \Gamma$.

Let v^* and v^q be the points where ∂T meets the halfline Γ and the halfline from 0 through $z^q = z\,(M_{k_q})$, respectively. We have $v^q \to v^*$ and

$$z^* = \zeta_* \; v^*, \; z^q = \zeta_q \, q^{v^q} \text{ for some } \zeta_* > 0, \; \zeta_q > 0.$$

If φ is the gauge of the convex set D, then $\varphi\,(z^*) = \varphi\,(\zeta_* \; v^*) = 1$, hence $\zeta_* = 1 \,/\, \varphi\,(v^*)$. Similarly, $\varphi\,(\zeta_q \, q^{v^q}) = 1$, hence $\zeta_q = 1 \,/\, \varphi\,(\,v^q\,)$. But, φ being continuous, $\zeta_q \to \zeta_*$, proving that $z^q \to z^*$.

Now denote by $v^{q,j}$ and t^q the vectors v^j, t constructed as indicated above for the cone $M = M_{k_q}$. Clearly, for every j, $v^{q,j} \to v^*$ as $q \to \infty$. Since $t^q \neq 0$, we may assume, by taking a subsequence if necessary that $t^q \,/\, \| t^q \| \to t^*$.

We have seen that $\langle t^q , x - z^q \rangle \leqslant 0$ for all $x \in D$. Hence, passing to the limit, we get $\langle t^*, x - z^* \rangle \leqslant 0$ for all $x \in D$, which implies, by making $x = 0$, $\langle t^*, z^* \rangle \geqslant 0$. Furthermore, we cannot have $\langle t^*, z^* \rangle = 0$ since then $\langle t^*, x \rangle \leqslant 0$ for all $x \in D$, which would conflict with the fact $0 \in \operatorname{int} D$ and $t^* \neq 0$. So $\langle t^*, z^* \rangle > 0$, and hence, $(t^*, v^*) > 0$. This implies that for all large enough q we must have $\langle t^q , v^{q,j} \rangle > 0$ for every j, i.e. we must be in the case 1) considered above. Denoting by $z^{q,j}$, $\zeta_{q,j}$ the z^j, ζ_j constructed for $M = M_{k_q}$, we can then write

$$z^{q,j} = \zeta_{q,j} \, v^{q,j}, \text{ with } \zeta_{q,j} = \frac{\langle t^q , z^q \rangle}{\langle t^q , v^{q,j} \rangle} \to \frac{\langle t^*, z^* \rangle}{\langle t^*, v^* \rangle} = \zeta_* ,$$

so that $z^{q,j} \to z^*$.

Since for all large enough q we shall be in the case 1), it follows that $\mu\,(M_{k_q}) \geqslant \alpha\,(M_{k_q})$ (see (6)). Noting that $\mu_q = \mu\,(M_{k_q}) < f\,(x^{k_q}) = \gamma_q < f\;(0)$ (because $M_{k_q} \in \mathcal{R}_{k_q}$), we deduce $f\,(z^{q,j}) < \gamma_q$ for at least some $j = j(q) \in \{1, \ldots, n\}$. Hence, passing to the limit, $f\,(z^*) \leqslant \gamma^*$ (the continuity of f follows from the fact that f is concave and defined throughout R^n). On the other hand, $\gamma_{q+1} \leqslant f\;(z^q)$, and consequently, $\gamma^* \leqslant f\;(z^*)$.

Therefore, $f\;(z^*) = \gamma^*$. But we have seen that $\gamma_q > \mu_q \geqslant f\;(z^{q,j})$.

This implies $\gamma_q - \mu_q \to 0$, completing the proof of the Lemma.

Remark 1. We have shown that the unpleasant case 2) where $\mu(M)$ is computed according to formula (7) cannot occur when the cones become sufficiently «thin». However, for the speed of convergence, it is important to avoid this case whenever possible, even in the early stage of the algorithm. This can be achieved by choosing the point $z = z(M)$ and the vector $t = t(M)$ that determines the supporting hyperplane to D at z, in such a way that this hyperplane meets all edges of the cone M and defines a simplex $[O, z^1, \dots, z^n]$ bounding M as tightly as possible. Often a fairly good bound may be obtained without much effort if z is taken to be the intersection of ∂D with the ray from 0 through the barycentre of the simplex $[v^1, \dots, v^n]$. To get a still better bound one can consider $\overline{z} = \arg\min \{ f(z^j) : j = 1, \dots, n \}$. If $\overline{z} \in D$, then $f(\overline{z})$ is the minimum of f over $D \cap M$. Otherwise, $g_{i_0}(\overline{z}) = \max_i g_i(z) > 0$. Let $\overline{t} \in \partial g_{i_0}(\overline{z})$. Then the hyperplane $\overline{H} = \{ \bar{x} : \langle \overline{t}, x - \overline{z} \rangle + g_{i_0}(\overline{z}) = 0 \}$ cuts off $\overline{z}$ from the simplex $[O, z^1, \dots, z^n]$. The remaining part of this simplex is a polytope Δ containing $D \cap M$, so we can take as $\mu(M)$ the minimum of f over Δ. This minimum can be computed by comparing the values of f at the vertices of Δ. Since the structure of Δ is simple, the computation of the vertex set of Δ presents no difficulty.

Remark 2. Following the bounding method developed in [5], instead of choosing first the point $z = z(M)$ where a supporting hyperplane to D must be drawn, one seeks the supporting hyperplane to D that is parallel to the hyperplane passing through the n points where the n edges of M cut the boundary of the level set $\{ x \in R^n : f(x) \leqslant f(x^k) \}$ (k being the index of the current step). As can easily be seen, when D is polytope this amounts to solving a linear subprogram (the solution of this linear subprogram is just the point $z = z(M)$).

This method requires much more computational efforts than the one presented above, even though it provides in general a better value of $\mu(M)$, i.e. a value closer to the exact minimum of $f(x)$ over $D \cap M$.

4. CASE WHERE NO INTERIOR POINT OF D IS AVAILABLE

So far we assumed that at the beginning an interior point of D is available. We now consider the general case.

I. If D is a *polytope* and we know a nondegenerate extreme point of D — without loss of generality it may be assumed that this extreme point is O — then we may take as the initial collection $\mathcal{M}_0 = \{ M_0 \}$ where M_0 is the smallest cone vertexed at O that contains D. The reader will have no difficulty in verifying that, starting from M_o all the procedure described above can be applied without any further modification.

II. Suppose now that D is a compact convex set defined by (2) and we do not know whether or not it possesses an interior point. Let us consider the auxiliary problem:

Minimize y, subject to $g_i(x) - y \leqslant o$ $(i = 1, \ldots, m)$ (9)

Let x^o an arbitrary point of R^n, y^o a number so large that $g_i(x^o) - y^o < 0$ $(i = 1, \ldots, m)$. Then $(x^o, y^o) \in R^{n+1}$ is an interior point of the constraint set of problem (9) and starting from it one can solve (9) by the previous method. If after a certain number of steps we obtain a feasible point $(\bar{x}, \bar{y}) \in R^{n+1}$ with $\bar{y} < o$, then $\bar{g}_i(x) \leqslant \bar{y} < 0$ $(i = 1, \ldots, m)$, which means that $\bar{x} \in$ int D. This event necessarily occurs provided int $D \neq \phi$. Starting now from $\bar{x}$ we can applyl the previous method to the original problem (1).

If the optimal value y^* of (9) is positive, this indicates that the origina problem has no feasible solution.

If $y^* = 0$ then int $D = \phi$. In that case we can always stop the process of solving (9) when a feasible point $(\bar{x}, \bar{y})$ is reached such that $\bar{y}$ is sufficiently small, for example $\bar{y} < \varepsilon$: since $g_i(\bar{x}) < \varepsilon$ (i = 1, ..., m), $\bar{x}$ is an interior point of the enlarged constraint set $D_\varepsilon = \{x : g_i(x) \leqslant \varepsilon \ (i = 1, \ldots, m)\}$. Starting from $\bar{x}$ we can then use the previous method to find the minimum of $f(x)$ over D_ε. This will provide an approximate minimum of $f(x)$ over D. *)

Received May 15, 1981

REFERENCES

[1]. R. Horst. *An algorithm for nonconvex programming problems.* Math. Programming 10 (1976), 312 — 321.

[2]. H. Tuy *Concave programming under linear constraints.* Doklady Abad. Nauk SSSR 159 (1964), 32 — 35 *(in Russian).*

[3]. Ng. v. Thoai and H. Tuy. *Convergent algorithms for minimizing a concave function.* Mathematics of Operations Research 5 (1980), 556 — 566.

*) The present method can be extended to the case of an unbounded constraint set was. A subsequent paper will be devoted to this extension.