JOURNAL OF APPROXIMATION THEORY 35, 111–126 (1982)

A Subgradient Projection Algorithm

V. P. SREEDHARAN

Department of Mathematics,
Michigan State University, East Lansing, Michigan 48824, USA

Communicated by E. W. Cheney

Received February 20, 1981

1. INTRODUCTION

This paper supplies proofs of various results announced in our earlier paper [11]. Some of the introductory statements from [11] are repeated here to make the presentation self-contained. We present an algorithm for minimizing a certain type of non-differentiable convex function constrained to a convex polytope in $\mathbb{R}^d$. The particular form of this problem was motivated by some questions in the equilibrium theory of mathematical economics. The economic motivation and some other mathematical considerations appear in Rubin and Sreedharan [8]. The problem is a generalization of the minimum norm problem of approximation theory [10]. As explained below it is also a generalization of results due to Rosen [6], Wolfe [12] and Lemarechal [2].

The problem of minimizing a convex function over a linearly constrained set has been the subject of much study. We single out the approach taken by Rosen [6], where the objective function is smooth. Rosen attempts to exploit the known convergence of the method of steepest descent in the unconstrained case. As is typical of so-called "feasible direction" methods, Rosen computes at each iteration a direction of descent which points into the constrained region from the current point. He searches in that direction until he reaches either a relative minimum along the ray of search or the boundary of the constrained region. The process then repeats. When possible Rosen uses the negative gradient as the direction. When this direction points out of the set, he projects it onto a face of the set. Rosen's contribution is the choice of projected direction. His method is susceptible to a phenomenon known variously as "jamming" or "zigzagging," in which the sequence generated clusters at, or even converges to, nonoptimal points. The trouble lies in the possibility that the sequence is alternating among two or more faces of the constrained region in such a way that the distance along the direction of search from the current point to the boundary is going to zero.

0021-9045/82/060111-16$02.00/0

Various modifications have been proposed to avoid this. In particular, Polak [3, 4] has adopted a technique which prohibits the sequence generated from approaching arbitrarily close to a face when conditions for a constrained minimum are not being met at the limit.

Rosen's method, and much of the other work in the area, requires that the objective function be differentiable. It turns out that in the economic problem alluded to earlier the objective function is not differentiable everywhere. Attention has recently been focused on algorithms for optimizing nondifferentiable convex functions. The algorithms of Wolfe [12] and Lemarechal [2] generalize classical methods for unconstrained optimization by replacing the gradient with a carefully chosen subgradient. These algorithms do not apply to the constrained case. The algorithm of Bertsekas and Mitter [1] is theoretically applicable to constraints even more general than linear ones. In spite of the ease of proof of convergence by these authors, their algorithm requires the computation of the "ε-subdifferential" of the objective function, a prohibitive task even in the linearly constrained case. The algorithm proposed here is prompted by those of Sreedharan [9, 10], Rosen [6] and Polak [3, 4]. Our algorithm, in contrast to the Bertsekas and Mitter algorithm, requires only the computation of a subset of the ε-subdifferential and is computationally feasible for a large class of problems. The actual computational details and experience with a Fortran program for our algorithm on a CDC 6500 computer will be given in the paper by Rubin [7].

2. Problem

In this paper we denote the standard Euclidean inner product of two vectors in $\mathbb{R}^d$ by simply juxtaposing them. The corresponding Euclidean length of a vector is denoted by $|\cdot|$. We now state the problem. Let $X \subset \mathbb{R}^d$ be a non-empty convex polytope defined by

$$X = \{x \in \mathbb{R}^d \mid a_i x \leqslant b_i,\ i = 1,\ldots, m\},$$

where a_i, $x \in \mathbb{R}^d$, $b_i \in \mathbb{R}$. By definition a convex polytope is the convex hull of a finite number of points in $\mathbb{R}^d$ and as such is bounded. Let f be a smooth strictly convex function on X; i.e.,

$$f((x+y)/2) < (f(n) + f(y))/2, \qquad x \neq y,\ x, y \in X,$$

and is of class C^1 in a neighborhood of X. Further, let v be a piecewise affine convex function on X; i.e.,

$$v(x) = \max\{v_j(x) \mid 1 \leqslant j \leqslant r\},$$

where

$$v_j(x) = g_j x + c_j, \qquad g_j,\ x \in \mathbb{R}^d,\ c_j \in \mathbb{R}.$$

Such a v is not differentiable everywhere except in trivial cases. The problem is to minimize $f(x) + v(x)$ subject to the constraint $x \in X$. Symbolically we have

$$\text{(P)} \begin{cases} a_i x \leqslant b_i, & i = 1,\ldots, m \\ f(x) + v(x) \text{ (min)}. \end{cases}$$

We shall refer to this as problem (P). This form of the objective function arises naturally, when v is obtained by solving a linear programming problem whose right-hand side, the constraint constants, depends linearly on a parameter x. In fact, this particular form arose from the economic model referred to earlier.

3. Notation

Let x be a point in the constraint set X and $\varepsilon \geqslant 0$. We first define the sets of indices

$$I_\epsilon(x) = \{1 \leqslant i \leqslant m \mid a_i x \geqslant b_i - \varepsilon\}; \tag{3.1}$$

$$J_\epsilon(x) = \{1 \leqslant j \leqslant r \mid v_j(x) \geqslant v(x) - \varepsilon\}. \tag{3.2}$$

Note that

$$I_0(x) = \{1 \leqslant i \leqslant m \mid a_i x = b_i\}, \tag{3.3}$$

and

$$J_0(x) = \{1 \leqslant j \leqslant r \mid v_j(x) = v(x)\}. \tag{3.4}$$

Using the above index sets we define the convex subsets of $\mathbb{R}^d$

$$C_\epsilon(x) = \text{cone}\{a_i \mid i \in I_\epsilon(x)\}; \tag{3.5}$$

and

$$K_\epsilon(x) = \text{conv}\{g_j \mid j \in J_\epsilon(x)\}. \tag{3.6}$$

Here and elsewhere we denote by cone S the convex cone generated by S with apex at the origin and by conv S the convex hull of the set S.

For any nonempty closed convex set $S \subset \mathbb{R}^d$ there is a unique point $a \in S$ nearest to the origin, which we denote by $N[S]$. The point $a = N[S]$ is characterized by the inequality

$$a(x-a) \geqslant 0 \qquad \text{for all } x \in S. \tag{3.7}$$

4. Algorithm

In this section we present a subgradient projection algorithm for solving problem (P).

Step 0. Start with arbitrary $x_0 \in X$, $\varepsilon_0 > 0$ and $k = 0$.

Step 1. Compute $y_0 = N[K_0(x_k) + \nabla f(x_k) + C_0(x_k)]$. If $y_0 = 0$, STOP; x_k is the solution of problem (P). If $y_0 \neq 0$, set $\varepsilon = \varepsilon_0$.

Step 2. Compute $y_\varepsilon = N[K_\varepsilon(x_k) + \nabla f(x_k) + C_\varepsilon(x_k)]$.

Step 3. If $|y_\varepsilon|^2 > \varepsilon$ set $\varepsilon_k = \varepsilon$, $s_k = y_\varepsilon$ and GO TO Step 5.

Step 4. Replace ε by $\varepsilon/2$ and GO TO Step 2.

Step 5. Compute $\bar{\alpha}_k = \max\{\alpha \in \mathbb{R} \mid x_k - \alpha s_k \in X\}$. (It will be shown that $\bar{\alpha}_k$ is positive.) Find $\alpha_k \in [0, \bar{\alpha}_k]$ such that there exists

$$z_k \in K_0(x_k - \alpha_k s_k) + \nabla f(x_k - \alpha_k s_k),$$

with

$$z_k s_k = 0.$$

If no such α_k exists, set $\alpha_k = \bar{\alpha}_k$.

Step 6. Define $x_{k+1} = x_k - \alpha_k s_k$. Increment k by 1 and GO TO Step 1.

Steps 1 and 2 can be implemented as special quadratic programs. Step 5 requires a properly constructed line search and some comparisons. In practice, Step 1 would be replaced by the statement: STOP if $|y_0|$ is sufficiently small. We refer to the paper of Rubin [7] for the computational details and experience. With small problems even hand calculations using the above algorithm yielded good answers.

5. Convergence of the Algorithm

The proof that the algorithm converges is somewhat involved and depends on a series of lemmas, some of which are of independent interest.

We need some more terminology and notation. When $F: \mathbb{R}^d \to [-\infty, \infty]$ is a convex function its *ε-subdifferential* $\partial_\varepsilon F(x)$, where $\varepsilon \geqslant 0$, is defined by saying

$$u \in \partial_\varepsilon F(x) \qquad \text{iff} \quad F(y) \geqslant F(x) + u(y-x) - \varepsilon, \ \forall y \in \mathbb{R}^d.$$

$\partial_0 F(x)$ is the *subdifferential* of F at x which we denote by $\partial F(x)$. Any $u \in \partial F(x)$ is referred to as a *subgradient* of F at x. More explicitly, u satisfies the subgradient inequality

$$F(y) \geqslant F(x) + u(y-x), \qquad \forall y \in \mathbb{R}^d.$$

Note, however, that $\partial F(x)$ can be empty. See Rockafellar [5] for all these and related notions. The constrained problem (P) is converted into an unconstrained one via the indicator function δ of the constraint set X which is defined by setting $\delta(x) = 0$ if $x \in X$ and $\delta(x) = \infty$ if $x \notin X$. Then $F = f + v + \delta$ is convex on $\mathbb{R}^d$ and the problem minimize $F(x)$, $x \in \mathbb{R}^d$ is equivalent to the constrained minimization problem (P). We now state a sequence of lemmas, using the earlier notation.

5.1. LEMMA. *For all $\varepsilon \geqslant 0$ and all $x \in \mathbb{R}^d$, $K_\varepsilon(x) \subset \partial_\varepsilon v(x)$.*

Proof. If $u \in K_\varepsilon(x)$, then by (3.6) there exist $\lambda_j \geqslant 0$, $j \in J_\varepsilon(x)$ such that $\sum \lambda_j = 1$ and

$$u = \sum_{j \in J_\varepsilon(x)} \lambda_j g_j. \tag{5.1.1}$$

For $j \in J_\varepsilon(x)$, we have

$$v_j(y) = v_j(x) + g_j(y-x) \geqslant v(x) - \varepsilon + g_j(y-x).$$

Therefore for every $j \in J_\varepsilon(x)$,

$$v(y) = \max_{i=1,\ldots,r} v_i(y) \geqslant v(x) - \varepsilon + g_j(y-x).$$

Multiplying these inequalities by λ_j and summing over the index set $J_\varepsilon(x)$ we arrive at the inequality

$$v(y) \geqslant v(x) - \varepsilon + u(y-x). \tag{5.1.2}$$

Since (5.1.2) holds for all $y \in \mathbb{R}^d$, the lemma is proved.

5.2. LEMMA. *Given $x \in \mathbb{R}^d$, there exists a neighborhood V of x such that $J_0(y) \subset J_0(x)$ for all $y \in V$.*

Proof. The functions $w_j = v - v_j$, $j = 1,\dots, r$ are continuous with $w_j(x) > 0$ iff $j \notin J_0(x)$. Thus there exists a neighborhood V of x such that w_j is positive throughout V for each $j \notin J_0(x)$. If $j \notin J_0(x)$ and $y \in V$, $w_j(y) > 0$, and so $j \notin J_0(y)$, proving the lemma.

5.3. In the next lemma we will use the notion of the support function of a set in $\mathbb{R}^d$. If S is a subset of $\mathbb{R}^d$, then its *support function* φ is defined by

$$\varphi(x) = \sup\{xy \mid y \in S\}, \qquad x \in \mathbb{R}^d.$$

It is well known that two closed convex subsets of $\mathbb{R}^d$ are identical iff their support functions are the same.

5.4. LEMMA. *$\partial v(x) = K_0(x)$ for every $x \in X$.*

Proof. Let $x \in X$. Since both $\partial v(x)$ and $K_0(x)$ are closed and convex, it suffices, in view of remarks in 5.3, to show that these sets have the same support function. In fact, we show that $v'(x; \cdot)$ is this support function, where $v'(x; y)$ is the directional derivative of v at x in the direction $y \in \mathbb{R}^d$, i.e.,

$$v'(x; y) = \lim_{\alpha \downarrow 0} \{v(x + \alpha y) - v(x)\}/\alpha. \tag{5.4.1}$$

Since v is an everywhere finite valued convex function, it is known that

$$v'(x; y) = \sup\{yu \mid u \in \partial v(x)\} \qquad \text{for all } y \in \mathbb{R}^d, \tag{5.4.2}$$

i.e., $v'(x; \cdot)$ is the support function of $\partial v(x)$. Given $y \in \mathbb{R}^d$, by Lemma 5.2 there exists an $\varepsilon > 0$ such that $J_0(x + \alpha y) \subset J_0(x)$ for all $\alpha \in [0, \varepsilon]$. Now

$$v(x + \alpha y) = v_j(x + \alpha y), \qquad \forall j \in J_0(x + \alpha y)$$

and

$$v(x) = v_j(x), \qquad \forall j \in J_0(x).$$

Hence we see that for $0 \leqslant \alpha \leqslant \varepsilon$,

$$\begin{aligned} v(x + \alpha y) - v(x) &= \max_{j \in J_0(x)} \{v_j(x + \alpha y) - v_j(x)\} \\ &= \max_{j \in J_0(x)} \alpha g_j y. \end{aligned}$$

This in view of (5.4.1) shows that

$$v'(x; y) = \max_{j \in J_0(x)} g_j y = \max\{uy \mid u \in K_0(x)\}, \tag{5.4.3}$$

and so $v'(x; \cdot)$ is also the support function of $K_0(x)$, completing the proof of the lemma.

We note that (5.4.3) proves the following statement:

5.5. COROLLARY. $v'(x; s) = \max\{su \mid u \in K_0(x)\}$.

5.6. LEMMA. *For each $x \in X$ and $\varepsilon > 0$ there exists a $\gamma > 0$ such that*

$$J_0(x) \subset J_\epsilon(y) \qquad \text{whenever} \quad |x - y| < \gamma.$$

Proof. Choose $\gamma > 0$ such that $|g_j| \gamma < \varepsilon/2$ for $j = 1,..., r$ and $|v(x) - v(y)| < \varepsilon/2$ if $|x - y| < \gamma$. Now if $j \in J_0(x)$ and $|x - y| < \gamma$, then

$$\begin{aligned} v(y) - v_j(y) &= v(y) - v(x) + v_j(x) - v_j(y) \\ &< \varepsilon/2 + |g_j(x - y)| < \varepsilon, \end{aligned}$$

and so $j \in J_\epsilon(y)$, proving the lemma.

5.7. LEMMA. $\partial F(x) = \nabla f(x) + K_0(x) + C_0(x)$ *for all* $x \in X$.

Proof. The indicator function δ of the set X is clearly proper and convex, while f and v are everywhere finite valued and convex. It is well known that for $x \in X$, $\partial\delta(x) = C_0(x)$. Moreover, any $x \in \operatorname{rel\,int} X$ belongs to rel int (eff dom f) $\cap$ rel int (eff dom v) $\cap$ rel int (eff dom δ), where rel int and eff dom denote relative interior and effective domain, respectively. Hence by Rockafellar [5] the lemma follows.

The next lemma shows that the stopping criterion in Step 1 of the algorithm is well chosen.

5.8. LEMMA. *If $y_0 = 0$ in Step 1 of Algorithm 4, then x_k is the minimizer of F.*

Proof. $y_0 = 0$ implies that $0 \in \partial F(x_k)$, a necessary and sufficient condition for x_k to minimize F. The strict convexity of f ensures that the minimizer of F is unique.

5.9. LEMMA. *Step 4 of Algorithm 4 is not executed infinitely often in any one iteration.*

Proof. If Step 4 is executed infinitely often, then $\varepsilon \to 0$ and $y_\epsilon \to 0$. Now $\varepsilon_1 > \varepsilon_2 \geqslant 0$ implies that $K_{\epsilon_2}(x_k) \subset K_{\epsilon_1}(x_k)$, $C_{\epsilon_2}(x_k) \subset C_{\epsilon_1}(x_k)$ and hence

$$|y_{\epsilon_1}| \leqslant |y_{\epsilon_2}|,$$

so that $y_\epsilon \to 0$ as $\varepsilon \to 0$ implies that $y_\epsilon = 0$ for every $\varepsilon \geqslant 0$. Hence $y_0 = 0$. In this event we could not have reached Step 4 in the first place, a contradiction.

We now show that Step 5 of the algorithm is well defined and can be implemented.

5.10. LEMMA. *If $s_k \neq 0$, then $-s_k$ is a feasible direction of strict descent at the point x_k.*

Proof. From the definition of s_k in Step 3 of the algorithm,

$$s_k = N[\nabla f(x_k) + K_{\epsilon_k}(x_k) + C_{\epsilon_k}(x_k)].$$

Let $i \in I_0(x_k) \subset I_{\epsilon_k}(x_k)$; then $a_i \in C_{\epsilon_k}(x_k)$ and so

$$s_k + a_i \in \nabla f(x_k) + K_{\epsilon_k}(x_k) + C_{\epsilon_k}(x_k),$$

using the fact that $C_{\epsilon_k}(x_k)$ is a convex cone. By (3.7) we have $s_k(s_k + a_i - s_k) \geqslant 0$. Thus $a_i s_k \geqslant 0$ for every $i \in I_0(x_k)$. Since $a_i x_k < b_i$ for $i \notin I_0(x_k)$, there exists $\alpha > 0$ such that $a_i(x_k - \alpha s_k) \leqslant b_i$ for all $i = 1,..., m$. Hence $-s_k$ is a feasible direction at x_k.

To show that $-s_k$ is a direction of strict descent, we show that

$$F'(x_k; -s_k) = \lim_{\alpha \downarrow 0} \{F(x_k - \alpha s_k) - F(x_k)\}/\alpha < 0. \tag{5.10.1}$$

From the first part of the proof, there exists $\bar{\alpha} > 0$ such that $x_k - \alpha s_k \in X$ for $0 \leqslant \alpha \leqslant \bar{\alpha}$. For α in this range, $F(x_k - \alpha s_k) = f(x_k - \alpha s_k) + v(x_k - \alpha s_k)$ and so by Corollary 5.5

$$\begin{aligned} F'(x_k; -s_k) &= f'(x_k; -s_k) + v'(x_k; -s_k) \\ &= -\nabla f(x_k)\, s_k + \max\{-s_k y \mid y \in K_0(x_k)\} \\ &= -\min\{(\nabla f(x_k) + y)\, s_k \mid y \in K_0(x_k)\}. \end{aligned} \tag{5.10.2}$$

When $y \in K_0(x_k) \subset K_{\epsilon_k}(x_k)$ we have

$$\nabla f(x_k) + y \in \nabla f(x_k) + K_{\epsilon_k}(x_k) + C_{\epsilon_k}(x_k)$$

and so by (3.7), $s_k(\nabla f(x_k) + y - s_k) \geqslant 0$ and consequently $(\nabla f(x_k) + y)s_k \geqslant |s_k|^2 > 0$. Combining this with (5.10.2) we have

$$F'(x_k; -s_k) \leqslant -|s_k|^2 < 0, \tag{5.10.3}$$

completing the proof.

From the first half of Lemma 5.10 we have the following corollary.

5.11. Corollary. *The number $\bar{\alpha}_k$ defined in Step 5 of Algorithm 4 is positive.*

The next two lemmas explain the choice of α_k and z_k in Step 5 of the algorithm.

5.12. Lemma. *Let $s_k \neq 0$ and define φ on $[0, \bar{\alpha}_k]$ by $\varphi(\alpha) = F(x_k - \alpha s_k)$. If $\bar{\alpha}_k$ is not a minimizer of φ on $[0, \bar{\alpha}_k]$, then z_k satisfying Step 5 of Algorithm 4 exists.*

Proof. By Lemma 5.10, $\varphi'(0) = F'(x_k; -s_k) < 0$, so that there is some $\alpha \in (0, \bar{\alpha}_k]$ such that $\varphi(\alpha) < \varphi(0)$. Since we have hypothesized that $\bar{\alpha}_k$ does not minimize φ, there exists $\alpha_k \in (0, \bar{\alpha}_k)$ minimizing φ over $[0, \bar{\alpha}_k]$. Set $y = x_k - \alpha_k s_k$. There exists $\varepsilon > 0$ such that $F(y) \leqslant F(y + \lambda s_k)$ for $|\lambda| \leqslant \varepsilon$. It follows that

$$\{F(y + \lambda s_k) - F(y)\}/\lambda \geqslant 0 \qquad (5.12.1)$$

and

$$\{F(y - \lambda s_k) - F(y)\}/\lambda \geqslant 0, \qquad (5.12.2)$$

$0 < \lambda \leqslant \varepsilon$. Since F is convex, the directional derivatives $F'(y; s_k)$ and $F'(y; -s_k)$ both exist, and from (5.12.1) and (5.12.2) we conclude that $F'(y; s_k) \geqslant 0$ and $F'(y; -s_k) \geqslant 0$. Using Corollary 5.5

$$\begin{aligned} F'(y; \pm s_k) &= f'(y; \pm s_k) + v'(y; \pm s_k) \\ &= \pm \nabla f(y)\, s_k + \max\{\pm u s_k \mid u \in K_0(y)\}. \end{aligned}$$

Since $K_0(y)$ is compact, there exist $u, w \in K_0(y)$ such that

$$\nabla f(y)\, s_k + u s_k = F'(y; s_k) \geqslant 0$$

and

$$\nabla f(y)\, s_k + w s_k = -F'(y; -s_k) \leqslant 0.$$

So for an appropriately chosen convex combination h of u and w we have $h \in K_0(y)$ and

$$\nabla f(y)\, s_k + h s_k = 0.$$

Taking $z_k = \nabla f(y) + h \in \nabla f(y) + K_0(y)$ satisfies the requirement in Step 5 of the algorithm.

The number α_k determined in Step 5 of Algorithm 4 has the following property.

5.13. LEMMA. *Let $s_k \neq 0$ and φ be as in the previous lemma. Then α_k is the unique minimizer of φ on $[0, \bar{\alpha}_k]$. Moreover, α_k is positive.*

Proof. Since $F'(x_k; -s_k) < 0$ by Lemma 5.10, the conclusion that $\alpha_k > 0$ follows immediately once we show that α_k minimizes φ over $[0, \bar{\alpha}_k]$. Uniqueness of this minimizer follows from the strict convexity of F.

If z_k satisfying Step 5 of the algorithm cannot be found, then by Lemma 5.12, $\bar{\alpha}_k$ minimizes φ over $[0, \bar{\alpha}_k]$, and in Step 5 we would have set $\alpha_k = \bar{\alpha}_k$ so that the lemma is verified. So consider the case when $\alpha_k \in (0, \bar{\alpha}_k]$ is located such that an appropriate vector z_k exists. Set now $y = x_k - \alpha_k s_k$. Since $z_k \in \nabla f(y) + K_o(y) \subset \partial F(y)$, for any $\alpha \in [0, \bar{\alpha}_k]$ we have by the subgradient inequality that

$$\varphi(\alpha) = F(x_k - \alpha s_k) \geqslant F(y) + (\alpha_k - \alpha)\, z_k s_k = F(y) = \varphi(\alpha_k),$$

so that α_k minimizes φ over $[0, \bar{\alpha}_k]$, completing the proof of the lemma.

5.14. COROLLARY. *Let $s_k \neq 0$ and $x_{k+1} = x_k - \alpha_k s_k$ as in Step 6 of Algorithm 4. Then $F(x_{k+1}) < F(x_k)$.*

Proof. This follows from Lemma 5.13 and the observation that $F'(x_k; -s_k) < 0$.

The lemmas stated up to this point prove that the algorithm is feasible and that F decreases at each iteration. We now turn to lemmas leading to a convergence proof.

5.15. LEMMA. *Let $\bar{x} \in X$ be the minimizer of F and $\hat{x}$ be a cluster point of the sequence (x_k). Then (x_k) converges to $\bar{x}$.*

Proof. Let $\hat{x}$ be any cluster point of (x_k). We have $F(\hat{x}) = F(\bar{x})$. Since F is strictly convex, $\bar{x}$ is the unique minimizer, so that $\hat{x} = \bar{x}$. Hence the sequence (x_k) has a unique cluster point $\bar{x}$. Due to the compactness of X, we now conclude that the sequence (x_k) converges to $\bar{x}$.

5.16. LEMMA. *Let 0 be a cluster point of the sequence (s_k). Then the sequence (x_k) converges to $\bar{x}$, the minimizer of F.*

Proof. We pass to corresponding subsequences $(s_{k'})$ and $(x_{k'})$ such that $s_{k'} \to 0$ and $x_{k'} \to \hat{x} \in X$. We shall show that $\hat{x}$ minimizes F, so that by the previous lemma $x_k \to \bar{x}$. Since the restriction of F to X is continuous from within X, to prove that $\hat{x}$ is a minimizer of F, it suffices to show that $F(y) \geqslant F(\hat{x})$ for all $y \in \operatorname{rel\,int} X$. Let $y \in \operatorname{rel\,int} X$. For all $i \in I_0(\hat{x})$, $a_i y < b_i = a_i \hat{x}$, and so for k' sufficiently large

$$a_i(y - x_{k'}) < 0 \qquad \text{for all } i \in I_0(\hat{x}). \tag{5.16.1}$$

Since $s_{k'} \to 0$ and $\varepsilon_{k'} \leqslant |s_{k'}|^2$, $\varepsilon_{k'} \to 0$, so that for k' sufficiently large

$$I_{\epsilon_{k'}}(x_{k'}) \subset I_0(\hat{x}). \tag{5.16.2}$$

Now there exist $u_{k'} \in K_{\epsilon_{k'}}(x_{k'})$ and $w_{k'} \in C_{\epsilon_{k'}}(x_{k'})$ such that $s_{k'} = \nabla f(x_{k'}) + u_{k'} + w_{k'}$. By Lemma 5.1, $K_{\epsilon_k}(x_k) \subset \partial_{\epsilon_k} v(x_k)$, so that

$$v(y) - v(x_{k'}) \geqslant u_{k'}(y - x_{k'}) - \varepsilon_{k'}. \tag{5.16.3}$$

Since f is convex, it follows that

$$\begin{aligned} F(y) - F(x_{k'}) &\geqslant \nabla f(x_{k'})(y - x_{k'}) + u_{k'}(y - x_{k'}) - \varepsilon_{k'} \\ &= s_{k'}(y - x_{k'}) - w_{k'}(y - x_{k'}) - \varepsilon_{k'}. \end{aligned} \tag{5.16.4}$$

Assume that k' is large enough that (5.16.1) and (5.16.2) hold. Since $w_{k'}$ belongs to the convex cone generated by $\{a_i \mid i \in I_{\epsilon_{k'}}(x_{k'})\}$, in view of (5.16.1) and (5.16.2) we have $w_{k'}(y - x_{k'}) \leqslant 0$, and so from (5.16.4)

$$F(y) - F(x_{k'}) \geqslant s_{k'}(y - x_{k'}) - \varepsilon_{k'} \tag{5.16.5}$$

when k' is sufficiently large. In the limit (5.16.5) gives

$$F(y) - F(\hat{x}) \geqslant 0,$$

proving the lemma.

5.17. Lemma. *If* 0 *is a cluster point of the sequence* (ε_k) *defined in Algorithm* 4, *then the sequence* (x_k) *converges to* $\bar{x}$, *the minimizer of* F.

Proof. Passing to corresponding subsequences $(\varepsilon_{k'})$ and $(x_{k'})$, we may assume that $\varepsilon_{k'} \to 0$ and $x_{k'} \to \hat{x} \in X$. By Lemma 5.9, Step 4 of the algorithm is executed finitely often per iteration, and hence the subsequence $(\varepsilon_{k'})$ can be chosen such that

$$|y_\epsilon|^2 \leqslant \varepsilon, \qquad y_\epsilon = N[\nabla f(x_{k'}) + K_\epsilon(x_{k'}) + C_\epsilon(x_{k'})]$$

and

$$|y_{\epsilon/2}|^2 > \varepsilon/2 = \varepsilon_{k'}, \qquad y_{\epsilon/2} = N[\nabla f(x_{k'}) + K_{\epsilon/2}(x_{k'}) + C_{\epsilon/2}(x_{k'})].$$

From these we see that $|y_{2\epsilon_{k'}}|^2 \leqslant 2\varepsilon_{k'}$, showing that $y_{2\epsilon_{k'}} \to 0$. We now repeat the proof of Lemma 5.16, replacing $s_{k'}$ with $y_{2\epsilon_{k'}}$, concluding that $x_k \to \bar{x}$.

5.18. Lemma. *The sequence* (s_k) *is bounded.*

Proof. Note that

$$\nabla f(x_k) + K_0(x_k) \subset \nabla f(x_k) + K_{\epsilon_k}(x_k) + C_{\epsilon_k}(x_k),$$

so that

$$|s_k| \leqslant |\nabla f(x_k)| + |N[K_0(x_k)]|. \tag{5.18.1}$$

$K_0(x_k)$ is one of a finite number of possible polytopes, so that there is an upper bound on $|N[K_0(x_k)]|$ independent of k. As f is of class C^1 on the compact set X, the right-hand side of (5.18.1) is bounded, proving the lemma.

5.19. LEMMA. *If the sequence* (s_k) *is bounded away from* 0, *then the sequence* (α_k) *converges to zero.*

Proof. Suppose that (s_k) is bounded away from 0 and that $\alpha_k \nrightarrow 0$. Since $\alpha_k |s_k|$ is bounded above by the diameter of X and (s_k) is bounded away from 0, (α_k) is bounded. Given this and the compactness of X, we can pass to corresponding subsequences $(s_{k'})$, $(\alpha_{k'})$ and $(x_{k'})$ such that $s_{k'} \to s \neq 0$, $\alpha_{k'} \to \alpha > 0$ and $x_{k'} \to x \in X$. By Corollary 5.14, the sequence $(F(x_k))$ is monotone decreasing, so that all of its subsequences have the same limit, namely, $F(x)$. In particular, $F(x_{k'+1}) \to F(x)$; but $x_{k'+1} = x_{k'} - \alpha_{k'} s_{k'} \to x - \alpha s$, so that

$$F(x - \alpha s) = F(x). \tag{5.19.1}$$

Since F is convex and $F(x_{k'} - \alpha_{k'} s_{k'}) \leqslant F(x_{k'} - \lambda s_{k'})$ for all $\lambda \in [0, \bar{\alpha}_{k'}]$, we have

$$F(x_{k'} - \alpha_{k'} s_{k'}) \leqslant F(x_{k'} - \alpha_{k'} s_{k'}/2) \leqslant F(x_{k'})$$

and so in the limit

$$F(x - \alpha s) \leqslant F(x - \alpha s/2) \leqslant F(x). \tag{5.19.2}$$

Since $\alpha > 0$ and $s \neq 0$, (5.19.1) and (5.19.2) taken together contradict the strict convexity of F.

5.20. LEMMA. *Let the sequence* (ε_k) *defined in Algorithm* 4 *be such that there exists* $\varepsilon \geqslant 0$ *satisfying* $\varepsilon_k \geqslant \varepsilon$ *for every* k. *For any index* i, *the inequality*

$$b_i - a_i x_k \leqslant \varepsilon \tag{5.20.1}$$

implies the inequality

$$b_i - a_i x_k \leqslant b_i - a_i x_{k+1}. \tag{5.20.2}$$

Proof. If (5.20.1) holds, then $i \in I_{\epsilon_k}(x_k)$, and so $a_i \in C_{\epsilon_k}(x_k)$. As noted in the course of proving Lemma 5.10, we get $a_i s_k \geqslant 0$. Since $x_{k+1} = x_k - \alpha_k s_k$, we now see that (5.20.2) holds.

5.21. LEMMA. *Assume that the following hold.*

(i) *The sequence* (ε_k) *in Algorithm* 4 *is such that there exists* $\varepsilon > 0$ *with* $\varepsilon_k \geqslant \varepsilon$ *for all* k.

(ii) *The sequence* (α_k) *converges to* 0.

(iii) *Some subsequence* $(x_{k'})$ *of* (x_k) *converges to the point* x.

Then there exists a subsequence of $(x_{k'})$, *again denoted* $(x_{k'})$, *such that* $I_0(x_{k'}) = I_0(x)$ *for every index* k'.

Proof. Assume that (i), (ii) and (iii) hold. Since the index sets $I_0(x_{k'})$ are subsets of the finite set $\{1,..., m\}$, we can pass to a subsequence of $(x_{k'})$, again denoted $(x_{k'})$, such that for some subset I of $\{1,..., m\}$ we have $I_0(x_{k'}) = I$ for all k'. We will show that $I_0(x) = I$. If $i \in I$, then $a_i x_{k'} = b_i$ for all k', so that in the limit $a_i x = b_i$. Therefore $I \subset I_0(x)$. Now suppose that $i \in I_0(x) \backslash I$. We derive a contradiction. Since $x_{k+1} = x_k - \alpha_k s_k$, with (s_k) shown bounded in Lemma 5.18 and $\alpha_k \to 0$, we see that $|x_{k+1} - x_k| \to 0$ as $k \to \infty$. Hence there exists k_0 such that

$$a_i(x_{k+1} - x_k) < \varepsilon/2 \qquad \text{for all } k \geqslant k_0. \tag{5.21.1}$$

Choose $p \geqslant k_0$ such that $I_0(x_p) = I$ and

$$\hat{\varepsilon} = b_i - a_i x_p < \varepsilon. \tag{5.21.2}$$

Such an index p exists because $i \in I_0(x)$ implies that $b_i - a_i x_{k'} \to b_i - a_i x = 0$. Also $\hat{\varepsilon} > 0$, since $i \notin I$. Let q be the first index such that $q > p$ and

$$b_i - a_i x_q \leqslant \hat{\varepsilon}/2. \tag{5.21.3}$$

Now by (5.21.1), (5.21.2) and (5.21.3)

$$b_i - a_i x_{q-1} = b_i - a_i x_q + a_i(x_q - x_{q-1}) < \hat{\varepsilon}/2 + \varepsilon/2 < \varepsilon,$$

and so by Lemma 5.20

$$b_i - a_i x_{q-1} \leqslant b_i - a_i x_q \leqslant \hat{\varepsilon}/2. \tag{5.21.4}$$

Note that $q - 1 \geqslant p$. If $q - 1 = p$, then (5.21.4) contradicts (5.21.2). If $q - 1 > p$, then (5.21.4) contradicts the choice of q as the smallest index greater than p such that (5.21.3) holds.

5.22. COROLLARY. *Suppose that the following hold.*

(i) *There exists* $\varepsilon > 0$ *such that* $\varepsilon_k \geqslant \varepsilon$ *for all* k.

(ii) *There exists* $\eta > 0$ *such that* $|s_k| \geqslant \eta$ *for all* k.

(iii) *Some subsequence* $(x_{k'})$ *of* (x_k) *converges to* x.

Then there is a subsequence of $(x_{k'})$, *again denoted* $(x_{k'})$, *such that* $I_0(x_{k'})=I_0(x)$ *for all* k'.

Proof. Hypothesis (ii) of this corollary implies hypothesis (ii) of Lemma 5.21 by Lemma 5.19.

We are at last prepared to prove the convergence of our algorithm.

5.23. THEOREM. *Algorithm* 4 *generates either a terminating sequence whose last term is the minimizer of problem* (P) *or an infinite sequence converging to the minimizer of problem* (P).

Proof. In view of Lemma 5.8, we need only consider the case in which Algorithm 4 generates an infinite sequence (x_k). In this case, $s_k \neq 0$ for every k. We assume that (x_k) fails to converge to the solution of (P) and derive a contradiction.

By Lemma 5.16 we may suppose that there exists $\eta > 0$ such that $|s_k| \geqslant \eta$ for all k. Similarly, by Lemma 5.17 we may assume that there exists $\varepsilon > 0$ such that $\varepsilon_k \geqslant \varepsilon$ for all k. By Lemma 5.18, (s_k) is bounded. Also X is compact. Hence we may pass to a subsequence (k') of positive integers such that

$$x_{k'} \to x \in X \qquad \text{and} \qquad s_{k'} \to s \neq 0. \tag{5.23.1}$$

From Step 6 of the algorithm, $x_{k'+1}=x_{k'}-\alpha_{k'}s_{k'}$. Since $|s_k| \geqslant \eta$ for all k, Lemma 5.19 ensures that $\alpha_k \to 0$, and so $x_{k'+1} \to x$. Passing to a subsequence of (k'), again denoted (k'), we may suppose that there exist sets I, J and J' of indices such that

$$I_{\epsilon_{k'}}(x_{k'})=I, \qquad J_{\epsilon_{k'}}(x_{k'})=J, \qquad J_0(x_{k'+1})=J' \tag{5.23.2}$$

for all k'. We assert that $J' \subset J$. Since $x_{k'+1} \to x$, $J_0(x_{k'+1}) \subset J_0(x)$ for k' sufficiently large. Moreover, since $x_{k'} \to x$ and $\varepsilon > 0$, by Lemma 5.6, $J_0(x) \subset J_\epsilon(x_{k'})$ for k' large enough. As $\varepsilon_{k'} \geqslant \varepsilon$, we must have $J_\epsilon(x_{k'}) \subset J_{\epsilon_{k'}}(x_{k'})$. Thus for k' sufficiently large

$$J'=J_0(x_{k'+1}) \subset J_0(x) \subset J_\epsilon(x_{k'}) \subset J_{\epsilon_{k'}}(x_{k'})=J,$$

and so $J' \subset J$.

Using Corollary 5.22, since $x_{k'+1} \to x$ we can pass to yet another subsequence, again denoted (k'), such that

$$I_0(x_{k'})=I_0(x)=I_0(x_{k'+1}) \tag{5.23.3}$$

for all k'. Now set

$$K=\operatorname{conv}\{g_j \mid j \in J\} \qquad \text{and} \qquad C=\operatorname{cone}\{a_i \mid i \in I\}. \tag{5.23.4}$$

Due to (5.23.2) and (5.23.4), we see that for all k'

$$K_{\epsilon_{k'}}(x_{k'}) = K \qquad \text{and} \qquad C_{\epsilon_{k'}}(x_{k'}) = C. \tag{5.23.5}$$

From (5.23.3) we deduce that $\alpha_{k'} < \bar{\alpha}_{k'}$ for each k'; for if $\alpha_{k'} = \bar{\alpha}_{k'}$, some constraint inactive at $x_{k'}$ becomes active at $x_{k'+1}$ and so $I_0(x_{k'}) \neq I_0(x_{k'+1})$. Thus for each k', the vector $z_{k'}$ specified in Step 5 of Algorithm 4 must exist, i.e.,

$$z_{k'} \in \nabla f(x_{k'+1}) + K_0(x_{k'+1})$$

and

$$z_{k'} s_{k'} = 0. \tag{5.23.6}$$

Taking into account (5.23.2) and (5.23.4) we have

$$K_0(x_{k'+1}) = \operatorname{conv}\{ g_j \mid j \in J' \} \subset \operatorname{conv}\{ g_j \mid j \in J \} = K$$

and so

$$z_{k'} \in \nabla f(x_{k'+1}) + K. \tag{5.23.7}$$

Since $\nabla f(x_{k'+1}) \to \nabla f(x)$ and K is compact, by passing to still another subsequence (k') and applying (5.23.7) we may assume that there exists $z \in \nabla f(x) + K$ such that $z_{k'} \to z$.

From Steps 2 and 3 of the algorithm, we have that

$$s_{k'} = N[\nabla f(x_{k'}) + K_{\epsilon_{k'}}(x_{k'}) + C_{\epsilon_{k'}}(x_{k'})]$$

and so, in view of (5.23.2) and (5.23.4),

$$s_{k'} = N[\nabla f(x_{k'}) + K + C].$$

Since $x_{k'} \to x$ and $s_{k'} \to s$, it follows easily that

$$s = N[\nabla f(x) + K + C]. \tag{5.23.8}$$

Now $z \in \nabla f(x) + K \subset \nabla f(x) + K + C$, and so by (3.7) we have $s(z - s) \geqslant 0$, i.e.,

$$zs \geqslant |s|^2. \tag{5.23.9}$$

As $|s_{k'}| \geqslant \eta$ for all k', clearly (5.23.9) implies that

$$zs \geqslant \eta^2 > 0. \tag{5.23.10}$$

On the other hand, letting $k' \to \infty$ in (5.23.6) yields

$$zs = 0, \tag{5.23.11}$$

contradicting (5.23.10). Thus our assumption that (x_k) fails to converge to the solution of (P) cannot be valid. The proof that the algorithm generates a sequence converging to the optimal solution is now complete.

REFERENCES

1. D. P. BERTSEKAS AND S. MITTER, A descent method for optimization problems with nondifferentiable cost functionals, *SIAM J. Control.* **11** (1973), 637–652.
2. C. LEMARECHAL, An extension of Davidon methods to nondifferentiable problems, *Math. Programming Stud.* **3** (1975), 95–109.
3. E. POLAK, On the convergence of optimization algorithms, *Rev. Fr. Inform. Rech. Operation* (16–R1) (1969), 17–34.
4. E. POLAK, "Computational Methods in Optimization," Academic Press, New York, 1971.
5. R. T. ROCKAFELLAR, "Convex Analysis," Princeton Univ. Press, Princeton, N. J., 1970.
6. J. B. ROSEN, The gradient projection method for nonlinear programming, Part I: Linear constraints, *J. SIAM* **8** (1960), 181–217.
7. P. RUBIN, Implementation of a subgradient projection algorithm, in preparation.
8. P. RUBIN AND V. P. SREEDHARAN, Existence of equilibria in a PIES-type economic model, in preparation.
9. V. P. SREEDHARAN, Least-squares algorithms for finding solutions of overdetermined linear equations which minimize error in an abstract norm, *Numer. Math.* **17** (1971), 387–401.
10. V. P. SREEDHARAN, Least-squares algorithms for finding solutions of overdetermined systems of linear equations which minimize error in a smooth strictly convex norm, *J. Approx. Theory* **8** (1973), 46–61.
11. V. P. SREEDHARAN, Note on a subgradient projection algorithm, *in* "Approximation Theory III" (E. W. Cheney, Ed.), pp. 857–864, Academic Press, New York 1980.
12. P. WOLFE, A method of conjugate subgradients for minimizing nondifferentiable functions, *Math. Programming Stud.* **3** (1975), 145–173.