JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 46, No. 1, MAY 1985

Lower Subdifferentiable Functions and Their Minimization by Cutting Planes[1]

F. PLASTRIA[2]

Communicated by O. L. Mangasarian

Abstract. This paper introduces lower subgradients as a generalization of subgradients. The properties and characterization of boundedly lower subdifferentiable functions are explored. A cutting plane algorithm is introduced for the minimization of a boundedly lower subdifferentiable function subject to linear constraints. Its convergence is proven and the relation is discussed with the well-known Kelley method for convex programming problems. As an example of application, the minimization of the maximum of a finite number of concave-convex composite functions is outlined.

Key Words. Lower subgradients, boundedly lower subdifferentiable functions, quasiconvex functions, Lipschitz functions, cutting plane algorithm.

1. Introduction

One of the earliest methods for nonlinear optimization was the cutting plane method of Kelley (Ref. 1) and Cheney and Goldstein (Ref. 2). At each iteration of this algorithm, a linear program is solved obtained by the linearization of the nonlinear function(s) defining the problem. The exactness and convergence of these algorithms were only ensured for convex functions. It is indeed only possible to construct linear lower approximating functions at each point when the source function is convex.

However, when a minimization problem is concerned, the only points of interest are those where the objective function is less than the values observed previously. Based on this idea, the notion of lower subdifferential functions arises naturally as those functions that can be approximated

[1] The author thanks the referees for several constructive remarks.

[2] Assistant Lecturer, Centrum voor Statistiek en Operationeel Onderzoek, Vrije Universiteit Brussel, Brussels, Belgium.

below, on the set of these points of interest, by a linear function. It appears that this class of functions is strictly larger than the class of convex functions.

In Sections 2 and 3, the characterization of this class and the properties with respect to some operators are studied. Section 4 discusses a cutting plane method for the minimization of a boundedly lower subdifferentiable function subject to linear constraints. Kelley's cutting plane method (Ref. 1) may be viewed as a special case of our method when the objective is convex and all constraints are linear.

In Section 5, we indicate some possible applications for our method. Section 6 suggests future work concerning the extension of our cutting plane method to problems with nonlinear constraints.

2. Lower Subdifferentiability

A real-valued function f defined on $K \subset E^n$ is called *subdifferentiable* at the point $x \in K$ if there exists a vector $x^0 \in E^n$ such that, for any $y \in K$, we have

$$\langle y-x, x^0\rangle + f(x) \leq f(y). \tag{1}$$

In this case, x^0 is a *subgradient* of f at x, and the set of all such vectors is the *subdifferential* of f at x, denoted by $\partial f(x)$.

If f admits a subgradient at any point of K, it is *convex* on K.

The theory of subgradients is well known, and Rockafellar (Ref. 3) gives a thorough treatment of the subject. We extent this notion as follows.

The function f is *lower subdifferentiable* at x on $K \subset E^n$ if there exists a vector x^0 such that (1) holds for any $y \in K$ with $f(y) < f(x)$. We will then call x^0 a *lower subgradient* of f at x on K and denote the set of all these by $\partial^- f(x)$. A function is *lower subdifferentiable* (*lsd*) on $K \subset E^n$ if it admits at least one lower subgradient at each point of K.

Note the semiglobal character of lower subgradients. Their definition involves all points of lower functional value. This is similar to the global character of ε-subgradients (see, e.g., Hiriart-Urruty, Ref. 4). Other generalized subgradients, as introduced by Clarke (Ref. 5) and Rockafellar (Ref. 6), are of local character. Therefore, there does not seem to be a close conversely, as the real function $|x|^{1/2}$ shows. The class of lsd functions is in

It is clear that any convex function is lsd, since $\partial f(x) \subset \partial^- f(x)$, but not conversely, as the real function $|x|^{1/2}$ shows. The class of lsd functions is in fact strictly enclosed between quasiconvex functions and convex functions.

A function f is *quasiconvex* on K when all level sets

$$S_c = \{x \in K \mid f(x) \leq c\}$$

are convex (Mangasarian, Ref. 7).

Theorem 2.1. If K is convex and closed, then every lsd function on K is quasiconvex and lower semicontinuous on K.

Proof. If S_c is void, it is convex. If not, then, for any $y \in K$ outside S_c, we can choose a lower subgradient y^0 for f at y. Set

$$d_y = c - f(y),$$

and let

$$H^-(y) = \{x \mid \langle x - y, y^0 \rangle \leq d_y\}.$$

Then, $y \notin H^-(y)$; and, for any $x \in S_c$, we have

$$f(x) \leq c < f(y),$$

and so

$$\langle x - y, y^0 \rangle \leq f(x) - f(y) \leq d_y.$$

This shows that S_c is the intersection of the closed convex sets $H^-(y) \cap K$, where y ranges over $K \setminus S_c$. Hence, S_c is both closed and convex. □

There exist continuous quasiconvex functions that are not lsd; for example, the function $-(1-x^2)^{1/2}$, $x \in [-1, 1]$, and zero elsewhere is not lower subdifferentiable at 1 nor at -1.

We say that f is *boundedly lower subdifferentiable* (*blsd*) on K, if, at each point of K, there exists a lower subgradient of f of norm not exceeding a constant N, which will be called the *blsd-bound* of f.

Theorem 2.2. Every blsd function f on K is a Lipschitz function on K.

Proof. It is sufficient to consider two points x and y with different function values, and without loss of generality we may suppose $f(x) < f(y)$. There then exists, by the blsd property of f at y, a vector y^0, with $\|y^0\| \leq N$, such that

$$0 < f(y) - f(x) \leq \langle y - x, y^0 \rangle \leq \|y - x\| \cdot N,$$

which proves the theorem. □

Theorem 2.3. Every quasiconvex function f on E^n satisfying a Lipschitz condition with constant N is blsd on E^n with blsd-bound N.

Proof. Let a be a point of E^n. Then, the strict level set

$$S = \{x \mid f(x) < f(a)\}$$

is convex due to the quasiconvexity of f (see Ponstein, Ref. 8). Since f is Lipschitz, it is continuous, and S is an open convex set not containing a. There then exists a separating hyperplane for a and S: there exists u^0, with $\|u^0\| = 1$, such that, for any x in S,

$$\langle x-a, u^0\rangle < 0.$$

Set $a^0 = Nu^0$. We proceed to show that $a^0 \in \partial^- f(a)$. Since $\|a^0\| = N$, this will terminate the proof.

For any x in S, call x' the orthogonal projection of x on the hyperplane

$$\langle z-a, a^0\rangle = 0.$$

Then x' lies outside S, or

$$f(x') \geq f(a);$$

furthermore,

$$\langle a-x, a^0\rangle = \|x-x'\| \cdot \|a^0\| = N\|x-x'\|.$$

Thus, we have

$$f(a)-f(x) \leq f(x')-f(x) \leq N\|x-x'\| = \langle a-x, a^0\rangle,$$

showing that a^0 is a lower subgradient of f at x. □

It follows that, for functions with domain E^n, blsd is equivalent to being quasiconvex and Lipschitz.

This easy characterization of blsd functions is unfortunately not true for a general domain K. The following function is indeed a counterexample.

Define $K = [-1, 1] \times E$; and, for any $i \in \mathbb{N}$, call p_i the point $(0, 3i)$. For any $a = (a_1, a_2) \in K$, define i_a as the integer part of $a_2/3$. Denote by d_a the Euclidean distance of a and p_{i_a}. Consider now, if $a_2 \geq 0$,

$$g(a) = \begin{cases} i_a + 1, & \text{if } d_a < 1, \\ i_a + d_a, & \text{if } 1 \leq d_a \leq 2, \\ i_a + 2, & \text{if } d_a > 2, \end{cases}$$

and by symmetry

$$g(a_1, a_2) = g(a_1, -a_2).$$

It is easy to see that $g: K \to E$ is quasiconvex, and the verification that g is Lipschitz with constant 1 is straightforward. Furthermore, this function possesses no quasiconvex extension to E^2, since for any $m \in \mathbb{N}$ the level sets S_m cannot be convexly increased: for example, at the points $(1, 3(m-1))$, the only line of support to S_m is vertical! It will follow from Theorem 2.4

below, by taking the restriction of g to $[-1, 1]\times[0, 6]$, that g then cannot be blsd on K.

When f is one-dimensional, however, it is easy to see that Theorem 2.3 is true for any $K \subset E$.

Theorem 2.4. Let $f: K \subset E^n \to E$ be bounded above. Then, f is blsd on K if and only if there exists a quasiconvex Lipschitz function $g: E^n \to E$ extending f.

Proof. If such a g exists, it is blsd on E^n by Theorem 2.3; hence, any restriction of g is also blsd on its domain.

Inversely, let f be blsd on K with blsd-bound N, such that for all $x \in K$, $f(x) \leq M$. For each $x \in K$ and $x^0 \in \partial^- f(x)$, with $\|x^0\| \leq N$, define the function

$$g_{x,x^0}: E^n \to E: y \mapsto \min\{f(x), f(x) + \langle y - x, x^0\rangle\}.$$

Then, g_{x,x^0} is quasiconvex and bounded above by $f(x)$.

Since f is bounded above on K, the function

$$g: E^n \to E: y \mapsto \sup\{g_{x,x^0}(y) \mid x \in K, x^0 \in \partial^-(f(x), \|x^0\| \leq N\}$$

is defined everywhere. One easily sees that g extends f. Furthermore g, being a pointwise supremum of quasiconvex functions, is quasiconvex. In order to terminate the proof, it only remains to be shown that g is Lipschitz.

Consider any $y, z \in E^n$, with $g(y) < g(z)$. Let us first suppose that, for some $x \in K$, we have $g(z) = f(x)$. Then, there exists a $x^0 \in \partial^- f(x)$, with $\|x^0\| \leq N$, such that

$$\langle z - x, x^0\rangle \geq 0;$$

and, since $g(y) < f(x)$,

$$\langle y - x, x^0\rangle < 0.$$

By definition of g,

$$g(y) \geq f(x) + \langle y - x, x^0\rangle.$$

Hence,

$$g(z) - g(y) = f(x) - g(y) \leq \langle x - y, x^0\rangle.$$

Let now t be the orthogonal projection of y on the hyperplane H_x with equation

$$\langle u - x, x^0\rangle = 0;$$

and let s be the point of intersection of H_x with the line segment joining z and y. We then have

$$\langle x-y, x^0\rangle = \langle t-y, x^0\rangle = \|t-y\| \cdot \|x^0\|,$$

$$\|t-y\| \leq \|s-y\| \leq \|z-y\|.$$

Hence,

$$g(z)-g(y) \leq \langle x-y, x^0\rangle = \|t-y\| \cdot \|x^0\| \leq N \cdot \|z-y\|.$$

In the case where $g(z) \neq f(x)$ for all $x \in K$, then, for every $\epsilon > 0$, there exists $x \in K$ and $x^0 \in \partial^- f(x)$, with $\|x^0\| \leq N$, such that

$$0 \leq g(z) - (f(x) + \langle z-x, x^0\rangle) < \epsilon.$$

However,

$$g(y) \geq f(x) + \langle y-x, x^0\rangle,$$

and thus

$$\begin{aligned} g(z)-g(y) &< f(x)+\langle z-x, x^0\rangle + \epsilon - f(x) - \langle y-x, x^0\rangle \\ &= \langle z-y, x^0\rangle + \epsilon \\ &\leq \|z-y\| \cdot \|x^0\| + \epsilon \\ &\leq N\|z-y\| + \epsilon. \end{aligned}$$

Since ϵ is arbitrary, we must have

$$g(z)-g(y) \leq N\|z-y\|,$$

and g is Lipschitz with constant N. □

By continuity, this result applies to any blsd function with compact domain. It remains an open problem whether the theorem is valid for unbounded blsd functions on unbounded domains.

3. Properties and Examples

Theorem 3.1. The lower subdifferential $\partial^- f(a)$ of f at a is a closed convex set. For any $\lambda \geq 1$ and any $a^0 \in \partial^- f(a)$, one has a $\lambda a^0 \in \partial^- f(a)$. $0 \in \partial^- f(a)$ if and only if a is a global minimum of f and then $\partial^- f(a) = E^n$.

Proof. The proof is straightforward. □

Since any lsd function f is quasiconvex, any strict local minimum of f is a global minimum (see Ponstein, Ref. 8). In order to obtain equivalence of local and global minima, the lsd property has to be strengthened to strict lsd, where (1) holds for all y with $f(y) \leq f(x)$.

Theorem 3.2. If for each $i \in I$ the function f_i is lsd on K, and if for any $x \in K$ the supremum $g(x)$ of $f_i(x)$ over I is reached, then g is lsd on K. If all f_i are blsd with blsd-bound L_i, and if the L_i are uniformly bounded above by L, then g is blsd with blsd-bound L.

Proof. The proof is straightforward but requires explicitly that each supremum is reached. That this hypothesis is necessary is shown by the following counterexample. □

Let $I = \mathbb{N}_0$, the set of positive natural numbers, and define $f_i : E \to E$ by

$$f_i(x) = \begin{cases} i^{-1} - \left(\left(\dfrac{i+1}{i}\right)^2 - x^2\right)^{1/2}, & x^2 \leq (i+2)/i, \\ 0, & \text{elsewhere.} \end{cases}$$

All these f_i are blsd on E. Their supremum is, however, the example cited above of a quasiconvex function which is not lsd.

The following counterexample shows that all hypotheses of the second part are mandatory. Let $I = E_0^+$, the set of positive real numbers, and define $f_i : E^+ \to E$, where E^+ denotes the nonnegative reals, by

$$f_i(x) = \begin{cases} i^{-1/2} x, & x \leq i, \\ i^{1/2}, & x > i. \end{cases}$$

Every f_i is blsd on E^+ with blsd-bound $i^{-1/2}$. However,

$$\max\{f_i(x) \mid i \in I\} = x^{1/2},$$

and this function is not blsd.

Theorem 3.3. If f is lsd on K and g is lsd and nondecreasing on $f(K) \subset E$, then the composite function $g \circ f$ is lsd on K and

$$\partial^-(g \circ f)(x) \supset \{\lambda^0 x^0 \mid \lambda^0 \in \partial^- g(f(x)),\, x^0 \in \partial^- f(x)\}.$$

If they are both blsd, then $g \circ f$ is blsd.

Proof. Let x be a point in K and $x^0 \in \partial^- f(x)$. If $f(x)$ is a minimum of g, then x is also a minimum of $g \circ f$; thus, for any $\lambda^0 \in \partial^- g(f(x)) = E^n$, we have

$$\lambda^0 x^0 \in \partial^-(g \circ f)(x) = E^n.$$

Otherwise, let $\lambda^0 \in \partial^- g(f(x))$ and $y \in E$ be such that

$$g(y) < g(f(x)).$$

But then, we have

$$\lambda^0(y - f(x)) \leq g(y) - g(f(x)) < 0.$$

Since g is nondecreasing, $y < f(x)$, which shows that $\lambda^0 > 0$.

Now, let $z \in K$ satisfy

$$g \circ f(z) < g \circ f(x).$$

Since g is nondecreasing, this implies

$$f(z) < f(x),$$

and we have both

$$\lambda^0(f(z) - f(x)) + g(f(x)) \leq g(f(z))$$

and

$$\langle z - x, x^0 \rangle \leq f(z) - f(x).$$

Since $\lambda^0 > 0$, it follows that

$$\langle z - x, \lambda^0 x^0 \rangle + g(f(x)) \leq \lambda^0(f(z) - f(x)) + g(f(x)) \leq g(f(z)),$$

showing that $\lambda^0 x^0 \in \partial^-(g \circ f)(x)$.

The second part is now straightforward. □

In order to simplify the statement of some of the next theorems, we introduce the following notations.

Suppose $\lambda = (\lambda^1, \ldots, \lambda^k) \in E^k$ and, for each $i = 1, \ldots, k$, $x^i \in E^{n_i}$. Then, $X = (x^1, \ldots, x^k) \in E^n$, with $n = \sum_{i=1}^k n_i$. We denote by $[\lambda, X]$ the vector $(\lambda^1 x^1, \ldots, \lambda^k x^k)$ of E^n. If $A \subset E^k$ and $B \subset E^n$, then

$$[A, B] = \{[\lambda, X] \mid \lambda \in E^k, X \in E^n\}.$$

Theorem 3.4. If $h: K \subset E^k \to E$ is lsd and nondecreasing and, for each $i = 1, \ldots, k$, $f_i: K_i \subset En_i \to E$ is convex and subdifferentiable, and $F: \prod_{i=1}^k K_i \to E^k$ denotes the set theoretical product $\prod_{i=1}^k f_i$, then $\Phi = h \circ F$ is lsd on

$$L = F^{-1}\left[\prod_{i=1}^{k} f_i(K_i) \cap K\right]$$

and, for $X \in L$, one has

$$\partial^{-}\Phi(X) \supset [\partial^{-}h(F(X)), \prod_{i=1}^{k} \partial f_i(x^i)].$$

Moreover, if h is blsd and all f_i are convex and blsd, then Φ is blsd.

Proof. For any $X \in L$, we have $f_i(x^i) \in K_i$ and $F(X) \in K$. Hence, there exists $\lambda^0 \in \partial^{-}h(F(X))$ and $x_i^0 \in \partial f_i(x^i)$, $i = 1, \ldots, k$. Since h is nondecreasing, an analogous argument as was used in the foregoing proof shows that $\lambda_i^0 \geq 0$, $i = 1, \ldots, k$.

Let now $Z \in L$, with $\Phi(Z) < \Phi(X)$. Then, since $x_i^0 \in \partial f_i(x^i)$ and $\lambda_i^0 \geq 0$, and since $\lambda^0 \in \partial^{-}h(F(X))$ and $h(F(Z)) < h(F(X))$, we have

$$\begin{aligned}
\langle Z, [\lambda^0, X^0]\rangle &= \sum_{i=1}^{k} \lambda_i^0 \langle z^i, x_i^0\rangle \\
&\leq \sum_{i=1}^{k} \lambda_i^0(\langle x^i, x_i^0\rangle + f_i(z^i) - f_i(x^i)) \\
&= \langle X, [\lambda^0, X^0]\rangle + \langle F(Z) - F(X), \lambda^0\rangle \\
&\leq \langle X, [\lambda^0, X^0]\rangle + h(F(Z)) - h(F(X)) \\
&= \langle X, [\lambda^0, X^0]\rangle + \Phi(Z) - \Phi(X).
\end{aligned}$$

Hence, $[\lambda^0, X^0] \in \partial^{-}\Phi(X)$.

The second part is immediate. □

Theorem 3.5. If $f: K \subset E^n \to E$ is lsd and $T: E^m \to E^n$ is linear, then $f \circ T: T^{-1}(K) \subset E^m \to E$ is lsd; and we have, for any x with $T(x) \in K$,

$$\partial^{-}(f \circ T)(x) \supset T^*(\partial^{-}f(T(x))),$$

where T^* denotes the dual linear function of T. If f is blsd with blsd-bound N, then $f \circ T$ is blsd with bound $N \cdot \|T\|$.

Proof. Let $x \in E^m$, with $T(x) \in K$, and suppose $x^0 \in \partial^{-}f(T(x))$. For any $y \in T^{-1}(K)$, with $f(T(y)) < f(T(x))$, we have, on account of the definition of T^*, that

$$\begin{aligned}
\langle T^*(x^0), y - x\rangle &= \langle x^0, T(y - x)\rangle \\
&= \langle x^0, T(y) - T(x)\rangle \\
&\leq f(T(y)) - f(T(x)),
\end{aligned}$$

showing that $T^*(x^0) \in \partial^-(f \circ T)(x)$. If $\|x^0\| \le N$, then

$$\|T^*(x^0)\| \le \|T^*\| \cdot \|x^0\| \le \|T\| \cdot N.$$

For the properties of T^* and $\|T\|$, see Yosida, Ref. 9. □

Corollary 3.1. If $h: K \subset E^k \to E$ is lsd and nondecreasing and if, for each $i = 1, \ldots, k$, $f_i: K_i \subset E^n \to E$ is convex and subdifferentiable, then the function defined by $G(x) = h(f_1(x), \ldots, f_k(x))$ is lsd on

$$M = \left\{ x \in \bigcap_{i=1}^{k} K_i \,|\, (f_1(x), \ldots, f_k(x)) \in K \right\}.$$

If $x \in M$, $x_i^0 \in \partial^- f_i(x)$, for $i = 1, \ldots, k$, and $\lambda^0 \in \partial^- h(f_1(x), \ldots, f_k(x))$, we have

$$\sum_{i=1}^{k} \lambda_i^0 x_i^0 \in \partial^- G(x).$$

If, furthermore, h and all f_i are blsd, then G is blsd.

Proof. We have

$$G = h \circ F \circ \Lambda$$

in the following diagram:

$$E^n \xrightarrow{\Lambda} (E^n)^k \xrightarrow{F} E^k \xrightarrow{h} E$$

$$x \mapsto (x, \ldots, x) \mapsto (f_1(x), \ldots, f_k(x)) \mapsto G(x);$$

or, using the notations of Theorem 3.4,

$$G = \Phi \circ \Lambda,$$

where Φ is lsd. Λ being linear, Theorem 3.5 applies, showing that $\partial^- G(x) \supset \Lambda^*(\partial^- \Phi(x))$. One easily sees that Λ^* is nothing but the sum operator

$$\sum : (E^n)^k \to E^n : (y_1, \ldots, y_k) \mapsto \sum_{i=1}^{k} y_i.$$

Combining with the results of Theorem 3.4, one obtains the desired result. □

As with general quasiconvex functions, the class of lsd functions is not closed under addition (see, e.g., Greenberg and Pierskalla, Ref. 10).

The following examples show that no general relation seems to exist between pseudoconvexity and lsd. The function x^3 is lsd on every interval that is bounded below. If the interval contains 0, then the function is not pseudoconvex (Mangasarian, Ref. 7). The function $x+x^3$ is pseudoconvex on E but nowhere lsd on E. However, any function defined on E^n which is Lipshitz and pseudoconvex is blsd, since, for any x, we then have

$$N \cdot \nabla f(x)/\|\nabla f(x)\| \in \partial^- f(x).$$

Every convex function is lsd on its domain, since $\partial f(x) \subset \partial^- f(x)$. In order to be blsd, it is sufficient that subgradients of bounded norm exist everywhere on the domain. Hence, by Rockafellar (Ref. 3, p. 237), every convex function is blsd on any compact subset of its domain.

Let K by any interval of E with minimum value m, and let $f: K \to E$ be concave and nondecreasing. Then, f is lsd on K, since, for any $a \in K$,

$$(f(m)-f(a))/(m-a) \in \partial^- f(a) \quad \text{and} \quad \partial^- f(m) = E.$$

If the one-sided derivative of f at m exists ($<+\infty$), then f is blsd on K.

An application of this result is as follows. Suppose $f: K \subset E^n \to E$ is r-convex (Avriel, Ref. 11) and bounded below on the compact K by the value m; then, f is blsd on K. Indeed, f is r-convex if and only if $\exp(rf)$ is convex (see Avriel, Ref. 11) or if $f = r^{-1}\log(g)$, where g is convex. By the foregoing, both log and g are blsd on their respective domains. It follows from Theorem 3.3 that f is blsd on K. In order to calculate lower subgradients of f, it is, however, necessary to know a value for m.

4. Minimizing a BLSD Function under Linear Constraints

Let f be any real-valued function defined on E^n that is blsd on K, a compact polynomial set in E^n. We develop an algorithm that constructs a sequence of points in K such that any accumulation point is an optimal solution to the problem

$$\min\{f(x) \mid x \in K\}. \tag{2}$$

Our algorithm is of the general cutting plane type, as defined by Eaves and Zangwill (Ref. 12). Other well-known cutting plane algorithms, those of Kelley (Ref. 1), Cheney and Goldstein (Ref. 2), and Veinott (Ref. 13) fail to solve our problem, since they require the objective function to be convex. Indeed, the problem of minimizing f under convex or quasiconvex constraints is transformed into the problem of minimizing t under the same constraints and the additional constraint

$$f(x) - t \leq 0.$$

This last constraint defines a convex set if and only if f is convex. This shows that the device cannot be applied to our problem, since the objective may be nonconvex. The relation between our cutting plane method and Kelley's method will be discussed further at the end of the present section.

The algorithm is as follows. Let N be a blsd-bound of f on K. Let x_0 be any point in K. Choose any $x_0^0 \in \partial^- f(x_0)$, with $\|x_0^0\| \leq N$, and solve the following linear program for $j=0$:

$$(P_{j+1}) \qquad \min t,$$
$$\text{s.t.} \qquad t \geq \langle x - x_k, x_k^0 \rangle + f(x_k), \qquad k = 0, \ldots, j,$$
$$x \in K.$$

The compactness of K ensures the existence of an optimal solution (t_{j+1}, x_{j+1}), since (P_{j+1}) is equivalent to the minimization of the continuous function

$$\max\{\langle x - x_k, x_k^0 \rangle + f(x_k) \mid k = 0, \ldots, j\}.$$

Choose again a $x_{j+1}^0 \in \partial^- f(x_{j+1})$, with $\|x_{j+1}^0\| \leq N$, and solve (P_{j+1}) after increasing j by one.

Each program (P_{j+1}) is identical to (P_j), except for one additional constraint: the cut which cuts off the point (t_j, x_j), if x_j was not optimal. Thus, the sequence $(t_j)_{j>0}$ is nondecreasing; and, by the definition of lower subgradients and the existence of a minimum of f on K (f is continuous on K), it is easy to see that every t_j is a *lower bound* on the minimal function value of f on K. Furthermore, all generated points x_j are in K; thus, the sequence $(x_j)_{j>0}$ possesses accumulation points in K.

Theorem 4.1. Every accumulation point of the sequence $(x_j)_{j>0}$ minimizes f on K.

Proof. Suppose $x \in K$ does not minimize f on K. Let M denote the minimal value of f on K. Then, there is an $\epsilon > 0$ such that

$$f(x) - 2\epsilon > M.$$

By the continuity of f on K, there exists a ball with center x and radius δ such that, for any $y \in K$ in this ball,

$$f(y) - 2\epsilon > M$$

holds. Let further N denote a blsd-bound of f, and define

$$\eta = \min\{\epsilon/N, \delta\}.$$

We will show that either no iterate x_j or exactly one x_j is within distance η of x, which will terminate the proof.

If there is a k such that $\|x_k - x\| < \eta$, then, for any $y \in K$ with $\|y - x\| < \eta$, we have

$$|\langle y - x_k, x_k^0 \rangle| \leq \|y - x_k\| \cdot \|x_k^0\| \leq 2\eta N \leq 2\epsilon,$$

and thus also

$$f(x_k) + \langle y - x_k, x_k^0 \rangle \geq f(x_k) - 2\epsilon > M.$$

For any $m > k$, however, the following holds:

$$M \geq t_m \geq \langle x_m - x_k, x_k^0 \rangle + f(x_k),$$

which shows that

$$\|x_m - x\| \geq \eta.$$

Hence, if an iterate x_k is within distance η of x, then no subsequent iterate is. □

Theorem 4.2. The sequence $(t_j)_{j>0}$ converges to the minimal value M of f on K.

Proof. Let $(x_{k_s})_s$ be any convergent subsequence of $(x_k)_{k>0}$. This is a Cauchy sequence; and, for any positive value ϵ, we can choose $m > n$ such that

$$\|x_{k_m} - x_{k_n}\| < \epsilon / N,$$

where N is the blsd-bound of f. Now, we have

$$M \geq t_{k_m} \geq \langle x_{k_m} - x_{k_n}, x_{k_n}^0 \rangle + f(x_{k_n}),$$

and so

$$\begin{aligned} 0 \leq M - t_{k_m} &\leq \langle x_{k_n} - x_{k_m}, x_{k_n}^0 \rangle + M - f(x_{k_n}) \\ &\leq \|x_{k_n} - x_{k_m}\| \cdot \|x_{k_n}^0\| \leq \epsilon. \end{aligned}$$

Since $(t_j)_{j\geq 0}$ is a nondecreasing sequence, the result follows. □

The convergence proof requires explicitly that all previously generated constraints should be retained, as is the case with all aforementioned cutting plane methods. The question whether old cuts can be dropped without losing convergence was studied by Eaves and Zangwill (Ref. 12) and Topkis (Refs. 14 and 15). The techniques proposed in the first paper for dropping some of the old cuts can be applied to our algorithm. But the strong properties needed in order to permit dropping all inactive cuts are not verified here in general, although for one-dimensional functions this may always be done. In Ref. 16, we discuss another technique for testing which inactive constraints may be dropped. This yields a substantial reduction of the number of cuts that have to be retained in order to obtain the same sequence of approximations to the optimal solution.

In comparison to gradient search methods and their generalizations, the cutting plane method seems to involve a greater computational effort due to the optimization step of the successive linear programs. However, this step is effectively carried out by way of the dual simplex algorithm (Dantzig, Ref. 17), and often only one pivoting will suffice. It must also be observed that many recently proposed algorithms for convex programming require differentiability, often of second order, and include the solution of a linear program at each iteration (see, e.g., Bazaraa and Goode, Ref. 18). Furthermore, two compensations are obtained in the cutting plane algorithm for the increase in the number of calculations.

A first compensation stems from the automatic generation of the lower bounds t_k. These permit the implementation of good *stopping rules*, e.g., if a function value is found that is within a prespecified tolerance of the lower bound. We advocate the use of a marginal precision stopping rule such as

$$F - t_k \leq \epsilon \cdot t_k,$$

where F denotes the best function value found during the iterations.

The second compensation is obtained by the possibility to carry out a *post-optimal analysis* with respect to the near-optimal region. Let T denote the highest lower bound found, and define

$$h_\epsilon(x) = \max\{f(x_k) + \langle x - x_k, x_k^0\rangle \mid f(x_k) > T + \epsilon\}, \qquad \text{for each } x \in K.$$

The intersection H_ϵ of the epigraph of h_ϵ with the hyperplane $t = T + \epsilon$ is defined by linear constraints. The extreme points of this convex polytope are then easily calculated by a simplex technique, such as in Dyer and Proll (Ref. 19). This set H_ϵ is an outer approximation of the set

$$\min_\epsilon f = \{x \in K \mid f(x) \leq T + \epsilon\}.$$

The maximal value of f on the set of extreme points of H_ϵ, as compared to $T + \epsilon$, may serve to evaluate how well H_ϵ approximates $\min_\epsilon f$. When not satisfactory, this approximation may be improved by the addition of new cutting planes at the extreme points of H_ϵ, followed by a restart of the construction of a new H'_ϵ.

When the objective f is convex, one may use subgradients as lower subgradients. In this case, our algorithm is equivalent to the Kelley method (Ref. 1) in the case of linear constraints, after masking the objective as a nonlinear constraint (see the introduction to this section). This latter version of the Kelley method is the one studied by Wolfe (Ref. 20) in order to investigate convergence rates. Wolfe's results indicate linear convergence at best. Possibly, this could be improved by some interpolatory step, as suggested by Wolfe. It may be observed here that the rate of increase of the sequence of lower bounds is improved by choosing, at each x_k, a lower subgradient of smallest possible norm.

5. Application

Let $f_1, \ldots, f_r$ be *convex differentiable functions* defined on the compact polyhedral subset K of E^n. Let $g_1, \ldots, g_r$ be *concave and nondecreasing functions* on $f_1(K), \ldots, f_r(K)$ respectively, which are nowhere vertical. Consider the problem of minimizing on K the function

$$f: K \to E: x \mapsto \max\{g_1(f_1(a)), \ldots, g_r(f_r(x))\}.$$

From the results of Section 3, it follows that f is blsd on K. For each $x \in K$, a lower subgradient to f can be calculated as follows. Denote by i_x an index such that

$$f(x) = g_{i_x}(f_{i_x}(x))$$

and denote by m_i a lower bound of f_i on K_i. Then, an $x^0 \in \partial^- f(x)$ is given by

$$\frac{g_{i_x}(m_{i_x}) - f(x)}{m_{i_x} - f_{i_x}(x)} \cdot \nabla f_{i_x}(x), \qquad \text{if } f(x) \neq m_{i_x}.$$

In the other case, when

$$f(x) = m_{i_x},$$

x is optimal.

Using these lower subgradients, the cutting plane algorithm easily applies. This can be used to solve linearly constrained minimax facility location problems with concave cost functions and mixed norms. This extends the range of the problems solved by Jacobsen (Ref. 21). Some computational results in this latter context have been obtained by the author and are described in Ref. 22.

6. Concluding Remarks

In this paper, we have extended the class of convex functions to the class of boundedly lower subdifferentiable functions as application range of a cutting plane method for minimization under linear constraints. The extension of our cutting plane method to problems involving nonlinear constraints, as in the Kelley method (Ref. 1), is currently under consideration. This is, however, not a trivial matter.

Indeed, when the nonlinear constraints are locally linearized by the construction of supplementary cutting planes, the optimal solution x_k to the successive linear subproblems will mostly be unfeasible. A lower subgradient at x_k to the objective may then fail to yield a useful cutting plane if the value $f(x_k)$ remains below the constrained optimal value. Therefore, one must first construct from x_k a feasible point y_k and make use only of a lower subgradient of the objective at y_k.

Such an extended method for nonlinearly constrained problems will be described in a forthcoming paper.

References

1. KELLEY, J. E., *The Cutting Plane Method for Solving Convex Programs*, SIAM Journal on Applied Mathematics, Vol. 8, pp. 703–712, 1960.
2. CHENEY, E. W., and GOLDSTEIN, A. A., *Newton's Method of Convex Programming and Tchebycheff Approximation*, Numerische Mathematik, Vol. 1, pp. 253–268, 1959.
3. ROCKAFELLAR, R. T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
4. HIRIART-URRUTY, J. B., *ϵ-Subdifferential Calculus*, Convex Analysis and Optimization, Edited by J. P. Aubin and R. B. Vinter, Pitman, Boston, Massachusetts, pp. 147–152, 1982.
5. CLARKE, F. H., *Generalized Gradients and Applications*, Transactions of the American Mathematical Society, Vol. 205, pp. 247–262, 1975.
6. ROCKAFELLAR, R. T., *The Theory of Subgradients and Its Applications to Problems of Optimization: Convex and Nonconvex Functions*, Helderman Verlag, Berlin, Germany, 1981.
7. MANGASARIAN, O. L., *Nonlinear Programming*, McGraw-Hill, New York, New York, 1969.
8. PONSTEIN, J., *Seven Kinds of Convexity*, SIAM Review, Vol. 9, pp. 115–119, 1967.
9. YOSIDA, K., *Functional Analysis*, Springer-Verlag, Berlin, Germany, 1978.
10. GREENBERG, H. J., and PIERSKALLA, W. P., *A Review of Quasiconvex Functions*, Operations Research, Vol. 19, pp. 1553–1570, 1971.
11. AVRIEL, M., *R-Convex Functions*, Mathematical Programming, Vol. 2, pp. 309–323, 1972.
12. EAVES, B. C., and ZANGWILL, W. I., *Generalized Cutting Plane Algorithms*, SIAM Journal on Control, Vol. 9, pp. 529–542, 1971.
13. VEINOTT, A. F., JR., *The Supporting Hyperplane Method for Unimodal Programming*, Operations Research, Vol. 15, pp. 147–152, 1967.
14. TOPKIS, D. M., *Cutting Plane Methods without Nested Constraint Sets*, Operations Research, Vol. 18, pp. 404–413, 1970.

15. TOPKIS, D. M., *A Note on Cutting Plane Methods without Nested Constraint Sets*, Operations Research, Vol. 18, pp. 1216-1220, 1970.
16. PLASTRIA, F., *Testing Whether a Cutting Plane May Be Dropped*, Revue Belge de Statistique, d'Informatique, et de Recherche Opérationelle, Vol. 22, pp. 11-21, 1982.
17. DANTZIG, G. B., *Linear Programming and Extensions*, Princeton University Press, Princeton, New Jersey, 1963.
18. BAZARAA, S. M., and GOODE, J. J., *An Algorithm for Solving Linearly Constrained Minimax Problems*, European Journal of Operational Research, Vol. 11, pp. 158-166, 1982.
19. DYER, M. E., and PROLL, L. G., *An Improved Vertex Enumeration Algorithm*, European Journal of Operational Research, Vol. 9, pp. 359-368, 1982.
20. WOLFE, P., *Convergence Theory in Nonlinear Programming*, Integer and Nonlinear Programming, Edited by J. Abadie, North-Holland Publishing Company, Amsterdam, Netherlands, pp. 1-36, 1970.
21. JACOBSEN, S. K., *An Algorithm for the Minimax Weber Problem*, European Journal of Operational Research, Vol. 6, pp. 144-148, 1981.
22. PLASTRIA, F., *Continuous Location Problems Solved by Cutting Planes, I: Unconstrained Single Facility Location*, Vrije Universititeit Brussel, Report CSOOTW No. 177, 1982.