Annual Princeton Conf. on Information Sciences and Systems, March 1971.

STEEPEST DESCENT FOR OPTIMIZATION PROBLEMS WITH NONDIFFERENTIABLE COST FUNCTIONALS*

Dimitri P. Bertsekas and Sanjoy K. Mitter
Massachusetts Institute of Technology, Cambridge, Mass.

1. Introduction

The steepest descent method[1] is a commonly used algorithm for finding the minimum of a differentiable cost functional. At each iteration a descent is made at the direction of the negative gradient according to some step size selection scheme. Convergence of the algorithm to a point satisfying the necessary conditions for a local minimum can be proved under quite general assumptions. Cost functionals which are not differentiable arise naturally in some situations including minimax problems[2-6]. In addition some constrained minimization problems can be converted into unconstrained minimization problems with a nondifferentiable cost functional. For example[2,8] the problem

$$\min_x f_o(x) \quad \text{subject to } f_i(x) \le 0, \quad i=1,2,\ldots N \tag{1}$$

is equivalent to the unconstrained problem

$$\min_x \{f_o(x) + K \sum_{i=1}^{N} \max[0, f_i(x)]\} \tag{2}$$

whenever Kuhn-Tucker multipliers $\{\lambda_1,\lambda_2,\ldots,\lambda_N\}$ for problem (1) exist and it is $K \ge \max\{\lambda_1,\ldots,\lambda_N\}$. However the cost functional in (2) is nondifferentiable even if the functions f_i are differentiable.

This paper has two objectives. First to examine the natural extension of the steepest descent algorithm for minimizing a directionally differentiable function mapping R^n (n-dimensional Euclidean space) into the real line. For the case where this function is convex we demonstrate a new characterization of the direction of steepest descent of the cost function at a point in terms of the subdifferential[9] of the function at this point. Using this characterization we generalize some results recently obtained by Luenberger[2] which concern optimal control problems. We also discuss the difficulties for proving convergence of this method to a minimum which are due to the nondifferentiability of the cost function. The second objective of the paper is to propose a new descent algorithm for minimizing an extended real valued convex function. This algorithm is a large step, double iterative algorithm and it is convergent under the sole assumption that the cost function is bounded below. The algorithm is based on the notion of the ε-subgradient of a convex function[9] and contains as a special case the algorithm of Pshenischnyi[6].

*This research was carried out at the Decision and Control Sciences Group of the M.I.T. Electronic Systems Laboratory with support extended by the Air Force Office of Scientific Research under grants AFOSR 70-1941 and by NASA NGL-22-009-124.

2. Steepest Descent for a Directionally Differentiable Cost Function

The problem that we consider in this section is to find $\inf_x f(x)$ where $f: R^n \to R$ is a directionally differentiable function. The function f is called directionally differentiable if the one-sided directional derivative with respect to any vector $y \in R^n$

$$f'(x;y) = \lim_{\lambda\to 0^+} \frac{f(x+\lambda y)-f(x)}{\lambda} \tag{3}$$

exists for all $x \in R^n$. The direction of steepest descent of f at x is defined[10] by the vector $\bar{y}(x)$ where

$$f'[x;\bar{y}(x)] = \min_{||y||\le 1} f'(x;y) \tag{4}$$

The steepest descent algorithm is specified by the iteration

$$x_{n+1} = x_n + \lambda_n \bar{y}(x_n) \tag{5}$$

where $\lambda_n > 0$ is the step size of the iteration and is chosen according to some scheme. Three problems of interest are: a) The characterization of $\bar{y}(x)$. b) The specification of the method for selecting the step size λ_n. c) the conditions under which $\lim_{n\to\infty} f(x_n) = \min_x f(x)$. We examine these problems in turn in what follows.

We first consider the case where the function f is convex, i.e. it is $f[\lambda x_1+(1-\lambda)x_2] \le \lambda f(x_1) + (1-\lambda)f(x_2)$ for all $0 \le \lambda \le 1$, $x_1,x_2 \in R^n$. Then f is directionally differentiable[9] and the function $f'(x;y)$ is a convex positively homogeneous convex function. The subdifferential of f at a point x is defined[9] as the set

$$\partial f(x) = \{x^*|\ f(z) \ge f(x) + \langle z-x,x^*\rangle, \forall z\in R^n\} \tag{6}$$

and it is a convex and compact set. The support function of $\partial f(x)$ is given by[9]

$$\sigma[y|\partial f(x)] = \max_{x^*\in\partial f(x)} \langle x^*,y\rangle = f'(x;y) \tag{7}$$

The subdifferential $\partial f(x)$ is a generalization of the notion of the ordinary gradient. If f is differentiable at x, then $\partial f(x)$ consists of a single point, the gradient $\nabla f(x)$. At points where f is not differentiable, the calculation of $\partial f(x)$ can be facilitated by the following expressions[9]

$$\partial(f_1+f_2+\ldots+f_N)(x) = \partial f_1(x)+\ldots+\partial f_N(x) \tag{8}$$

where $f_i: R^n\to R$ are convex functions for all i.

$$\partial f(x) = \{x^*|\ f^*(x^*) + f(x) - \langle x,x^*\rangle \le 0\} \tag{9}$$

where $f^*(x^*) = \sup_x\{<x,x^*> - f(x)\}$ is the conjugate convex function of f[9].

The direction of steepest descent $\bar{y}(x)$ can be conveniently characterized in terms of the set $\partial f(x)$.

Proposition 1: It is $\bar{y}(x) = -\frac{\bar{x}^*}{||\bar{x}^*||}$ where $\bar{x}^*$ is the vector of minimum (Euclidean) norm in the subdifferential $\partial f(x)$.

Proof: By (7) $\min_{||y||\le 1} f'(x;y) = \min_{||y||\le 1} \max_{x^*\epsilon\partial f(x)} <x^*,y>$ and since the constraint sets are convex and compact[11] we can interchange max and min in the above relation

$$\min_{||y||\le 1} f'(x;y) = \max_{x^*\epsilon\partial f(x)} \min_{||y||\le 1} <x^*,y> =$$

$$\max_{x^*\epsilon\partial f(x)} <x^*, -\frac{x^*}{||x^*||}> = -\min_{x^*\epsilon\partial f(x)} ||x^*|| =$$

$$= -||\bar{x}^*||$$

and the minimum in (4) is attained for

$$\bar{y}(x) = -\frac{\bar{x}^*}{||\bar{x}^*||}. \qquad \text{Q.E.D.}$$

When the function f is not convex but instead it is of the form $f = g\cdot h$ where $g: R^m \to R$ is a convex function and h is a differentiable mapping $h: R^n \to R^m$ with the value of its (Frechet) derivative at a point x denoted as $\frac{\partial h}{\partial x} = H_x: R^n \to R^m$, f is directionally differentiable with directional derivative given by the equation

$$f'(x;y) = g'[h(x); H_x y] \tag{8}$$

The above equation is a generalization of the well known chain rule and follows from Th. 3.2 and p. 48 in[12] and Th. 10.4 in [9]. From equation (8) it can be proved similarly as in Proposition 1 that the direction of steepest descent of the function $f = g\cdot h$ at a point x is $\bar{y}(x) = -\frac{\bar{x}^*}{||\bar{x}^*||}$ where $\bar{x}^*$ is the vector of minimum norm in the set $H_x'\partial g[h(x)]$ (where prime denotes transposition of the matrix). Since at an optimal point $\bar{x}$ it must be $\bar{y}(\bar{x}) = 0$ we have the necessary condition for optimality $0 \,\epsilon\, H_x'\partial g[h(x)]$.

The characterization of the direction of steepest descent in terms of the subdifferential is particularly useful for optimal control problems. Consider the minimization of the functional

$$J(u_o,u_1,\ldots,u_{N-1}) = \sum_{k=1}^{N} g_k(x_k,u_{k-1})$$

where $g_k: R^n \times R^m \to R$ are convex functions, subject to the system equation constraints

$$x_{k+1} = f_k(x_k,u_k) \qquad k=0,1,\ldots,N-1$$

where the initial state x_o is given and f_k are differentiable functions, $f_k: R^n \times R^m \to R$, with the values of their derivatives denoted as $\partial f_k/\partial x_k = F_{x_k}$, $\partial f_k/\partial u_k = F_{u_k}$. Then by using arguments very similar to the above we obtain that the direction of steepest descent at a point $\{u_o,u_1,\ldots,u_{N-1}\}$ with corresponding system trajectory $\{x_o,x_1,\ldots,x_N\}$ is specified by

$$\bar{y}(u_o,u_1,\ldots,u_{N-1}) = -\frac{\bar{u}^*}{||\bar{u}^*||}$$

where $\bar{u}^* \,\epsilon\, R^{Nm}$ is the vector

$$\bar{u}^* = \{(F'_{u_o}\bar{\lambda}_1 + \bar{q}_o),(F'_{u_1}\bar{\lambda}_2 + \bar{q}_1),\ldots,(F'_{u_{N-1}}\bar{\lambda}_N + \bar{q}_{N-1})\}$$

where $\bar{\lambda}_1,\bar{\lambda}_2,\ldots,\bar{\lambda}_N$, $\bar{q}_o,\bar{q}_1,\ldots,\bar{q}_{N-1}$ are vectors such that

$$\sum_{k=0}^{N-1} ||F'_{u_k}\bar{\lambda}_{k+1} + \bar{q}_k||^2 = \min_{(p_{k+1},q_k)\epsilon\partial g_{k+1}(x_{k+1},u_k)} \sum_{k=0}^{N-1} ||F'_{u_k}\lambda_{k+1} + q_k||^2$$

subject to the adjoint equation

$$\lambda_k = F'_{x_k}\lambda_{k+1} + p_k, \qquad \lambda_N = p_N$$

where the derivatives are evaluated along the point $\{u_o,u_1,\ldots,u_{N-1}\}$ and the corresponding trajectory $\{x_o,x_1,\ldots,x_N\}$. A necessary condition for optimality of the control sequence $\{\bar{u}_o,\bar{u}_1,\ldots,\bar{u}_{N-1}\}$ with corresponding trajectory $\{x_o,\bar{x}_1,\ldots,\bar{x}_N\}$ is that there exist vectors $\bar{\lambda}_1,\bar{\lambda}_2,\ldots,\bar{\lambda}_N$ such that

$$F'_{u_k}\bar{\lambda}_{k+1} + \bar{q}_k = 0 \qquad k=0,1,\ldots,N-1$$

$$\bar{\lambda}_k = F'_{x_k}\bar{\lambda}_{k+1} + \bar{p}_k \qquad k=1,2,\ldots,N-1$$

$$\bar{\lambda}_N = \bar{p}_N$$

for some vectors $\bar{q}_o,\bar{q}_1,\ldots,\bar{q}_{N-1}$, $\bar{p}_1,\bar{p}_2,\ldots,\bar{p}_N$ with $(\bar{p}_k,\bar{q}_{k-1}) \,\epsilon\, \partial g_k(\bar{x}_k,\bar{u}_{k-1})$, $k=1,2,\ldots,N$.

The above equations generalize to a larger class of problems the results obtained by Luenberger[2]. For related necessary conditions see also[13]. In the case where in the above problem there are state and control constraints a reformulation of the problem to an unconstrained problem is possible, whenever Kuhn-Tucker multipliers exist, as in equation (2).

We now turn our attention to various methods for selecting the step size λ_n in equation (5). The natural method would be to choose λ_n from $f[x_n+\lambda_n \bar{y}(x_n)] = \min_{\lambda\geq 0} f[x_n+\lambda\bar{y}(x_n)]$ assuming that such a λ_n exists. Despite the fact that the value of the cost function decreases at each iteration this method of selection of λ_n may not lead to convergence to the optimum even when the function f is convex. As reported for some minimax problems[4], this is due to the fact that the directional derivative $f'(x;y)$ is not in general a continuous function of x for any fixed y, and the algorithm may "jam" at points where f is not differentiable. For this reason in the minimax algorithms of[4,5,6] antijamming techniques are used to make the corresponding algorithms convergent. A different method of selection of the step size is to select a priori a sequence $\{\lambda_n\}$ where $\lambda_n > 0$, $\{\lambda_n\}\to 0$, $\sum_{n=1}^{\infty} \lambda_n = \infty$. With this selection of λ_n it can be proved that if f is convex and if the points $\{x_n\}$ generated by equation (5) are in some bounded set or if the function f is Lipschitz continuous, then for every $\varepsilon>0$ there exists an N such that $|f(\bar{x})-f(x_N)| < \varepsilon$ where $f(\bar{x}) = \min_x f(x)$. The resulting algorithm is an adaptation of an algorithm reported by Ermolev[7] and attributed to Shor. The convergence proof follows similar arguments as in[7]. The drawback of this algorithm is that a decrease of the value of the cost function is not generally observed at each iteration, possibly resulting to slow convergence, and that there is no good criterion for terminating the iterations. These drawbacks are avoided in the algorithm we describe in the next section.

3. A Descent Algorithm for Minimizing Convex Cost Functionals

In this section we present a new descent algorithm for minimizing an extended real valued convex function. We will consider a convex function $f: R^n\to(-\infty,\infty]$ which is lower semicontinuous and such that $f(x)<\infty$ for at least one $x\in R^n$, i.e. f is a closed proper convex function according to[9]. The problem is to find $\inf_x f(x)$ and a vector $\bar{x}$ (if it exists) such that $f(\bar{x}) = \min_x f(x)$. The only assumption that we will need is that $-\infty<\inf_x f(x)$. It should be noted that the use of the extended real line permits the incorporation of constraints into the cost functional. Thus the problem of finding the minimum of a function f over a set X is equivalent to finding the unconstrained minimum of the function $f(x) + \delta(x|X)$ 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 $x\notin X$. A basic concept for the algorithm that we present is the notion of the ε-subgradient[9].

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

$$f(z) \geq f(x)-\varepsilon + \langle z-x,x^*\rangle,\ \forall\, z\,\varepsilon\, R^n$$

The set $\partial_\varepsilon f(x)$ of all ε-subgradients at x will be called the <u>ε-subdifferential</u> at x. The set $\partial_\varepsilon f(x)$ is a nonempty closed convex set and is characterized by the equation[9]

$$\partial_\varepsilon f(x) = \{x^*|\ f^*(x^*) + f(x) - \langle x,x^*\rangle \leq \varepsilon\} \quad (9)$$

where $f^*(x^*) = \sup_x\{\langle x,x^*\rangle -f(x)\}$ is the conjugate convex function of f. It is evident that for $0<\varepsilon_1<\varepsilon_2$ it is $\partial_{\varepsilon_1} f(x)\subset \partial_{\varepsilon_2} f(x)$. The support function of $\partial_\varepsilon f(x)$ is given by[9]

$$\sigma[y|\partial_\varepsilon f(x)] = \sup_{x^*\varepsilon\partial_\varepsilon f(x)} \langle x^*,y\rangle = \inf_{\lambda>0} \frac{f(x+\lambda y)-f(x)+\varepsilon}{\lambda} \quad (10)$$

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

<u>Proposition 2</u>: Let x be such that $f(x) < \infty$. Then

$$0 \leq f(x) - \inf_x f(x) \leq \varepsilon \iff 0\,\varepsilon\,\partial_\varepsilon f(x)$$

<u>Proof</u>: Let $0 \leq f(x) - \inf f(x) \leq \varepsilon$. Then for any $\lambda>0$ and any $y\,\varepsilon\, R^n$ it is $\frac{f(x+\lambda y)-f(x)+\varepsilon}{\lambda} \geq \frac{\inf_x f(x)-f(x)+\varepsilon}{\lambda} \geq 0$. Taking the infimum over $\lambda>0$ we have $\inf_{\lambda>0} \frac{f(x+\lambda y)-f(x)+\varepsilon}{\lambda} \geq 0$ for all $y\varepsilon R^n$, or using (10) $\sigma[y|\partial_\varepsilon f(x)] \geq 0$ for all $y\varepsilon R^n$ which implies $0\varepsilon\partial_\varepsilon f(x)$. Conversely let $0\varepsilon\partial_\varepsilon f(x)$ from which $\inf_{\lambda>0} \frac{f(x+\lambda y)-f(x)+\varepsilon}{\lambda} \geq 0$ for all y. Then $f(x+y)-f(x)+\varepsilon\geq 0$ for all y and taking the infimum over y we obtain $\inf_x f(x)-f(x)+\varepsilon \geq 0$. Q.E.D.

<u>Proposition 3</u>: Let x be point such that $f(x) < \infty$ and $0 \notin \partial_\varepsilon f(x)$. Let y be any vector such that $\sup_{x^*\varepsilon\partial_\varepsilon f(x)} \langle y,x^*\rangle < 0$. Then

$$f(x) - \inf_{\lambda\geq 0} f(x+\lambda y) > \varepsilon$$

<u>Proof</u>: Assume the contrary, i.e. $\inf_{\lambda\geq 0} f(x+\lambda y) - f(x)+\varepsilon \geq 0$. Then for any $\lambda>0$ it will be $\frac{f(x+\lambda y)-f(x)+\varepsilon}{\lambda} \geq 0$ or $\inf_{\lambda>0} \frac{f(x+\lambda y)-f(x)+\varepsilon}{\varepsilon} \geq 0$ or or using (10) $\sup_{x^*\varepsilon\partial_\varepsilon f(x)} \langle x^*,y\rangle \geq 0$ a contradiction. Q.E.D.

It should be noted that if $0 \notin \partial_\varepsilon f(x)$ a vector y such that $\sup_{x^*\varepsilon\partial_\varepsilon f(x)} \langle y,x^*\rangle < 0$ is the vector $y=-\bar{x}^*$ where $\bar{x}^*$ is the (unique) vector of minimum Euclidean norm of $\partial_\varepsilon f(x)$, since it is $\langle\bar{x}^*,x^*\rangle \geq ||\bar{x}^*||^2 > 0$ for all $x^*\varepsilon\partial_\varepsilon f(x)$ from which $\sup_{x^*\varepsilon\partial_\varepsilon f(x)} - \langle\bar{x}^*,x^*\rangle \leq -||\bar{x}^*||^2 < 0$.

The Prop. 1 provides a termination criterion for any iterative minimization algorithm. The Prop. 2 states that if the value f(x) exceeds the optimal value by more than ε, then by a descent along the negative vector of minimum norm in the ε-subdifferential $\partial_\varepsilon f(x)$ we can decrease the value of the cost by at least ε. This fact provides the basis for the following algorithm.

1) Select initially a scalar $\varepsilon_o>0$, a vector x_o such that $f(x_o)<\infty$, and a number a, $0<a<1$.

2) Given $\varepsilon_n>0$ and x_n, if $0\notin\partial_{\varepsilon_n}(x_n)$ set $\varepsilon_{n+1}=\varepsilon_n$. If $0\varepsilon\partial_{\varepsilon_n}(x_n)$ implying $f(x_n)-\inf_x f(x)\le\varepsilon_n$ multiply ε_n consecutively by a and set $\varepsilon_{n+1}=a^k\varepsilon_n$ for the smallest integer $k>0$ for which $0\notin\partial_{\varepsilon_{n+1}} f(x_n)$. If $f(x_n)\ne\min_x f(x)$ there exists such a k. Then set $x_{n+1}=x_n+\lambda_n y_n$ where y_n is a vector such that $\sup\{<y_n,x^*>|\ x^*\varepsilon\partial_{\varepsilon_{n+1}} f(x_n)\}<0$ ($-y_n$ can be the vector of minimum norm in $\partial_{\varepsilon_{n+1}} f(x_n)$), and $\lambda_n>0$ is such that $f(x_n)-f(x_n+\lambda_n y_n)>\varepsilon_{n+1}$. By the fact that $0\notin\partial_{\varepsilon_{n+1}} f(x_n)$ and Proposition 3 such a scalar λ_n exists and can be found by a one-dimensional search. Another possible method is to select λ_n such that $f(x_n+\lambda_n y_n)=\min_{\lambda\ge 0} f(x_n+\lambda y_n)$ provided the minimum is attained. It can be easily proved that this can be guaranteed wherever the set of optimal points $M=\{\bar{x}|\ f(\bar{x})=\min_x f(x)\}$ is non-empty and bounded.

The following proposition gives the convergence properties of the algorithm:

<u>Proposition 4</u>: Consider the vectors x_n generated by the above algorithm. Then either $f(x_N)=\min_x f(x)$ for some $N\ge 0$, or the following statements hold: a) $\lim_{n\to\infty} f(x_n)=\inf_x f(x)$ b) $f(x_n)-f(x_{n+1})>\varepsilon_{n+1}>0$ c) $(a^{-1}-1)\varepsilon_n\ge f(x_n)-\inf_x f(x)>\varepsilon_{n+1}$ for all n such that $\varepsilon_n<\varepsilon_o$. If in addition the set $M=\{\bar{x}|\ f(\bar{x})=\min_x f(\bar{x})\}$ is nonempty and bounded then: d) Every convergent subsequence of $\{x_n\}$ has its limit in M, and at least one such subsequence exists. e) For every $\varepsilon>0$ there exists an $N>0$ such that $x_n\varepsilon M+\varepsilon B$ for all $n\ge N$ where $B=\{x|\ ||x||\le 1\}$ is the unit ball in R^n. f) If the minimum of f is attained at a single point $\bar{x}$ then $\{x_n\}\to\bar{x}$.

<u>Proof</u>: b) and c) follow directly from the construction of the scalars ε_n and Prop. 2 and 3. To prove a) in view of c) it is sufficient to prove $\{\varepsilon_n\}\to 0$. We have $\varepsilon_n\ge\varepsilon_{n+1}>0$ for all n. Therefore $\{\varepsilon_n\}\to\varepsilon$ where $\varepsilon\ge 0$ is some scalar and $\varepsilon_n\ge\varepsilon$ for all n. From b) we have $f(x_o)-f(x_1)\ge\varepsilon_1\ge\varepsilon$, $f(x_1)-f(x_2)\ge\varepsilon_2\ge\varepsilon,\dots,f(x_{N-1})-f(x_N)\ge\varepsilon_N\ge\varepsilon$ and by adding these relations $f(x_o)-f(x_N)\ge N\varepsilon$ or $f(x_o)-N\varepsilon\ge f(x_N)>\inf_x f(x)$ for all $N>0$. If it were $\varepsilon>0$ then the left hand side of the last inequality would decrease without bound. Therefore $\varepsilon=0$ and a) is proved. To prove d) notice that $x_n\varepsilon F_o$ where $F_o=\{x|\ f(x)\le f(x_o)\}$ and since M is nonempty and bounded, F_o is compact (see[9] 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[9]. Part e) follows from a) and Th. 27.2 in[9]. Part f) follows from a) and Cor. 27.2.2 in[9].

Q.E.D.

It can be seen that at each step of the above algorithm a double iteration must be done. The first iteration is to determine ε_{n+1} and y_n from ε_n and x_n. The vector of minimum norm on the sets $\partial_{\varepsilon_n} f(x_n)$, $\partial_{a\varepsilon_n} f(x_n)$, $\partial_{a^2\varepsilon_n} f(x_n),\dots$ is determined and for the first k for which $\partial_{a^k\varepsilon_n} f(x_n)$ does not contain the origin we set $\varepsilon_{n+1}=a^k\varepsilon_n$ and take y_n to be the negative vector of minimum norm on $\partial_{\varepsilon_{n+1}} f(x_n)$. This can be done by using standard techniques provided that set $\partial_{\varepsilon_n} f(x_n)$ can be characterized. The second iteration is a one dimensional search along the direction of y_n. At each iteration the cost is decreased by at least ε_{n+1} and if in fact the scalar $(a^{-1}-1)$ is selected to be sufficiently small, it can be seen from c) in Prop. 4 that we can get arbitrarily close to the optimal value in a single step. This would of course result in a large number of iterations to find ε_{n+1}. The condition c) can serve as a termination criterion of the algorithm.

A question which requires extensive discussion and cannot be examined within the space limit of this paper is the convenient characterization of the set $\partial_\varepsilon f(x)$. The characterization of $\partial_\varepsilon f(x)$, or some suitable approximation of it, is possible for a large class of functions. We will only mention here that, if necessary, the algorithm can be modified so that it is not necessary to calculate exactly the set $\partial_{\varepsilon_{n+1}} f(x_n)$ but instead it is possible to find the direction of descent y_n from the vector of minimum norm in a set S_n where $\partial_{\varepsilon_{n+1}} f(x_n)\subset S_n\subset\partial_{m\varepsilon_{n+1}} f(x_n)$ where $m>1$ is some scalar. Convergence of the algorithm will still be maintained. An important application of this modification of the algorithm is when f is of the form $f=f_1+f_2+\dots+f_N$. Then, under some inessential assumptions, it can be proved that $\partial_\varepsilon f(x)\subset\partial_\varepsilon f_1(x)+\dots+\partial_\varepsilon f_N(x)\subset\partial_{N\varepsilon} f(x)$. In this case if the sets $\partial_\varepsilon f_i(x)$ can be more easily characterized, the algorithm may be easier to apply in its modified form.

As an example consider now the minimization of the function $f(x)=\sigma(x|X)+\delta(x|A)$ where

$\sigma(\cdot|X)$ is the support function of a given set X and $\delta(\cdot|A)$ is the indicator function of a given hyperplane A. This is the problem considered by Pshenichnyi[5]. It can be proved that $\partial_\varepsilon f(x) = \{x^*|x^* \varepsilon X, \sigma(x|X) - \langle x,x^*\rangle \leq \varepsilon\} + A^\perp$ where $A^\perp$ is the one-dimensional subspace orthogonal to A. The algorithm in[5] then becomes identical to the algorithm of this section as applied to this particular problem.

References

1. Zangwill, W. I., Nonlinear Programming, Prentice Hall, 1969.

2. Luenberger, D. G., "Control Problems with Kinks," IEEE Trans., AC-15, No. 5, 1970.

3. Danskin, J. M., "The Theory of Max-Min with Applications," J. SIAM Appl. Math., Vol. 14, No. 4, 1966.

4. Dem'yanov, V. F., "The Solution of Several Minimax Problems," Kibernetica, Vol. 2, No. 6, 1966.

5. Pshenichnyi, B. N., "Dual Methods in Extremum Problems," Kibernetica, No. 3, 1965.

6. Birzak, B., and Pshenichnyi, B. N., "Some Problems of the Minimization of Unsmooth Functions," Kibernetica, Vol. 2, No. 6, 1966.

7. Ermol'ev, Y. M., "Methods of Solution of Nonlinear Extremal Problems," Kibernetica, Vol. 2, No. 4, 1966.

8. Ermol'ev, Y. M., and Shor, N. Z., "On the Minimization of Nondifferentiable Functions," Kibernetica, Vol. 3, No. 1, 1967.

9. Rockafellar, R. T., Convex Analysis, Princeton, 1970, pp. 217, 220.

10. Kantorovich, L. V., and Akilov, K. P., Functional Analysis in Normed Spaces, Pergamon, New York, 1965, Ch. 15.

11. Luenberger, D. G., Optimization by Vector Space Methods, Wiley, 1969, p. 208.

12. Dem'yanov, V. F., and Rubinov, A. M., Approximate Methods in Optimization Problems, Am. Elsevier Publ. Co., New York, 1970.

13. Rockafellar, R. T., "Conjugate Convex Functions in Optimal Control and the Calculus of Variations," Journal of Math. Analysis and Applications, Vol. 32, pp. 174-222, 1970.