SIAM J. CONTROL
Vol. 11, No. 4, November 1973

A DESCENT NUMERICAL METHOD FOR OPTIMIZATION PROBLEMS WITH NONDIFFERENTIABLE COST FUNCTIONALS*

DIMITRI P. BERTSEKAS† AND SANJOY K. MITTER‡

Abstract. In this paper we consider the numerical solution of convex optimization problems with nondifferentiable cost functionals. We propose a new algorithm, the ε-subradient method, a large step, double iterative algorithm which converges rapidly under very general assumptions. We discuss the application of the algorithm in some problems of nonlinear programming and optimal control and we show that the ε-subgradient method contains as a special case a minimax algorithm due to Pshenichnyi [5].

1. General remarks. One of the most common approaches toward the numerical solution of optimization problems with or without constraints is the use of descent algorithms such as the steepest descent, conjugate gradient, quasi-Newton methods, methods of feasible directions, etc. These decent methods have enjoyed a great deal of popularity due to their reliability, simplicity, and good convergence properties. In their usual form all these algorithms require the existence of the gradient of the function to be minimized both for explicit use in the calculations and as a guarantee of their convergence to a local minimum. In many optimization problems, however, often arising in an economics framework, the natural cost functional of the problem turns out to be nondifferentiable. Such problems have received considerable attention recently and are the subject of this paper.

Early work on optimization problems with nondifferentiable cost functionals can be traced to the early sixties with the research of Dubovitskii and Milyutin [1], [2] which apparently served as a starting point for subsequent work of Soviet scientists [3]–[6]. At about the same time the theory of subdifferentiability of convex functions was developed by Moreau [7], [8], Rockafellar [9], [10], and Brøndsted and Rockafellar [11]. The notion of the subdifferential of a convex function (set of all supporting hyperplanes to the graph of the function) provided an efficient generalization of the notion of the ordinary gradient and formed the basis for the development of generalized necessary and sufficient conditions for optimality (see e.g. [10]). Necessary conditions which generalize the Pontryagin maximum principle of optimal control in very elegant form have been given by Neustadt [12], Heins and Mitter [13], and Rockafellar [14]. The latter reference contains also some generalizations of known results in the calculus of variations.

* Received by the editors March 27, 1972, and in revised form November 12, 1972.

† Department of Engineering-Economic Systems, Stanford University, Stanford, California 94305. The work of this author was supported by the National Science Foundation under Grant NSF-GR-29237.

‡ Decision and Control Sciences Group, Electronic Systems Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139. The work of this author was supported by the Air Force Office of Scientific Research under Grant AFOSR 70-1941 and by NASA NGL-22-009-124.

Further necessary conditions for optimal control problems with nondifferentiable cost functionals were given by Luenberger [15]. Some additional results along the same lines can be found in the thesis by Ghanem [16]. Luenberger's results were somewhat generalized for the case of discrete-time systems using subdifferential theory by the authors [17]. Questions related to stochastic optimization problems with nondifferentiable cost functionals have been examined in [35], [36]. Such problems occur often in stochastic programming. A method for approximating a nondifferentiable convex function by a smooth function was also given in reference [35]. Necessary conditions for optimality for nonlinear, nonconvex programming problems without differentiability were obtained by Bazaraa, Goode and Shetty [18], [19] and for minimax problems by Danskin, Dem'yanov and Pschenichyni [20], [21], [5]. Among existing nonlinear programming algorithms, the convex cutting plane algorithm [25], [37] can be used for the solution of convex nondifferentiable optimization problems.

In the area of descent numerical methods a minimization algorithm has been reported by Ermol'ev [22], [23] and credited to Shor [24]. This algorithm is applicable to unconstrained convex programming problems with nondifferentiable cost. It reportedly has slow convergence properties [33] although computational examples using the algorithm are not available in the English literature. A similar algorithm has been proposed by Polyak [33]. Decent algorithms for the solution of minimax problems have been given by Dem'yanov [21], Pshenichnyi [5], Birzak and Pshenichnyi [26], and Levitin [34]. It should be noted that many optimization problems with nondifferentiable cost functionals can be converted into minimax problems. The generalization of the steepest descent method for the numerical solution of optimization problems with nondifferentiable cost functions was given by Luenberger [15]; however, a proof of convergence of this algorithm is not presently available. The problem appears to be that the algorithmic map in this algorithm is not closed (using Zangwill's terminology [25]). The ε-subgradient method, first presented in [17], circumvents this closure problem as will be seen in what follows. Other papers related to optimization problems with nondifferentiable cost functionals include those of Polyak [38], Minch [39], Auslender [40], [41], and Butz [42].

In this paper we present a new descent algorithm for constrained or unconstrained minimization problems where the cost function is convex but not necessarily differentiable. This algorithm, the ε-subgradient method, is a large step, double iterative algorithm that converges rapidly under very general assumptions. The algorithm was first presented in [17] and is based on the notion of the ε-subgradient of a convex function. In § 2 we describe the algorithm and we prove its convergence. In § 3 we consider some practical aspects of the algorithm and we demonstrate by means of examples its application. Finally, in § 4 we delineate some classes of problems for which the ε-subgradient method compares favorably with existing algorithms. In addition we show that the ε-subgradient method contains as a special case a minimax algorithm due to Pshenichnyi [5].

2. The ε-subgradient method. In this section we describe a descent algorithm for the minimization of a convex function subject to convex constraints. Rather than considering explicitly the constraints, however, we shall allow the function to

be minimized to take the value $+\infty$. Thus the problem of finding the minimum of a function $g(\cdot)$ over a set X is equivalent to finding the minimum of the extended real-valued function $f(x) = g(x) + \delta(x|X)$, where $\delta(\cdot|X)$ is the indicator function of X, i.e., $\delta(x|X) = 0$ for $x \in X$, $\delta(x|X) = \infty$ for $x \notin X$. Stating the problem formally:

Find $\inf_x f(x)$ where $f: R^n \to (-\infty, +\infty]$ is a convex function which is lower semicontinuous with $\inf_x f(x) > -\infty$ and $f(x) < +\infty$ for at least one $x \in R^n$.

With the above assumptions, the function f is a closed proper convex function as defined in [10]. A detailed discussion of closed proper convex functions can be found in the same reference. A basic concept for the algorithm that we shall present is the notion of ε-subgradient. This notion was introduced in [9], [11] in connection with investigations related to the existence and characterization of subgradients of convex functions.

Let x be a point such that $f(x) < \infty$ and $\varepsilon > 0$ any positive scalar. A vector $x^* \in R^n$ is said to be an *ε-subgradient* of f at x if

$$f(z) \geqq f(x) - \varepsilon + \langle z - x, x^* \rangle \quad \text{for all } z \in R^n, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product in R^n. The set $\partial_\varepsilon f(x)$ of all ε-subgradients at x will be called the *ε-subdifferential* of f at x. This set is nonempty, closed and convex. It is evident that for $0 < \varepsilon_1 < \varepsilon_2$ we have

$$\partial f(x) \subset \partial_{\varepsilon_1} f(x) \subset \partial_{\varepsilon_2} f(x).$$

A useful characterization of the set $\partial_\varepsilon f(x)$ is given by the equation [10, p. 220]

$$\partial_\varepsilon f(x) = \{x^* | f^*(x^*) + f(x) - \langle x, x^* \rangle \leqq \varepsilon\}, \tag{2}$$

where

$$f^*(x^*) = \sup_x \{\langle x, x^* \rangle - f(x)\} \tag{3}$$

is the conjugate convex function of f [10]. The support function of $\partial_\varepsilon f(x)$ is given by the following useful equation [10, p. 220]:

$$\sigma[y|\partial_\varepsilon f(x)] = \sup_{x^* \in \partial_\varepsilon f(x)} \langle y, x^* \rangle = \inf_{\lambda > 0} \frac{f(x + \lambda y) - f(x) + \varepsilon}{\lambda}. \tag{4}$$

The set $\partial_\varepsilon f(x)$ has some interesting properties from the algorithmic point of view as shown by the following two propositions.

PROPOSITION 1. *Let x be a vector such that $f(x) < \infty$. Then*

$$0 \leqq f(x) - \inf_z f(z) \leqq \varepsilon \leftrightarrow 0 \in \partial_\varepsilon f(x).$$

Proof. By the definition (1),

$$0 \in \partial_\varepsilon f(x) \leftrightarrow f(z) \geqq f(x) - \varepsilon \quad \text{for all } z \in R^n,$$

which is equivalent to the desired relations. Q.E.D.

PROPOSITION 2. *Let x be a point such that $f(x) < \infty$ and $0 \notin \partial_\varepsilon f(x)$. Let y be any vector such that*

$$\sup_{x^* \in \partial_\varepsilon f(x)} \langle y, x^* \rangle < 0. \tag{5}$$

Then we have

$$f(x) - \inf_{\lambda \geqq 0} f(x + \lambda y) > \varepsilon. \tag{6}$$

Proof. Assume the contrary, i.e., $\inf_{\lambda \geqq 0} f(x + \lambda y) - f(x) + \varepsilon \geqq 0$. Then we have

$$\frac{f(x + \lambda y) - f(x) + \varepsilon}{\lambda} \geqq 0 \quad \text{for all } \lambda > 0.$$

This implies by using (4)

$$\sup_{x^* \in \partial_\varepsilon f(x)} \langle x^*, y \rangle = \inf_{\lambda > 0} \frac{f(x + \lambda y) - f(x) + \varepsilon}{\lambda} \geqq 0.$$

Since $\partial_\varepsilon f(x)$ is closed this implies that $0 \in \partial_\varepsilon f(x)$ which contradicts the hypothesis. Q.E.D.

In the case where $0 \notin \partial_\varepsilon f(x)$, a possible method for finding a vector $\bar{y}(x) \in R^n$ such that $\sup_{x^* \in \partial_\varepsilon f(x)} \langle x^*, \bar{y}(x) \rangle < 0$ is the following. Let $\| \cdot \|$ be the usual Euclidean norm in R^n and let $\bar{x}^*$ be the unique vector of minimum norm in $\partial_\varepsilon f(x)$. Then the vector

$$\bar{y}(x) = -(\bar{x}^* / \| \bar{x}^* \|) \tag{7}$$

satisfies $\sup_{x^* \in \partial_\varepsilon f(x)} \langle \bar{y}(x), x^* \rangle = - \| \bar{x}^* \| < 0$.

Propositions 1 and 2 form the basis for the algorithm that we shall present. The former provides a termination criterion for the algorithm. The latter states that whenever the value $f(x)$ exceeds the optimal value by more than ε, then by a descent along a vector y satisfying (5) we can decrease the value of the cost by at least ε. Consider the following algorithm.

SUBGRADIENT METHOD.

Step 1. Select a vector x_0 such that $f(x_0) < \infty$, a scalar $\varepsilon_0 > 0$ and a scalar a, $0 < a < 1$.

Step 2. Given x_n and $\varepsilon_n > 0$, set $\varepsilon_{n+1} = a^k \varepsilon_n$, where k is the smallest nonnegative integer such that $0 \notin \partial_{\varepsilon_{n+1}} f(x_n)$.

Step 3. Find a vector y_n such that

$$\sup_{x^* \in \partial_{\varepsilon_{n+1}} f(x_n)} \langle y_n, x^* \rangle < 0. \tag{8}$$

Step 4. Set $x_{n+1} = x_n + \lambda_n y_n$, where $\lambda_n > 0$ is such that

$$f(x_n) - f(x_{n+1}) > \varepsilon_{n+1}.$$

Return to Step 2.

It should be mentioned that if x_n is not a minimizing point of f there always exists a nonnegative integer k such that $0 \notin \partial_{a^k \varepsilon_n} f(x_n)$, since by Proposition 1 we

have

$$0 \notin \partial_{a^k\varepsilon_n} f(x_n) \leftrightarrow f(x_n) - \inf_x f(x) > a^k\varepsilon_n.$$

Also by Proposition 2 there exists a scalar λ_n such that

$$f(x_n) - f(x_n + \lambda_n x_n) > \varepsilon_{n+1}, \tag{9}$$

thus showing that Step 4 can always be carried out. In fact, one can show that the set of all scalars λ_n satisfying (9) is an open bounded interval or an open half-line. One way of finding a scalar λ_n satisfying (9) is by means of the one-dimensional minimization

$$f(x_n + \lambda_n y_n) = \min_{\lambda > 0} f(x_n + \lambda y_n),$$

assuming the minimum is attained. This in turn can be guaranteed whenever the set of minimizing points of f is nonempty and compact, since in this case all the level sets of f are compact [10, Cor. 8.7.1]. We note also that Steps 2 and 3 can be carried out by means of an auxiliary minimization problem as will be discussed in detail in the next section.

We now prove the convergence of the ε-subgradient method.

PROPOSITION 3. *Consider the vectors x_n generated by the ε-subgradient method. Then either $f(x_m) = \min_x f(x)$ for some $m \geqq 0$ or the generated infinite sequence $\{x_n\}$ satisfies*

(a) $\lim_{n\to\infty} f(x_n) = \inf_x f(x)$.

If, in addition, the set $M = \{\bar{x} | f(\bar{x}) = \min_x f(x)\}$ is nonempty and bounded, then:

(b) *Every convergent subsequence of $\{x_n\}$ has its limit in M, and at least one such subsequence exists.*

(c) *For every $\varepsilon > 0$ there exists an $m \geqq 0$ such that $x_n \in M + \varepsilon B$ for all $n \geqq m$, where $B = \{x | \|x\| \leqq 1\}$ is the unit ball in R^n.*

(d) *If the minimum of f is attained at a single point $\bar{x}$ then $\{x_n\} \to \bar{x}$.*

Proof. By Proposition 2 we have

$$f(x_n) - f(x_{n+1}) > \varepsilon_{n+1} \quad \text{for all } n \geqq 0$$

and hence,

$$f(x_0) - \sum_{i=1}^{n} \varepsilon_i > f(x_n) > \inf_x f(x) \quad \text{for all } n \geqq 1.$$

Since $\varepsilon_i > 0$ the above inequality implies $\{\varepsilon_i\} \to 0$. This implies that $\varepsilon_{i+1} < \varepsilon_i$ for an infinite number of integers i. In view of Step 2 of the algorithm we have for those integers: $0 < f(x_i) - \inf_x f(x) \leqq \varepsilon_i$. Since $\{f(x_n)\}$ is a decreasing sequence, it follows that $\lim_{n\to\infty} f(x_n) = \inf_x f(x)$, and (a) is proved. To prove (b) notice that $x_n \in F_0$, where $F_0 = \{x | f(x) \leqq f(x_0)\}$ and since M is nonempty and bounded, F_0 is compact (see [10, Cor. 8.7.1]). Therefore the sequence $\{x_n\}$ has at least one convergent subsequence. The fact that the limits of all convergent subsequences belong to M follows from (a) and Cor. 27.2.1 in [10]. Part (c) follows from (a) and Thm. 27.2 in [10]. Part (d) follows from (a) and Cor. 27.2.2 in [10]. Q.E.D.

The above proposition establishes that the ε-subgradient method has attractive convergence properties. In fact, it converges to the optimal value even if an

optimal solution does not exist. A further attractive feature of the method is that it guarantees substantial progress at every iteration (Step 4) and that the progress of the computation is monitored constantly via the parameter ε (Step 2). The price for this substantial progress is the computations necessary to find the direction of descent in Steps 2 and 3. In the next section we shall describe some practical aspects of the algorithm and demonstrate by means of examples its application.

3. Practical aspects of the ε-subgradient method. A cursory examination of the ε-subgradient method reveals that in fact the most difficult step in a single iteration is finding the direction of descent y_n. However, contrary to most descent algorithms, the chosen direction of descent in the ε-subgradient method can lead to guaranteed substantial reduction of the value of the cost functional in a single iteration. To demonstrate this fact consider the following lemma.

LEMMA. *Assume that the scalars ε_0 and a in the ε-subgradient method are such that*

$$f(x_0) - \inf_x f(x) \leqq \varepsilon_0, \qquad 1/2 \leqq a < 1. \tag{10}$$

Then for all $n \geqq 1$,

$$f(x_n) - \inf_x f(x) < ((1 - a)/a)\varepsilon_n \leqq (1 - a)\varepsilon_{n-1}. \tag{11}$$

Proof. We have $f(x_0) - \inf_x f(x) \leqq \varepsilon_0$ implying that $0 \in \partial_{\varepsilon_0}(x_0)$. Hence in Step 2 we have $\varepsilon_1 \neq \varepsilon_0$. This in turn implies that $0 \in \partial_{\varepsilon_1/a} f(x_0)$, or equivalently,

$$f(x_0) - \inf_x f(x) \leqq \varepsilon_1/a.$$

On the other hand,

$$f(x_0) - f(x_1) > \varepsilon_1.$$

Combining the last two inequalities we have

$$f(x_1) - \inf_x f(x) < ((1 - a)/a)\varepsilon_1,$$

proving (11) for $n = 1$. Since $1/2 \leqq a < 1$, the last inequality implies that $f(x_1) - \inf_x f(x) < \varepsilon_1$ and the same argument as above can be used to prove (11) for $n = 2$ and every n. Q.E.D.

It is evident now from (11) that a substantial reduction of the value of the cost functional is possible by choosing the value of the parameter a high enough. On the other hand, a value of the parameter a close to unity leads to an increased number of iterations in order to find the scalar ε_{n+1} from ε_n in Step 2 of the algorithm. Thus, in practice, one must settle on a compromise value for the parameter a depending on how difficult it is to carry out a single check $0 \in \partial_{a^k\varepsilon_n} f(x_n)$ in Step 2. Another possibility is to modify the algorithm so that the value of the parameter a is adjusted during the iterations in Step 2 on the basis of information already obtained. A number of convergent schemes are possible. We do not discuss these schemes since they are not theoretically interesting but rather relate to the intelligent programming of the method.

We now turn to the important question of how the calculation of the direction of descent is to be carried out once the value of the parameter a is selected. As

mentioned in the previous section it is possible to carry out Steps 2 and 3 of the algorithm by solving the following minimization problem:

$$\min_{x^* \in \partial_{a^k \varepsilon_n} f(x_n)} \|x^*\|. \tag{12}$$

Now clearly we have $0 \in \partial_{a^k \varepsilon_n} f(x_n)$ if and only if problem (12) has a zero optimal value and therefore Step 2 of the algorithm can be carried out by solving problem (12) successively for $k = 0, 1, \cdots$. There exists an integer k for which problem (12) has a nonzero optimal value. Let $\bar{x}^*$ be the optimal solution of problem (12) for the first such integer k. Then a suitable direction of descent y_n satisfying (8) in Step 3 of the algorithm is given by

$$y_n = -\bar{x}^*/\|\bar{x}^*\|. \tag{13}$$

One efficient method for solving the minimization problem (12) is to solve successively the unconstrained problem

$$\min_{x^*} \{\|x^*\|^2 + P_k(x^*)\}, \tag{14}$$

where $P_k(\cdot)$ is a (moderate) penalty function

$$\begin{aligned} &P_k(x^*) \geqq 0 \quad \text{for all } x^*, \\ &P_k(x^*) = 0 \quad \text{if and only if } x^* \in \partial_{a^k \varepsilon_n} f(x_n). \end{aligned} \tag{15}$$

It is clear that problem (14) has a zero optimal value if and only if problem (12) has a zero optimal value. Furthermore, when k is such that problem (12) has a nonzero value, problem (14) yields an approximate solution $\tilde{x}^*$ to problem (12). In this case one can either increase the penalty and obtain a more accurate solution or obtain an approximate direction of descent $\tilde{y}_n$ from

$$\tilde{y}_n = -\tilde{x}^*/\|\tilde{x}^*\|.$$

The approximate direction $\tilde{y}_n$ is considered acceptable if it yields a point x_{n+1} satisfying $f(x_n) - f(x_{n+1}) > \varepsilon_{n+1}$ in Step 4. If $\tilde{y}_n$ is not acceptable we increase the penalty in problem (14) and resolve the problem in order to obtain a more accurate direction of descent.

The preceding discussion clearly demonstrates that the application of the ε-subgradient method to a specific problem requires the solution of minimization problems of the form

$$\min_{x^* \in \partial_\varepsilon f(x)} \|x^*\|. \tag{16}$$

At first sight it would therefore appear that the ε-subgradient method can be applied only to the limited class of functions for which the ε-subdifferential $\partial_\varepsilon f(x)$ has a convenient characterization. We shall demonstrate in what follows in this section that this is not the case and, in fact, the method can be applied to most functions likely to be encountered in practice. This is due to the fact that problem (16) can be cast into the usual nonlinear programming framework even if a convenient closed form characterization of the set $\partial_\varepsilon f(x)$ is not available.

By making use of the characterization (2) of the ε-subdifferential $\partial_\varepsilon f(x)$ in terms of the conjugate convex function f^*, problem (16) can be written as

$$\text{minimize } \|x^*\| \tag{17}$$

subject to

$$f^*(x^*) + f(x) - \langle x, x^* \rangle \leqq \varepsilon.$$

Now there is a class of simple functions f for which the conjugate

$$f^*(x^*) = \sup_x \{\langle x, x^* \rangle - f(x)\}$$

has a convenient closed form. Such functions include:

(a) Positively homogeneous closed convex functions, i.e., support functions of given sets [10, § 13]. Thus if

$$f(x) = \sigma(x|X) = \sup_{x^* \in X} \langle x, x^* \rangle,$$

then

$$f^*(x^*) = \delta(x^*|\bar{X}) = \begin{cases} 0 & \text{if } x^* \in \bar{X}, \\ \infty & \text{if } x^* \notin \bar{X}, \end{cases}$$

where $\bar{X}$ is the closure of the convex hull of X. This class includes all norms and seminorms in R^n as well as linear functions. In addition, the conjugates of powers greater than one of norms and seminorms in R^n (including quadratic forms) are given in [10, § 15].

(b) Exponentials and logarithms of coordinates of x (see [10, § 12]).

(c) Indicator functions of affine sets (linear manifolds), convex cones and unit balls with respect to a norm or a seminorm [10, § 13].

(d) Indicator functions of sets with known support functions, [10, § 13]. If X is a closed convex set and

$$f(x) = \delta(x|X),$$

then

$$f^*(x^*) = \sigma(x^*|X) = \sup_{x \in X} \langle x, x^* \rangle.$$

We note that constraint sets which are characterized by their support function are encountered, for example, in some optimal control problems as will be discussed in some detail in § 4.

Now from this class of simple functions one can build more complicated functions by means of various operations such as summation, affine transformation, maximization, etc. The conjugates of such functions are characterized by the following well-known relations:

$$(f_1 + f_2 + \cdots + f_m)^*(x^*) = \min_{\sum_{i=1}^m x_i^* = x^*} \left\{ \sum_{i=1}^m f_i^*(x_i^*) \right\} \quad ([10, \text{Thm. } 16.4]), \tag{18}$$

where f_i, $i = 1, \cdots, m$, are closed proper convex functions with a common point in the relative interior of their effective domain, and the function $f_1 + \cdots + f_m$ is

defined by

$$(f_1 + f_2 + \cdots + f_m)(x) = f_1(x) + f_2(x) + \cdots + f_m(x),$$

$$(f \cdot A)^*(x^*) = \min_{A^* y^* = x^*} f^*(y^*), \tag{19}$$

where $f: R^m \to R_e$ is a closed proper convex function, A is a linear transformation from R^n to R^m, A^* denotes its adjoint, the function $f \cdot A$ is the composition of f and A, and, in addition, the range of A contains a point in the relative interior of the effective domain of f.

$$(\max\{f_1, \cdots, f_m\})^*(x^*) = \min_{\substack{x^* = \sum_{i=1}^m \lambda_i x_i^* \\ \lambda_i \geqq 0 \\ \sum_{i=1}^m \lambda_i = 1}} \left\{ \sum_{i=1}^m \lambda_i f_i^*(x_i^*) \right\} \quad \text{([10, Thm. 16.5])}, \tag{20}$$

where f_i, $i = 1, \cdots, m$, are convex real-valued functions and the function $\max\{f_1, \cdots, f_m\}$ is defined by

$$(\max\{f_1, \cdots, f_m\})(x) = \max\{f_1(x), \cdots, f_m(x)\},$$

$$g^*(x^*) = f^*(x^*) + \langle c, x^* \rangle, \tag{21}$$

where $g(x) = f(x - c)$, $f = R^n \to (-\infty, +\infty]$ is a closed proper convex function and $c \in R^n$ is a given vector.

The equations (18)–(21) can be used in order to put the minimization problem (17) in the standard nonlinear programming framework for a wide variety of functions. As an illustration, consider the case where the function f to be minimized by means of the ε-subgradient method has the form[9]

$$f(x) = f_1(x) + f_2(x) + \cdots + f_m(x).$$

By using (18) the optimization problem (17) can be written as

$$\text{minimize } \|x^*\|$$

subject to

$$\min_{\sum_{i=1}^m x_i^* = x^*} \left\{ \sum_{i=1}^m f_i^*(x_i^*) \right\} + f(x) - \langle x^*, x \rangle \leqq \varepsilon.$$

It can be easily seen that the above problem is equivalent to

$$\text{minimize } \left\| \sum_{i=1}^m x_i^* \right\|$$

subject to

$$\sum_{i=1}^m f_i^*(x_i^*) + f(x) - \sum_{i=1}^m \langle x_i^*, x \rangle \leqq \varepsilon.$$

This latter problem is in the standard nonlinear programming framework whenever the functions f_i belong to the class of simple functions mentioned earlier. As another example consider the case where the function f has the form

$$f(x) = \max\{f_1(A_1 x), \cdots, f_m(A_m x)\},$$

where $A_1, \cdots, A_m$ are linear transformations and $f_1, \cdots, f_m$ are real-valued convex functions. By using (19), (20), the optimization problem (17) for this function can be written as

$$\text{minimize } \|x^*\|$$

subject to

$$\min_{\substack{x^* = \sum_{i=1}^m \lambda_i x_i^* \\ \lambda_i \geqq 0 \\ \sum_{i=1}^m \lambda_i = 1}} \left\{ \sum_{i=1}^m \lambda_i \min_{A_i^* y_i^* = x_i^*} f_i^*(y_i^*) \right\} + f(x) - \langle x^*, x \rangle \leqq \varepsilon,$$

or equivalently,

$$\text{minimize } \left\| \sum_{i=1}^m \lambda_i A_i^* y_i^* \right\|$$

subject to

$$\sum_{i=1}^m \lambda_i f_i^*(y_i^*) + f(x) - \sum_{i=1}^m \lambda_i \langle A_i^* y_i^*, x \rangle \leqq \varepsilon,$$

$$\lambda_i \geqq 0, \quad \sum_{i=1}^m \lambda_i = 1.$$

Similarly, one can write the optimization problem (17) in standard form whenever the function to be minimized involves simultaneously sums, compositions with linear transformations and maxima of the basic simple functions referred to earlier. Thus the ε-subgradient method can be applied for the minimization of a wide class of functions. This class of functions can be further enlarged by making use of the following technique to eliminate some of the constraints of the minimization problem.

Consider the convex programming problem

(22) $$\text{minimize } f_0(x)$$

subject to

$$x \in X, \quad f_i(x) \leqq 0, \qquad i = 1, \cdots, m,$$

where $f_0, f_1, \cdots, f_m$ are real-valued convex functions and X is a closed convex set. Let $\bar{x}$ be an optimal solution of this problem and assume that there exists a point $\tilde{x} \in X$ such that $f_i(\tilde{x}) < 0$, $i = 1, \cdots, m$. Then there exist nonnegative Lagrange multipliers, $\lambda_1, \cdots, \lambda_m$, corresponding to $\bar{x}$ [25], [37] such that $\bar{x}$ minimizes

$$f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)$$

subject to $x \in X$. Furthermore, it is known [15] that if k is a scalar such that

(23) $$k > \max \{\lambda_1, \cdots, \lambda_m\}$$

then $\bar{x}$ is an optimal solution to the problem

$$\text{minimize } f_0(x) + k \sum_{i=1}^{m} \max [0, f_i(x)] \tag{24}$$

subject to $x \in X$.

Conversely, every optimal solution of problem (24) is an optimal solution of problem (22) so that the two problems are equivalent and either one of the two can be solved in place of the other. Concerning the selection of the scalar k, it can be easily proved that if μ is a strict lower bound for the optimal value of problem (22), then

$$k = \max\left\{\frac{f_0(\tilde{x}) - \mu}{-f_1(\tilde{x})}, \cdots, \frac{f_0(\tilde{x}) - \mu}{-f_m(\tilde{x})}\right\}$$

satisfies (23), where $\tilde{x}$ is a vector such that $\tilde{x} \in X$ and $f_i(\tilde{x}) < 0$, $i = 1, 2, \cdots, m$.

We shall close this section by showing explicitly the form of the auxiliary minimization problem (17) for a specific problem.

Example. Consider the problem

$$\text{minimize}\left\{\max_{\substack{y_i \geqq 0 \\ \|y\| \leqq 1}} \langle x, y\rangle + \max [0, \tfrac{1}{2}x'Qx + \langle c, x\rangle]\right\}$$

subject to $x \in X = \{x | x_i \geqq 0, i = 1, \cdots, m\}$.

In the above problem, x, y are vectors in R^n, $\|\cdot\|$ denotes the Euclidean norm in R^n, Q is a positive definite matrix and c is a given vector. By defining

$$f_1(x) = \max_{\substack{y_i \geqq 0 \\ \|y\| \leqq 1}} \langle x, y\rangle,$$

$$f_2(x) = (1/2)x'Qx + \langle c, x\rangle,$$

$$f_3(x) = \delta(x|X) = \begin{cases} 0 & \text{if } x \in X, \\ \infty & \text{if } x \notin X, \end{cases}$$

the problem is written

$$\text{minimize } f(x) = f_1(x) + \max [0, f_2(x)] + f_3(x).$$

The auxiliary optimization problem to be solved in Steps 2 and 3 of the ε-subgradient method is

$$\text{minimize } \|x^*\|$$

subject to $f^*(x^*) + f(x) - \langle x^*, x\rangle \leqq a^k\varepsilon$. By using (18) and (20) and the fact that the conjugate of the zero function is the function

$$(0)^*(x^*) = \begin{cases} 0 & \text{if } x^* = 0, \\ \infty & \text{if } x^* \neq 0, \end{cases}$$

the above problem is equivalent to

$$\text{minimize } \|x_1^* + \lambda x_2^* + x_3^*\|^2 \tag{25}$$

subject to

$$f_1^*(x_1^*) + \lambda f_2^*(x_2^*) + f_3^*(x_3^*) + f(x) - \langle x_1^* + \lambda x_2^* + x_3^*, x\rangle \leqq a^k\varepsilon, \qquad 0 \leqq \lambda \leqq 1.$$

We have

$$f_1^*(x_1^*) = \begin{cases} 0 & \text{if } x_1^* \geqq 0, \quad \|x_1^*\| \leqq 1, \\ \infty & \text{otherwise,} \end{cases}$$

$$f_2^*(x_2^*) = \tfrac{1}{2}(x_2^* - c)'Q^{-1}(x_2^* - c),$$

$$f_3^*(x_3^*) = \begin{cases} 0 & \text{if } x_3^* \leqq 0, \\ \infty & \text{otherwise,} \end{cases}$$

where the inequalities $x_1^* \geqq 0, x_3^* \leqq 0$ are interpreted to be componentwise. Hence problem (25) takes the form

$$\text{minimize } \|x_1^* + \lambda x_2^* + x_3^*\|^2$$

subject to

$$(\lambda/2)(x_2^* - c)'Q^{-1}(x_2^* - c) + f(x) - \langle x_1^* + \lambda x_2^* + x_3^*, x\rangle \leqq a^k\varepsilon,$$

$$0 \leqq x_1^*, \quad \|x_1^*\| \leqq 1, \quad x_3^* \leqq 0, \quad 0 \leqq \lambda \leqq 1,$$

a nonlinear program with linear and quadratic constraints. If $(\bar{x}_1^*, \bar{x}_2^*, \bar{x}_3^*, \bar{\lambda})$ is an optimal solution of the above problem then $\bar{x}^* = \bar{x}_1^* + \bar{\lambda}\bar{x}_2^* + \bar{x}_3^*$ is an optimal solution of the auxiliary optimization problem of Steps 2 and 3 of the ε-subgradient method.

4. Applications. In this section we attempt to delineate some classes of problems for which the ε-subgradient method compares favorably with existing methods. It is well known that many optimization problems with nondifferentiable cost functionals can be converted into nonlinear programming problems where all functions involved are differentiable. For example consider the problem

$$\text{minimize } \max\{f_1(x), \cdots, f_m(x)\}, \tag{26}$$

where the functions f_i are convex and differentiable. This problem is equivalent to the problem

$$\text{minimize } y \tag{27}$$

subject to

$$f_i(x) \leqq y, \qquad i = 1, \cdots, m,$$

where y is a scalar auxiliary variable. This latter problem can be solved by any of the existing algorithms for differentiable functions such as, for instance, the ε-perturbation feasible direction method [25]. Also problem (26) can be solved by using Dem'yanov's minimax algorithm [21] which is closely related to the feasible direction method mentioned above. It appears that either one of the two algorithms is preferable to the ε-subgradient method for the solution of problem (26). This is due to the considerable computation necessary in order to find the direction of descent in the ε-subgradient method. More generally, one can say that if the optimization problem can be converted to a nonlinear program where all functions

involved are differentiable, standard methods should, in most cases, be preferable over the ε-subgradient method.

The ε-subgradient method, however, should be considered advantageous when applied to problems which cannot be converted to nonlinear programming problems involving differentiable functions since it has the advantage of fast convergence. One class of such problems is characterized by the presence of terms of the form $\max_{y \in Y} \langle x, y \rangle$ either in the cost function or the constraints. The first known algorithm involving functions of the form $\max_{y \in Y} \langle x, y \rangle$ is the one of Pshenichnyi [5] who considered the problem

$$\text{minimize } \max_{y \in Y} \langle x, y \rangle \tag{28}$$

subject to

$$x \in A,$$

where Y is a convex compact set and A is a given hyperplane. When the ε-subgradient method is applied to problem (26), the direction of descent is determined by solving the auxiliary optimization problem

$$\text{minimize } \|x_1^* + x_2^*\|$$

subject to

$$x_1^* \in Y, \quad \max_{y \in Y} \langle x, y \rangle - \langle x_1^*, x \rangle \leqq \varepsilon,$$

$$x_2^* \in A^{\perp},$$

where $A^{\perp}$ is the one-dimensional subspace orthogonal to the hyperplane A. This is exactly the same optimization problem by means of which the direction of descent is determined in Pshenichnyi's method and thus the ε-subgradient method and Pshenichnyi's method are identical when applied to problem (28). The ε-subgradient method, however, can be applied to much more general problems involving terms of the form $\max_{y \in Y} \langle x, y \rangle$. One such example was given in the previous section. For such problems the ε-subgradient method compares favorably with, for example, Dem'yanov's minimax algorithm which involves comparable computations for finding the direction of descent but does not converge as fast as the ε-subgradient method.

The ε-subgradient method can also be used effectively for problems where some of the constraint sets are not given explicitly but instead can be specified from their support function. For such problems methods of feasible directions, for example, are not applicable. As an example, consider the following optimal control problem where some of the constraint sets are characterized as reachable sets of a differential system.

Consider the linear system

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \tag{29}$$

over the time interval $[t_0, T]$ which is controllable from t_0 to T and where $A(t)$ is a Lebesgue integrable $n \times n$ matrix, and $B(t)$ is a continuous $n \times m$ matrix function on $[t_0, T]$. The m-vector-valued function $u(t)$ is assumed to be measurable in $[t_0, T]$ and such that

$$u(t) \in U \quad \text{almost everywhere in } [t_0, T], \tag{30}$$

where U is a nonempty compact subset of R^n. Assume further that the initial condition is constrained to lie in X_0, a convex compact subset of R^n:

$$x(t_0) \in X_0. \tag{31}$$

Consider the problem of minimizing

$$J[x(t_0), u] = F[x(T)], \tag{32}$$

where F is a closed proper convex function in R^n subject to the constraints (29)–(31).

Then under our assumptions, for every pair $(x(t_0), u)$ satisfying (30) and (31), there exists a unique absolutely continuous solution of (29). The set $X(T)$ of reachable states $x(T)$ at time T corresponding to the constraints (30), (31) is convex and compact by a theorem of Neustadt [30], and its support function is given by ([31], [32])

$$\sigma[x^*|X(T)] = \sigma[\Phi'(t_0, T)x^*|X_0] + \int_{t_0}^{T} \sigma[B'(t)\Phi'(t, T)x^*|U]\, dt,$$

where $\Phi(t, \tau)$ is the unique absolutely continuous transition matrix corresponding to the matrix $A(t)$.

The problem can now be recast as one of minimizing the extended real-valued convex function

$$f[x(T)] = F[x(T)] + \delta[x(T)|X(T)]$$

and the ε-subgradient method can be used for its solution. The direction of descent is determined by solving the optimization problem

$$\text{minimize } \|x_1^* + x_2^*\|$$

subject to

$$F^*(x_1^*) + \sigma[x_2^*|X(T)] + F[x(T)] - \langle x_1^* + x_2^*, x(T)\rangle \leqq \varepsilon.$$

For the problem that we consider there is some difficulty associated with the one-dimensional line search in Step 4 of the ε-subgradient method since it is not easy to check feasibility of any given terminal state. This difficulty can be circumvented by finding a point along the direction of descent such that the value of the function F has decreased by ε or a little less. It can be easily seen that such a point is feasible and that the algorithm will still be convergent.

5. Conclusions. The ε-subgradient method is a descent algorithm which can solve efficiently some convex minimization problems with nondifferentiable cost functionals which cannot be solved by standard nonlinear programming methods. It converges fast under very general assumptions but requires the solution of an auxiliary optimization problem in order to determine the direction of descent at each iteration. Presently, we do not have any computational experience with the method. It is hoped that such computational experience will be gained in the near future.

REFERENCES

[1] A. Y. Dubovitskii and A. A. Milyutin, *Extremum problems in the presence of constraints*, Dokl. Akad. Nauk SSSR, 149 (1963), pp. 452–455; English transl., Soviet Math Dokl., 4 (1963), pp. 452–455.

[2] ———, *Extremum problems in the presence of constraints*, Zh. Vychisl. Mat. i Mat. Fiz., 5 (1965), pp. 395–453.

[3] V. F. Dem'yanov and A. M. Rubinov, *Minimization of functionals in normed spaces*, this Journal, 6 (1968), pp. 73–89.

[4] ———, *Approximate Methods in Optimization Problems*, American Elsevier, New York, 1970.

[5] B. N. Pshenichnyi, *Dual methods in extremum problems*, Kibernetika, 1 (1965), no. 3, pp. 89–95.

[6] ———, *Convex programming in a normed space*, Ibid., 1 (1965), no. 5, pp. 46–54.

[7] J. J. Moreau, *Fonctionelles sous-differentiables*, C. R. Acad. Sci. Paris, 257 (1963), pp. 4117–4119.

[8] ———, *Semi-continuité de sous-gradient d'une fonctionelle*, Ibid., 360 (1965), pp. 1057–1070.

[9] R. T. Rockafellar, *Characterization of the subdifferentials of convex functions*, Pacific J. Math., 17 (1966), pp. 497–510.

[10] ———, *Convex Analysis*, Princeton University Press, Princeton, N.J., 1970.

[11] A. Brøndsted and R. T. Rockafellar, *On the subdifferentiability of convex functions*, Proc. Amer. Math. Soc., 16 (1965), pp. 605–611.

[12] L. W. Neustadt, *A general theory of extremals*, J. Comput. System Sci., 3 (1969), pp. 57–92.

[13] W. Heins and S. K. Mitter, *Conjugate convex functions, duality, and optimal control problems, I. Systems governed by ordinary differential equations*, Information Sci., 2 (1970), pp. 211–243.

[14] R. T. Rockafellar, *Conjugate convex functions in optimal control and the calculus of variations*, J. Math. Anal. Appl., 32 (1970), pp. 174–222.

[15] D. G. Luenberger, *Control problems with kinks*, IEEE Trans. Automatic Control, AC-15 (1970), pp. 570–575.

[16] M. Z. E. Ghanem, *Optimal control problems with nondifferentiable cost functionals*, Ph.D. dissertation, Dept. of Engineering–Economic Systems, Stanford University, Stanford, Calif., 1970.

[17] D. P. Bertsekas and S. K. Mitter, *Steepest descent for optimization problems with nondifferentiable cost functionals*, Proc. 5th Annual Princeton Conference on Information Sciences and Systems, Princeton, N.J., 1971.

[18] M. S. Bazaraa, J. J. Goode and C. M. Shetty, *Optimality criteria in nonlinear programming without differentiability*, Operations Res., 19 (1971), pp. 77–86.

[19] M. S. Bazaraa, *Nonlinear programming: nondifferentiable functions*, Ph.D. dissertation, Georgia Institute of Technology, Atlanta, Ga., 1969.

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

[21] V. F. Dem'yanov, *The solution of several minimax problems*, Kibernetika, 2 (1966), no. 6, pp. 58–66.

[22] Y. M. Ermol'ev, *Methods of solution of nonlinear extremal problems*, Ibid., 2 (1966), no. 4, pp. 1–17.

[23] Y. M. Ermol'ev and N. Z. Shor, *On the minimization of nondifferentiable cost functions*, Ibid., 3 (1967), no. 1, pp. 101–102.

[24] N. Z. Shor, *On the structure of algorithms for the numerical solution of problems of optimal programming and design*, Dissertation, Kiev, 1964.

[25] W. I. Zangwill, *Nonlinear Programming: A Unified Approach*, Prentice-Hall, Englewood Cliffs, N.J., 1969.

[26] B. Birzak and B. N. Pshenichnyi, *Some problems of the minimization of unsmooth functions*, Kibernetika, 2 (1966), no. 6, pp. 43–46.

[27] L. V. Kantorovich and K. P. Akilov, *Functional Analysis in Normed Spaces*, Pergamon Press, New York, 1965, Chap. 15.

[28] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley, New York, 1969.

[29] D. P. Bertsekas, *Control of uncertain systems with a set-membership description of the uncertainty*, Ph.D. thesis, Dept. of Electrical Engineering, Mass. Inst. of Technology, Cambridge, Mass., 1971.

[30] L. W. NEUSTADT, *The existence of optimal controls in the absence of convexity conditions*, J. Math. Anal. Appl., 7 (1963), pp. 110–117.
[31] H. S. WITSENHAUSEN, *Minimax control of uncertain systems*, M.I.T. Electronics Systems Lab. Rep. ESL-R-269, Cambridge, Mass., 1966.
[32] ———, *A minimax control problem for sampled linear systems*, IEEE Trans. Automatic Control, AC-13 (1968), pp. 5–21.
[33] B. T. POLYAK, *Minimization of unsmooth functionals*, Zh. Vychisl. Mat. i Mat. Fiz., 9 (1969), no. 3, pp. 509–521.
[34] E. S. LEVITIN, *A general minimization method for unsmooth extremal problems*, Ibid., 9 (1969), no. 4, pp. 783–806.
[35] D. P. BERTSEKAS, *Stochastic optimization problems with nondifferentiable cost functionals*, J. Optimization Theory Appl., Aug., 1973.
[36] ———, *Stochastic optimization problems with nondifferentiable cost functionals with an application in stochastic programming*, Proc. 1972 Conference on Decision and Control, New Orleans, La., 1972.
[37] D. G. LUENBERGER, *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, Mass., 1973.
[38] B. T. POLYAK, *A general method for solving extremal problems*, Dokl. Akad. Nauk SSSR, 174 (1967), no. 1, pp. 33–36.
[39] R. A. MINCH, *Applications of symmetric derivatives in mathematical programming*, Math. Programming, 1 (1971), pp. 307–321.
[40] A. AUSLENDER, *Méthodes numeriques pour la décomposition et la minimization de fonctions non différentiables*, Numer. Math., 18(1971), pp. 213–223.
[41] ———, *Recherche de points de selle d'une fonction*, Cahiers Centre Etudes Recherche Opér., 12 (1970), no. 2.
[42] A. BUTZ, *Iterative saddle point techniques*, SIAM J. Appl. Math., 15 (1967), pp. 719–726.