JOURNAL OF APPROXIMATION THEORY 51, 27–46 (1987)

ϵ-Subgradient Projection Algorithm

V. P. SREEDHARAN

Department of Mathematics, Michigan State University, East Lansing, Michigan 48824, U.S.A.

Communicated by E. W. Cheney

Received August 20, 1984; revised June 25, 1985

Using only certain easily computable ε-subgradients, an implementable convergent algorithm for finding the minimizer of a non-differentiable objective function subject to a finite number of linear constraints in d-dimensional space is given. The particular objective function consists of the pointwise maximum of a finite system of pseudoconvex functions. At each iteration cycle certain projections are computed. The negatives of these directions are feasible directions of strict descent for the objective function. The convergence of the algorithm is proved. The algorithm has also been numerically tested. © 1987 Academic Press, Inc.

1. INTRODUCTION

This paper presents an implementable algorithm for minimizing a certain type of non-differentiable pseudoconvex function subject to a finite collection of linear constraints in $\mathbb{R}^d$. We actually encountered this particular form of the problem in a certain stochastic logistics model, though the abstract problem is an obvious generalization of the problem in [11] and others. The motivation and the derivation of this model will appear in Myhre and Sreedharan [5]. The approach we take is motivated by Sreedharan [11, 12, 13], Dem'yanov and Malozemov [1], and Rosen [9]. The algorithm proposed here avoids the possibility of "jamming," a situation where the generated sequence clusters or even converges to non-optimal points. The algorithm is a generalization of one in Sreedharan [11], and so may be viewed as constrained counterparts of algorithms of Lemarechal [3] and Wolfe [14]. We take the point of view of facing non-differentiability directly. Thus we arrive at lower dimensional subproblems, instead of a single high dimensional problem.

In Section 5 we have included some development of the required optimality criteria generalizing some criteria in [1], due to our inability in finding these in the literature. We prove the convergence of the algorithm.

0021-9045/87 $3.00

Copyright © 1987 by Academic Press, Inc.
All rights of reproduction in any form reserved.

The algorithm was also tested numerically as applied to the logistics model (See [5]).

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 is denoted by $|\cdot|$. Let $a_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$ be given for $i = 1,..., m$. The feasibility set

$$X = \{x \in \mathbb{R}^d | a_i x \leqslant b_i,\ i = 1,..., m\}, \tag{2.1}$$

is assumed to be nonempty but not necessarily bounded. Before we describe the objective function let us remind the classic definition of a pseudoconvex function defined in a neighborhood of X. See Mangasarian [4] or Ponstein [8]. A function f defined in a neighborhood of X is said to be *pseudoconvex* on X iff f is differentiable in a neighborhood of X and

$$\forall x, y \in X, \qquad \nabla f(x)(y - x) \geqslant 0 \Rightarrow f(y) \geqslant f(x).$$

f is said to be *strictly pseudoconvex* on X iff f is differentiable in a neighborhood of X and

$$\forall x, y \in X, \qquad x \neq y, \qquad \nabla f(x)(y - x) \geqslant 0 \Rightarrow f(y) > f(x).$$

Suppose that we are given pseudoconvex, continuously differentiable functions $f_1,..., f_r$ on X and let

$$f(x) = \max\{f_j(x) | 1 \leqslant j \leqslant r\}. \tag{2.2}$$

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

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

We shall refer to this as problem (P).

3. Notation

Let $x \in X$ and $\varepsilon \geqslant 0$. We define the sets of indices $I_\varepsilon(x)$ and $J_\varepsilon(x)$ by

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

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

Note that

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

and

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

$I_0(x)$ is an enumeration of the binding (i.e., active) constraints at x and $J_0(x)$ is the index set of binding maximands at x. So $I_\varepsilon(x)$ and $J_\varepsilon(x)$ are the indices of ε-binding (i.e., almost active) constraints and maximands, respectively. With the help of these index sets we define the following convex subsets of $\mathbb{R}^d$:

$$C_\varepsilon(x) = \operatorname{cone}\{a_i \mid i \in I_\varepsilon(x)\}, \tag{3.5}$$

and

$$K_\varepsilon(x) = \operatorname{conv}\{\nabla f_j(x) \mid j \in J_\varepsilon(x)\}. \tag{3.6}$$

Here and throughout this paper we denote by cone S the convex cone generated by set S with apex at 0 and by conv S the convex hull of the set S.

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

$$ax \geqslant |a|^2, \qquad \forall x \in S. \tag{3.7}$$

For purposes of proving that the algorithm of the next section is convergent, we shall make an additional assumption on the objective function f, namely that f be *coercive on* X, the feasibility set. One says that f is *coercive on* X iff $x_k \in X$, $|x_k| \to \infty$ implies that $f(x_k) \to \infty$. Note that this condition is automatically satisfied, if X is bounded.

4. Algorithm

In this section we present our ε-subgradient projection algorithm for solving problem (P). Discussions of the different steps of the algorithm are given in Sections 4.2 to 4.7.

4.1 Algorithm

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

Step 1. Compute $y_0 = N[K_0(x_k) + C_0(x_k)]$.

Step 2. If $y_0 = 0$, STOP; x_k is a solution of problem (P). Otherwise proceed.

Step 3. Compute $y_\varepsilon = N[K_\varepsilon(x_k) + C_\varepsilon(x_k)]$.

Step 4. If $|y_\varepsilon|^2 > \varepsilon$ set $\varepsilon_k = \varepsilon$, $s_k = y_\varepsilon$ and GO TO Step 6; otherwise proceed.

Step 5. Replace ε by $\varepsilon/2$ and RETURN to Step 3.

Step 6. If there exists i such that $a_i s_k < 0$, define $\bar{\alpha}_k$ by

$$\bar{\alpha}_k = -\max_{1 \leqslant i \leqslant m} \left\{ \frac{b_i - a_i x_k}{a_i s_k} \,\middle|\, a_i s_k < 0 \right\}.$$

If no such i exists, set $\bar{\alpha}_k = \infty$.

Step 7. Find $\alpha_k \in [0, \bar{\alpha}_k]$, $\alpha_k < \infty$ such that there exists

$$z_k \in K_0(x_k - \alpha_k s_k) \text{ with } z_k s_k = 0.$$

If no such α_k exists, set $\alpha_k = \bar{\alpha}_k$. (It will be shown in Lemma 6.5 that $0 < \alpha_k < \infty$.)

Step 8. Define $x_{k+1} = x_k - \alpha_k s_k$. Increment k by 1 and RETURN to Step 1.

4.2. Step 5 can be replaced by the statement: Replace ε by ε/δ_q, where (δ_q) is any sequence of numbers such that $\delta_q > 1$, $\forall q$ and is uniformly bounded away from 1. Thus we found it convenient to set $\delta_q = 10$, $\forall q$ in similar problems (see [6] and [10].)

4.3. One may also wish to reset $\varepsilon = \varepsilon_0$ at the end of Step 2 during the early cycles of the algorithm. This should avoid taking small steps when not "near" the optimal solution. After these iterations we revert back, i.e. set $\varepsilon = \varepsilon_0$ in Step 0, but instead of an arbitrary $x_0 \in X$ we take x_0 to be the last available x_k and proceed from Step 0 onwards without any change from the algorithm as given. This alteration has only an insignificant effect on the proof of convergence of the algorithm.

4.4. Obviously, in practive Step 2 will be replaceced by the statement: STOP, if $|y_0| \leqslant \eta$, where $\eta > 0$ is a stopping rule parameter.

4.5. *Quadratic Programs*

Steps 1 and 3 call for computing y_ε the projection of the origin on the set $K_\varepsilon(x_k) + C_\varepsilon(x_k)$, i.e., find $y_\varepsilon \in K_\varepsilon(x_k) + C_\varepsilon(x_k)$ of least norm. This can be accomplished by a special quadratic program. In fact, we seek

$$\begin{gathered} \lambda_i \geqslant 0, \quad \mu_j \geqslant 0, \quad i \in I_\varepsilon(x_k), \quad j \in J_\varepsilon(x_k), \\ \sum_j \mu_j = 1, \\ \left| \sum_{i \in I_\varepsilon(x_k)} \lambda_i a_i + \sum_{j \in J_\varepsilon(x_k)} \mu_j \nabla f_j(x_k) \right|^2 \text{ (min).} \end{gathered} \tag{4.5.1}$$

We can rewrite (4.5.1) more compactly as follows. Let card $I_\varepsilon(x_k) = p$ and card $J_\varepsilon(x_k) = q$. Let e be the row vector of dimension q, all of whose components are unity. Finally, let M be the $m \times (p+q)$ matrix whose first p columns are a_i, $i \in I_\varepsilon(x_k)$ and the remaining q columns are $\nabla f_j(x_k)$, $j \in J_\varepsilon(x_k)$. Now (4.5.1) is the same as

$$\begin{gathered} [0, e]\, u = 1, \qquad u \geqslant 0, \qquad u \in \mathbb{R}^{p+q}, \\ uM'Mu \text{ (min),} \end{gathered} \tag{4.5.2}$$

where M' denotes the transpose of M.

Rubin [10] has the same $C_\varepsilon(x_k)$ as we have here, but not the same $K_\varepsilon(x_k)$. The $K_\varepsilon(x_k)$ in Owens [6] is similar to ours here. So any of the methods used in [6] or [10], or even some variations thereof like those in Lawson and Hanson [2] are applicable. Refer to Myhre and Sreedharan [5] for details of computational experience.

4.6. *The Line Search*

Let us now consider the determination of α_k in Step 7 of Algorithm 4.1. By Lemma 6.5 (few pages hence) an alternate definition of α_k is

$$\alpha_k = \arg \min_{0 \leqslant \alpha \leqslant \bar{\alpha}_k} f(x_k - \alpha s_k), \tag{4.6.1}$$

where it is shown that $0 < \alpha_k < \infty$, whenever $s_k \neq 0$. The calculation of α_k using (4.6.1) can be accomplished by employing a Golden Section search technique as applied to a univariate unimodal function, e.g., the IMSL subroutine ZXGSP. See Owens [6], Myhre and Sreedharan [5] for further details.

Note that α_k is not unique, in general. Any α_k found in Step 7 is sufficient to guarantee convergence of (x_k), in the sense that any cluster point of (x_k) is a minimizer of problem (P). Yet, in practice one would generally pick the smallest α_k in case of multiple candidates. But if each f_j is in addition strictly pseudoconvex then α_k is unique and then the whole sequence (x_k) converges to the unique minimizer of problem (P).

4.7. Due to inequality (7.1.1) of Lemma 7.1 below, we will see that members of $K_\varepsilon(x_k)$ may rightly be called ε-subgradients of f in analogy with the convex case. Moreover, s_k of Step 4 of the algorithm is the least distance vector between the sets $-K_\varepsilon(x_k)$ and $C_\varepsilon(x_k)$.

5. Optimality Conditions

In this section we extend certain optimality conditions that apply to a convex program. In fact, Theorem 5.9 below, the main result of this section, is a generalization of Theorem 3.1 in Chapter 4 of Dem'yanov and Malozemov [1]. We need this result to show that the stopping criterion in Step 2 of Algorithm 4.1 is well chosen. Throughout this section we do not require f to be coercive on X, though we shall require this hypothesis in our algorithm and for its convergence proof given in Sections 6 and 7. We first record some properties of the index sets I_ε and J_ε for use in this and subsequent sections.

5.1 Lemma. *To each $x \in X$ and $\varepsilon \geqslant 0$, there is a neighborhood V of x such that*

$$I_\varepsilon(y) \subset I_\varepsilon(x), \qquad \forall y \in V \cap X, \tag{5.1.1}$$

and

$$J_\varepsilon(y) \subset J_\varepsilon(x), \qquad \forall y \in V \cap X. \tag{5.1.2}$$

5.2 Lemma. *Given $x \in X$, there exists $\rho > 0$ such that*

$$I_\varepsilon(x) = I_0(x), \qquad J_\varepsilon(x) = J_0(x) \quad \text{for} \quad 0 \leqslant \varepsilon \leqslant \rho. \tag{5.2.1}$$

5.3 Lemma. *Let $x_k \in X$ and $(x_{k'})$ a subsequence of (x_k) such that $x_{k'} \to x \in X$. Assume that $(\varepsilon_{k'})$ is a sequence such that $\varepsilon_{k'} \downarrow 0$. Then $I_{\varepsilon_{k'}}(x_{k'}) \subset I_0(x)$ and $J_{\varepsilon_{k'}}(x_{k'}) \subset J_0(x)$, for all sufficiently large k'.*

5.4 Lemma. *Let $x_k \in X$ and $\varepsilon_k \downarrow \varepsilon > 0$. Let $(x_{k'})$ be a subsequence of (x_k) such $x_{k'} \to x \in X$. Then $I_0(x) \subset I_{\varepsilon_{k'}}(x_{k'})$ and $J_0(x) \subset J_{\varepsilon_{k'}}(x_{k'})$, for all k' sufficiently large.*

5.5 LEMMA. *Given $x \in X$ and $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$J_0(x) \subset J_\varepsilon(y), \text{ whenever } |x - y| < \delta, \; y \in X.$$

The above lemmas are Lemmas 5.1 through 5.5 of [12], where complete proofs are provided. Strictly speaking, the statement regarding the sets $J_0(x)$ and $J_{\varepsilon_k}(x_{k'})$ in Lemma 5.3 above, does not appear in Lemma 5.3 of [12]. Its proof is entirely analogous to that concerning the sets $I_0(x)$ and $I_{\varepsilon_{k'}}(x_{k'})$ in Lemma 5.3 of [12].

It is standard practice to call points in X *feasible points*. Due to the convexity of X, $h \in \mathbb{R}^d$ is a *feasible direction* at $x \in X$ iff there exists some $\alpha > 0$ such that $x + \alpha h \in X$, i.e., $x + \alpha h$ is a feasible point. The following lemma characterizing the feasible directions at a feasible point is well known.

5.6 LEMMA. *Let $x \in X$. Then h is a feasible direction at x iff $a_i h \leqslant 0$, $\forall i \in I_0(x)$.*

We will need the following lemma in the proof of Theorem 5.9 and also in subsequent sections.

5.7 LEMMA. *Let h be a feasible direction at the feasible point x, and let*

$$\nabla f_j(x)\, h < 0, \qquad \forall j \in J_0(x). \tag{5.7.1}$$

Then there exists $\delta > 0$ such that

$$f(x + \alpha h) < f(x), \qquad \forall \alpha \in (0, \delta]. \tag{5.7.2}$$

Proof. Since h is a feasible direction at x, due to Lemma 5.6 and the convexity of X, there exists $\delta_1 > 0$ such that $x + \alpha h \in X$, $0 \leqslant \alpha \leqslant \delta_1$. Because of (5.7.1), there exists $\delta_2 > 0$ such that for $0 < \alpha \leqslant \delta_2$,

$$f_j(x + \alpha h) < f_j(x), \qquad \forall j \in J_0(x). \tag{5.7.3}$$

Also there exists a neighborhood V of x, such that

$$f_j(y) < f(x), \qquad \forall j \notin J_0(x), \quad \forall y \in V \cap X. \tag{5.7.4}$$

Due to (5.7.3) and (5.7.4), there exists $\delta > 0$, $\delta \leqslant \min(\delta_1, \delta_2)$ such that

$$f_j(x + \alpha h) < f(x), \qquad \forall j = 1, ..., r, \qquad \forall \alpha \in (0, \delta].$$

By the definition of f then (5.7.2) follows, completing the proof of the lemma.

5.8 THEOREM. *Every local minimizer of the problem* (P) *is a global minimizer. Moreover, if evry* f_i *is strictly pseudoconvex, then at most one global minimizer exists.*

Proof. Let $\bar{x} \in X$ be a local minimizer of f on X and let $h \neq 0$ be such that $\bar{x} + h \in X$. We shall show that $f(\bar{x}+h) \geqslant f(\bar{x})$, i.e., $\bar{x}$ is a global minimizer. Due to the convexity of X, $\bar{x} + \alpha h \in X$ for $0 \leqslant \alpha \leqslant 1$. Since $\bar{x}$ is local minimizer, there exists δ, $0 < \delta \leqslant 1$ such that

$$f(\bar{x} + \alpha h) \geqslant f(\bar{x}), \qquad \forall \alpha \in [0, \delta]. \tag{5.8.1}$$

By Lemma 5.1 we may, by reducing δ if necessary, assume that

$$J_0(\bar{x} + \alpha h) \subset J_0(\bar{x}), \qquad I_0(\bar{x} + \alpha h) \subset I_0(\bar{x}), \qquad \forall \alpha \in [0, \delta]. \tag{5.8.2}$$

From (5.8.1) and (5.8.2) we see that

$$\max\{f_j(\bar{x} + \alpha h) \mid j \in J_0(\bar{x})\} \geqslant f(\bar{x}), \qquad \forall \alpha \in [0, \delta].$$

This shows that there exists $j \in J_0(\bar{x})$ such that

$$f_j(\bar{x} + \alpha h) \geqslant f_j(\bar{x}), \qquad \forall \alpha \in [0, \delta].$$

Due to the differentiability of f_j at $\bar{x}$, this yields

$$\nabla f_j(\bar{x})\, h \geqslant 0.$$

Now invoking the fact that f_j is pseudoconvex we get

$$f_j(\bar{x} + h) \geqslant f_j(\bar{x}). \tag{5.8.3}$$

So

$$f(\bar{x} + h) \geqslant f_j(\bar{x} + h) \geqslant f_j(\bar{x}) = f(\bar{x}), \tag{5.8.4}$$

which proves that $\bar{x}$ is a global minimizer.

If every f_j is strictly pseudoconvex, then strict inequality prevails in (5.8.3) and therefore in (5.8.4) also, proving that $\bar{x}$ is the unique global minimizer. The proof of the theorem is now complete.

5.9 THEOREM. $\bar{x} \in X$ *solves the problem* (P) iff $0 \in K_0(\bar{x}) + C_0(\bar{x})$.

Proof. In view of the previous theorem it is sufficient to show that $\bar{x}$ is a local minimizer iff $0 \in K_0(\bar{x}) + C_0(\bar{x})$.

"*Only if*" *part*: Assume that $0 \notin K_0(\bar{x}) + C_0(\bar{x})$. We will show that $\bar{x}$ is not a local minimizer. Since $0 \notin K_0(\bar{x}) + C_0(\bar{x})$, there exists $h \in \mathbb{R}^d$ such that

$$uh + vh < 0, \qquad \forall u \in K_0(\bar{x}), \qquad \forall v \in C_0(\bar{x}).$$

This yields that

$$\nabla f_j(\bar{x})\, h < 0, \qquad \forall j \in J_0(\bar{x}) \tag{5.9.1}$$

and

$$a_i h \leqslant 0, \qquad \forall i \in C_0(\bar{x}). \tag{5.9.2}$$

By Lemma 5.6 and (5.9.2) we see that h is a nonzero feasible direction at x, whereas by Lemma 5.7 we see that there exists $\delta > 0$ such that

$$f(\bar{x} + \alpha h) < f(\bar{x}), \qquad \forall \alpha \in (0, \delta].$$

So $\bar{x}$ is not a constrained local minimizer of f, and the proof of the "Only if" part is complete.

"*If*" *part*: We will show that if $\bar{x}$ is not a local minimizer, then $0 \notin K_0(\bar{x}) + C_0(\bar{x})$. Using Lemma 5.1, let us choose a neighborhood V of $\bar{x}$ such that

$$J_0(x) \subset J_0(\bar{x}), \qquad I_0(x) \subset I_0(\bar{x}), \qquad \forall x \in V \cap X. \tag{5.9.3}$$

Also since $\bar{x}$ is not a local minimizer of f, to each $\varepsilon > 0$, there exists a feasible direction h, $0 < |h| \leqslant \varepsilon$, such that

$$f(\bar{x} + h) < f(\bar{x}). \tag{5.9.4}$$

From (5.9.3) and (5.9.4) we see that there exists a nonzero feasible direction h, such that $\bar{x} + h \in V$ and (5.9.4) holds. So

$$\max\{f_j(\bar{x} + h) | j \in J_0(\bar{x})\} = f(\bar{x} + h) < f(\bar{x}).$$

Since $f_j(\bar{x}) = f(\bar{x})$, $\forall j \in J_0(\bar{x})$, this yields the inequality

$$f_j(\bar{x} + h) < f_j(\bar{x}), \qquad \forall j \in J_0(\bar{x}).$$

But since each f_j is pseudoconvex,

$$\nabla f_j(\bar{x})\, h < 0, \qquad \forall j \in J_0(\bar{x}). \tag{5.9.5}$$

Also since h is a feasible direction at $\bar{x}$, by Lemma 5.6

$$a_i h \leqslant 0, \qquad \forall i \in I_0(\bar{x}). \tag{5.9.6}$$

Now if $0 \in K_0(\bar{x}) + C_0(\bar{x})$, then there exists, $\lambda_j \geqslant 0$, $\mu_i \geqslant 0$, $\Sigma \lambda_j = 1$, $j \in J_0(\bar{x})$, $i \in I_0(\bar{x})$ such that

$$0 = \sum_{j \in J_0(\bar{x})} \lambda_j \nabla f_j(\bar{x}) + \sum_{i \in I_0(\bar{x})} \mu_i a_i. \tag{5.9.7}$$

Since some λ_j is positive, we find, upon taking the inner product of both sides of Eq. (5.9.7) with h and using (5.9.5) and (5.9.6) that $0<0$, which is absurd. The theorem is now completely proven.

5.10 COROLLARY. *Assume further that each f_j in the above theorem is strictly pseudoconvex. Then $\bar{x}\in X$ is the unique solution of (P)* iff $0\in K_0(\bar{x})+C_0(\bar{x})$.

Proof. This follows immediately from Theorems 5.9 and 5.8.

As mentioned earlier, Theorem 5.9 with the assumption that each f_j is convex is stated and proved in [1].

6. FEASIBILITY OF THE ALGORITHM

In this section we show that the various steps of the algorithm are well-defined and implementable. In this and the subsequent section we will have the standing assumption that f is coercive on X and that $f_1,...,f_r$ are continuously differentiable, pseudoconvex functions on X.

6.1 LEMMA. *The stopping criterion in Step* 2 *of Algorithm* 4.1 *is well chosen.*

Proof. If $y_0=0$, then $0\in K_0(x_k)+C_0(x_k)$. *Theorem* 5.9 *now shows that x_k is a minimizer of f.*

6.2 LEMMA. *Step* 5 *of the Algorithm* 4.1 *is not executed infinitely often in any one iteration.*

Proof. If Step 5 is executed infinitely often in a certain iteration, then the index k remains unchanged from that iteration onwards. By Lemma 5.2, there exists arbitrarily small $\varepsilon>0$ such that $I_\varepsilon(x_k)=I_0(x_k)$ and $J_\varepsilon(x_k)=J_0(x_k)$. For such ε then $K_\varepsilon(x_k)+C_\varepsilon(x_k)=K_0(x_k)+C_0(x_k)$. This implies that $y_0=y_\varepsilon$ for arbitrarily small $\varepsilon>0$. Also $\varepsilon\downarrow 0$, since Step 5 is executed indefinitely. So $y_\varepsilon\to 0$, which shows that $y_0=0$. But in this case we would not have reached Step 5 at all, a contradiction.

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

Proof. Since

$$a_i+s_k\in K_{\varepsilon_k}(x_k)+C_{\varepsilon_k}(x_k),\qquad \forall i\in I_{\varepsilon_k}(x_k),$$

and

$$\nabla f_j(x_k)\in K_{\varepsilon_k}(x_k)+C_{\varepsilon_k}(x_k),\qquad \forall j\in J_{\varepsilon_k}(x_k),$$

from the least norm inequality (3.7) we see that

$$(a_i + s_k)\, s_k \geqslant |s_k|^2, \qquad \forall i \in I_{\varepsilon_k}(x_k), \tag{6.3.1}$$

and

$$\nabla f_j(x_k)\, s_k \geqslant |s_k|^2, \qquad \forall j \in J_{\varepsilon_k}(x_k). \tag{6.3.2}$$

Now $I_{\varepsilon_k}(x_k) \supset I_0(x_k)$ and $J_{\varepsilon_k}(x_k) \supset J_0(x_k)$, so that by (6.3.1) and (6.3.2) we get the inequalities

$$a_i s_k \geqslant 0, \qquad \forall i \in I_0(x_k), \tag{6.3.3}$$

and

$$\nabla f_j(x_k)\, s_k \geqslant |s_k|^2, \qquad \forall j \in J_0(s_k). \tag{6.3.4}$$

This shows that $-s_k$ is a feasible direction at x_k and that inequality (5.7.1) holds with $h = -s_k$ in the notation of Lemma 5.7. So by Lemma 5.7 there exists $\delta > 0$, such that $x_k - \alpha s_k \in X$ for $0 \leqslant \alpha \leqslant \delta$ and

$$f(x_k - \alpha s_k) < f(x_k), \qquad \forall \alpha \in (0, \delta].$$

The proof of the lemma is now complete.

6.4 LEMMA. *Let $s_k \neq 0$. The $\bar{\alpha}_k$ defined in Step* 6 *of Algorithm* 4.1 *has the property that*

$$\bar{\alpha}_k = \sup\{\alpha \in \mathbb{R} \mid x_k - \alpha s_k \in X\}. \tag{6.4.1}$$

Proof. Let us define δ by $\delta = \sup\{\alpha \in \mathbb{R} \mid x_k - \alpha s_k \in X\}$, then by Lemma 6.3 we see that there exists $\alpha > 0$ such that $x_k - \alpha s_k \in X$, so that $\delta > 0$, possibly $+\infty$. Since $x_k - \alpha s_k \in X$ iff $a_i x_k - \alpha a_i s_k \leqslant b_i$, $\forall i = 1, \ldots, m$, we now see the equivalence of δ with $\bar{\alpha}_k$ as defined in Step 6 of Algorithm 4.1.

6.5 LEMMA. *Let $s_k \neq 0$ and define φ on interval $\mathscr{I}$ by $\varphi(\alpha) = f(x_k - \alpha s_k)$, where $\mathscr{I} = [0, \infty)$ if $\bar{\alpha}_k = \infty$, and $\mathscr{I} = [0, \bar{\alpha}_k]$ if $\bar{\alpha}_k < \infty$. If $\bar{\alpha}_k = \infty$, or in case $\bar{\alpha}_k < \infty$ and $\bar{\alpha}_k$ is not a minimizer of φ on $\mathscr{I}$, then α_k, z_k satisfying Step* 7 *of Algorithm* 4.1 *exists. Moreover, if α_k, z_k satisfying Step* 7 *have been found then α_k is a minimizer of φ on $\mathscr{I}$.*

Proof. By Lemma 6.3, 0 is not a minimizer of φ on $\mathscr{I}$. The hypotheses of this lemma and the fact that f is coercive implies that there exists a minimizer $\alpha_k \in (0, \bar{\alpha}_k)$, i.e., there exists $\varepsilon > 0$ such that

$$f(y + \lambda s_k) \geqslant f(y), \qquad y = x_k - \alpha_k s_k, \quad |\lambda| \leqslant \varepsilon \tag{6.5.1}$$

By reducing $\varepsilon > 0$ if necessary (using Lemma 5.1), there exists $j_1, j_2 \in J_0(y)$, ($j_1 = j_2$ permitted) such that

$$f_{j_1}(y + \lambda s_k) \geqslant f_{j_1}(y), \qquad 0 \leqslant \lambda \leqslant \varepsilon,$$
$$f_{j_2}(y + \lambda s_k) \geqslant f_{j_2}(y), \qquad -\varepsilon \leqslant \lambda \leqslant 0.$$

From this we see that

$$\nabla f_{j_1}(y)\, s_k \geqslant 0 \qquad \text{and} \qquad \nabla f_{j_2}(y)\, s_k \leqslant 0.$$

We now take z_k to be an appropriate convex combination of $\nabla f_{j_1}(y)$ and $\nabla f_{j_2}(y)$ so that $z_k s_k = 0$. Since $j_1, j_2 \in J_0(y)$, $z_k \in K_0(y)$; completing the proof of existence of α_k and z_k in Step 7.

To prove the second part of the lemma, suppose that α_k, z_k satisfying Step 7 has been found. We shall show that α_k is positive, and that α_k is a minimizer of φ on $[0, \bar{\alpha}_k]$.

Let $y = x_k - \alpha_k s_k$. Then there exists $\lambda_j \geqslant 0$, $\sum \lambda_j = 1$, $j \in J_0(y)$ such that

$$z_k = \sum \lambda_j \nabla f_j(y).$$

Since $z_k s_k = 0$, there exists $j_1, j_2 \in J_0(y)$, ($j_1 = j_2$ permitted) such that

$$\nabla f_{j_1}(y)\, s_k \geqslant 0 \qquad \text{and} \qquad \nabla f_{j_2}(y)\, s_k \leqslant 0 \tag{6.5.2}$$

Since each f_j is pseudoconvex this implies that

$$f_{j_1}(y + \lambda s_k) \geqslant f_{j_1}(y), \qquad 0 \leqslant \lambda \leqslant \alpha_k, \tag{6.5.3}$$

and

$$f_{j_2}(y - \lambda s_k) \geqslant f_{j_2}(y), \qquad 0 \leqslant \lambda \leqslant \bar{\alpha}_k - \alpha_k, \tag{6.5.4}$$

if $\bar{\alpha}_k < \infty$. But if $\bar{\alpha}_k = \infty$, then

$$f_{j_2}(y - \lambda s_k) \geqslant f_{j_2}(y), \qquad 0 \leqslant \lambda < \infty. \tag{6.5.5}$$

Using the definition of f and recalling that because $j_1, j_2 \in J_0(y)$, $f_{j_1}(y) = f_{j_2}(y) = f(y)$, when $\bar{\alpha}_k < \infty$ we find from (6.5.3) and (6.5.4) that

$$f(y + \lambda s_k) \geqslant f(y), \qquad 0 \leqslant \lambda \leqslant \alpha_k, \tag{6.5.6}$$

and

$$f(y - \lambda s_k) \geqslant f(y), \qquad 0 \leqslant \lambda \leqslant \bar{\alpha}_k - \alpha_k. \tag{6.5.7}$$

When $\bar{\alpha}_k = \infty$, (6.5.7) is replaced by

$$f(y - \lambda s_k) \geqslant f(y), \qquad 0 \leqslant \lambda < \infty. \tag{6.5.8}$$

If $\alpha_k = 0$, then $y = x_k$ and (6.5.7) or (6.5.8), whichever is applicable, asserts that $f(x_k - \lambda s_k) \geqslant f(x_k)$, $0 \leqslant \lambda \leqslant \bar{\alpha}_k < \infty$ or $0 \leqslant \lambda < \infty$, as the case may be. This contradicts Lemma 6.3, where we proved that $-s_k$ is a feasible direction of strict descent at x_k. So α_k is positive.

Inequalities (6.5.6), (6.5.7), and (6.5.8) assert that if $\bar{\alpha}_k < \infty$, then

$$\varphi(\alpha) \geqslant \varphi(\alpha_k), \qquad 0 \leqslant \alpha \leqslant \bar{\alpha}_k;$$

whereas if $\bar{\alpha}_k = \infty$, then

$$\varphi(\alpha) \geqslant \varphi(\alpha_k), \qquad 0 \leqslant \alpha < \infty.$$

In other words, α_k is a minimizer of φ on $[0, \bar{\alpha}_k]$, if $\bar{\alpha}_k < \infty$; whereas if $\bar{\alpha}_k = \infty$, then α_k is a minimizer of φ on $[0, \infty)$.

6.6 COROLLARY. *α_k is positive and finite. Moreover, α_k is unique if each f_j is strictly pseudoconvex.*

6.7 LEMMA. *Let $s_k \neq 0$ and $x_{k+1} = x_k - \alpha_k s_k$ as in Step 8 of Algorithm 4.1. Then $f(x_{k+1}) < f(x_k)$.*

Proof. This is clear from Lemmas 6.3 and 6.5.

6.8 LEMMA. *The sequence (x_k) generated by Algorithm 4.1 is bounded. Moreover, f takes the same value $v = \lim_{k \to \infty} f(x_k)$ at all cluster points of (x_k).*

Proof. By Lemma 6.7, the sequence $(f(x_k))$ is clearly bounded from above. Since f is coercive on X, it follows that (x_k) is bounded. Since the sequence $(f(x_k))$ is monotone decreasing, all its subsequences converge to the same limit v. So, if $(x_{k'})$ is a subsequence of (x_k) such that $x_{k'} \to x$, then $f(x) = v$.

7. CONVERGENCE OF THE ALGORITHM

In the previous section we showed that the various steps in the algorithm are implementable and that f decreases at each iteration. We now turn to the task of proving that the algorithm converges to a solution of the problem in the sense that every cluster point of the generated sequence is a minimizer of f. The reason for assuming that f is coercive is to ensure that (x_k) has at least one cluster point as guaranteed by Lemma 6.8. In fact, any hypotheses on f and X which will do this is sufficient, say for example that the set $\{x \in X \mid f(x) \leqslant f(x_k)\}$ is bounded, for some k.

7.1 LEMMA. *Let $\varepsilon \geqslant 0$ and $u \in K_\varepsilon(x)$. Suppose that $x+h \in X$ and $uh \geqslant 0$. Then*

$$f(x+h) \geqslant f(x) - \varepsilon. \tag{7.1.1}$$

Proof. Since there exists $\lambda_j \geqslant 0$, $\sum \lambda_j = 1$, $j \in J_\varepsilon(x)$ such that

$$u = \sum_{j \in J_\varepsilon(x)} \lambda_j \nabla f_j(x),$$

we see that there is some $j \in J_\varepsilon(x)$ with the property $\nabla f_j(x) \geqslant 0$. Since f_j is pseudoconvex, $f_j(x+h) \geqslant f_j(x) \geqslant f(x) - \varepsilon$. Hence $f(x+h) \geqslant f(x) - \varepsilon$.

7.2 COROLLARY. *Let $u \in K_0(x)$, $x+h \in X$ and $uh \geqslant 0$. Then $f(x+h) \geqslant f(x)$.*

7.3 LEMMA. *Let 0 be a cluster point of the sequence (s_k) and $\bar{x}$ any cluster point of (x_k). Then $\bar{x}$ is a minimizer of f.*

Proof. We pass to corresponding subsequences $(s_{k'})$ and $(x_{k'})$ such that $s_{k'} \to 0$ and $x_{k'} \to x \in X$. We shall show that both x and $\bar{x}$ are minimizers of f.

Let $y \in X$ be arbitrary. Then

$$a_i y \leqslant b_i = a_i x, \qquad \forall i \in I_0(x).$$

So

$$a_i(y-x) \leqslant 0, \qquad \forall i \in I_0(x). \tag{7.3.1}$$

Let us write

$$s_{k'} = u_{k'} + w_{k'}, \quad u_{k'} \in K_{\varepsilon_{k'}}(x_{k'}), \quad w_{k'} \in C_{\varepsilon_{k'}}(x_{k'}). \tag{7.3.2}$$

Note that $\varepsilon_{k'} \downarrow 0$, so that by Lemma 5.3, for all sufficiently large k' we have

$$I_{\varepsilon_{k'}}(x_{k'}) \subset I_0(x) \qquad \text{and} \qquad J_{\varepsilon_{k'}}(x_{k'}) \subset J_0(x). \tag{7.3.3}$$

The first inclusion in (7.3.3) yields the containment

$$C_{\varepsilon_{k'}}(x_{k'}) \subset C_0(x), \tag{7.3.4}$$

whereas the second inclusion in (7.3.3) gives the containment

$$K_{\varepsilon_{k'}}(x_{k'}) \subset \operatorname{conv}\{\nabla f_j(x_{k'}) \mid j \in J_0(x)\}. \tag{7.3.5}$$

So, there exists $\lambda_{j,k'} \geqslant 0$, $\sum_j \lambda_{j,k'} = 1$, $j \in J_0(x)$ such that

$$u_{k'} = \sum_{j \in J_0(x)} \lambda_{j,k'} \nabla f_j(x_{k'}).$$

By passing to a further subsequence, again denoted by (k'), we can require $\lambda_{j,k'} \to \lambda_j$ as $k' \to \infty$, for each $j \in J_0(x)$. By the continuity of ∇f_j at x, then

$$u_{k'} \to u = \sum_{j \in J_0(x)} \lambda_j \nabla f_j(x) \in K_0(x). \tag{7.3.6}$$

By (7.3.4), $w_{k'} \in C_0(x)$ so that by (7.3.1) we see that

$$w_{k'}(y - x) \leqslant 0, \qquad \forall k'. \tag{7.3.7}$$

Note that in case $I_0(x)$ is empty then $C_0(x) = \{0\}$ and every $w_{k'} = 0$, so (7.3.7) holds in this case also. By (7.3.2) we now see that

$$\begin{aligned} u_{k'}(y-x) &= s_{k'}(y-x) - w_{k'}(y-x) \\ &\geqslant s_{k'}(y-x), \qquad \text{by} \quad (7.3.7). \end{aligned}$$

Allowing $k' \to \infty$, since $s_{k'} \to 0$ we arrive at the relation $u(y-x) \geqslant 0$. Since $u \in K_0(x)$, by Corollary 7.2 we now conclude that $f(x) \leqslant f(y)$. But since x and $\bar{x}$ are both cluster points of $(x_{k'})$, by Lemma 6.8 we see that $f(x) = f(\bar{x})$, completing the proof of the lemma.

7.4 LEMMA. *If the sequence (ε_k) defined in Algorithm 4.1 converges to zero and $\bar{x}$ is any cluster point of $(x_{k'})$ then $\bar{x}$ is a minimizer of f.*

Proof. By Lemma 6.2, Step 5 of the algorithm is executed finitely often in each iteration. Hence a subsequence $(\varepsilon_{k'})$ of (ε_k) can be found such that

$$\varepsilon_{k'+1} = \varepsilon_{k'}/2 \qquad \text{and} \qquad |y_{\varepsilon_{k'}}|^2 \leqslant \varepsilon_{k'},$$

where y_ε was defined in Step 3 of the algorithm. Since $\varepsilon_k \to 0$, $y_{\varepsilon_{k'}} \to 0$. We replace all the occurrences of s_k, in the proof of the previous lemma by $y_{\varepsilon_{k'}}$ and repeat the reasoning therein to see the validity of the present lemma.

7.5 LEMMA. *The sequence (s_k) is bounded.*

Proof. Since $K_{\varepsilon_k}(x_k) + C_{\varepsilon_k}(x_k) \supset K_0(x_k)$,

$$\begin{aligned} |s_k| &= |N[K_{\varepsilon_k}(x_k) + C_{\varepsilon_k}(x_k)]|, \\ &\leqslant |N[K_0(x_k)]|, \\ &\leqslant \max\{|\nabla f_j(x_k)| \mid j \in J_0(x_k)\}, \\ &\leqslant \max\{|\nabla f_j(x_k)| \mid 1 \leqslant j \leqslant r\}, \\ &\leqslant \max_{x \in X_0} \max_{1 \leqslant j \leqslant r} |\nabla f_j(x)|, \end{aligned}$$

where X_0 is the closure of the set $\{x_0, x_1, x_2, \ldots\}$. The right-hand side of the above inequality is finite due to the fact that each f_j is of class C^1 on X_0, which by virtue of Lemma 6.8 is compact.

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

Proof. Suppose that $\alpha_k \nrightarrow 0$. Since $\alpha_k |s_k| = |x_{k+1} - x_k|$, by Lemma 6.8 we see that $\alpha_k |s_k|$ is bounded above. But (s_k) is bounded away from zero and hence the sequence (α_k) is bounded. Let us then 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 $x_{k'} \to x \in X$. Now

$$x_{k'+1} = x_{k'} - \alpha_{k'} s_{k'} \to x - \alpha s.$$

So $x - \alpha s$ and x are both cluster points of the sequence (x_k). By Lemma 6.8, we then have

$$f(x - \alpha s) = f(x). \tag{7.6.1}$$

Since (s_k) is bounded away from zero, form Algorithm 4.1 we see that there exists $\varepsilon > 0$ such that $\varepsilon_k = \varepsilon$ for all sufficiently large k. Passing to a further subsequence of (k'), again denoted by (k'), we may assume that

$$I_\varepsilon(x_{k'}) = I \qquad \text{and} \qquad J_\varepsilon(x_{k'}) = J, \qquad \forall k'. \tag{7.6.2}$$

Recall that

$$s_{k'} = N[K_\varepsilon(x_{k'}) + C_\varepsilon(x_{k'})]. \tag{7.6.3}$$

Since $s_{k'} \in K_\varepsilon(x_{k'}) + C_\varepsilon(x_{k'})$, we see that $a_i + s_{k'} \in K_\varepsilon(x_{k'}) + C_\varepsilon(x_{k'})$, $\forall i \in I$. Also $\nabla f_j(x_{k'}) \in K_\varepsilon(x_{k'}) \subset K_\varepsilon(x_{k'}) + C_\varepsilon(x_{k'})$, $\forall j \in J$. So, by (3.7) we get

$$(a_i + s_{k'})\, s_{k'} \geqslant |s_{k'}|^2, \qquad \forall i \in I, \tag{7.6.4}$$

and

$$\nabla f_j(x_{k'})\, s_{k'} \geqslant |s_{k'}|^2, \qquad \forall j \in J. \tag{7.6.5}$$

Allowing $k' \to \infty$ in (7.6.4) and (7.6.5) we arrive at the inequalities

$$a_i s \geqslant 0, \qquad \forall i \in I, \tag{7.6.6}$$

and

$$\nabla f_j(x)\, s \geqslant |s|^2, \qquad \forall j \in J. \tag{7.6.7}$$

But now $\varepsilon > 0$ and $x_{k'} \to x$. So by Lemma 5.4 and (7.6.2) we see that $I_0(x) \subset I$ and $J_0(x) \subset J$. From (7.6.6) and (7.6.7) we get the inequalities

$$a_i s \geqslant 0, \qquad \forall i \in I_0(x), \tag{7.6.8}$$

and

$$\nabla f_j(x)\, s \geqslant |s|^2, \qquad \forall j \in J_0(x). \tag{7.6.9}$$

In view of Lemma 5.6 and inequality (7.6.8), $-s$ is a nonzero feasible direction at x. This fact combined with inequality (7.6.9) and Lemma 5.7 shows that there exists $\delta > 0$ such that

$$f(x - \lambda x) < f(x), \qquad \forall \lambda \in (0, \delta]. \tag{7.6.10}$$

By Step 7 of Algorithm 4.1 and Lemma 6.5

$$f(x_{k'} - \alpha_{k'} s_{k'}) \leqslant f(x_{k'} - \lambda s_{k'}), \qquad \forall \lambda \in [0, \bar{\alpha}_{k'}] \cap [0, \infty). \tag{7.6.11}$$

Note that $\bar{\alpha}_{k'} \geqslant \alpha_{k'}$ and so $(\bar{\alpha}_{k'})$ is also bounded away from zero. Hence from (7.6.10) and (7.6.11) we conclude that there exists λ, $0 < \lambda \leqslant \delta$ satisfying (7.6.11) and the condition

$$f(x_{k'} - \alpha_{k'} s_{k'}) \leqslant f(x_{k'} - \lambda s_{k'}), \quad \forall k'. \tag{7.6.12}$$

Allowing $k' \to \infty$ in (7.6.12) yields the inequality

$$f(x - \alpha s) \leqslant f(x - \lambda s),$$

which by virtue of (7.6.10) shows that $f(x - \alpha s) < f(x)$, contradicting (7.6.1). So we have to conclude that (α_k) converges to zero, and the proof of the lemma is complete.

7.7. It is of some interest to remark that, if all the f_j's were known to be strictly pseudoconvex, then a simpler argument may be used to complete the proof of previous lemma after having arrived at equation (7.6.1). For purposes of clarity, we isolate this fact as a lemma.

7.8 Lemma. *Assume that each f_j is strictly pseudoconvex on X and that x, $x - h \in X$, $h \neq 0$, with $f(x) = f(x - h)$. Then there exists $\mu \in (0, 1)$ such that $f(x - \mu h) < f(x)$.*

Proof. If not, $f(x - \mu h) \geqslant f(x)$, $\forall \mu \in [0, 1]$. By Lemma 5.1 we can choose $\delta \in (0, 1)$ such that $J_0(x - \mu h) \subset J_0(x)$, $\forall \mu \in [0, \delta]$. Since $f(x - \mu h) \geqslant f(x)$, there exists $j \in J_0(x)$ such that $f_j(x - \mu h) \geqslant f_j(x)$, which implies that $\nabla f_j(x)\, h \leqslant 0$. Due to the strict pseudoconvexity of f_j, we now have $f_j(x - h) > f_j(x)$, resulting in a contradiction so that the lemma follows.

7.9. Under the stronger assumptions in the above lemma, we simply point out that, due to the definition of α_k,

$$f(x_{k'} - \alpha_{k'} s_{k'}) \leqslant f(x_{k'} - \lambda s_{k'}), \qquad \forall \lambda \in [0, \bar{\alpha}_{k'}] \cap [0, \infty). \qquad (7.9.1)$$

By Lemma 7.8 we can choose μ, $0 < \mu < 1$ such that $f(x - \mu\alpha s) < f(x)$. Then by (7.9.1)

$$f(x_{k'} - \alpha_{k'} s_{k'}) \leqslant f(x_{k'} - \mu\alpha_{k'} s_{k'}).$$

Allowing $k' \to \infty$, we get $f(x - \alpha s) \leqslant f(x - \mu\alpha s)$ and so $f(x - \alpha s) < f(x)$, which contradicts (7.6.1).

7.10 LEMMA. *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. This is Corollary 5.22 in [11] and follows immediately from Lemma 5.21 in [11].

7.11 THEOREM. *Algorithm* 4.1 *generates either a terminating sequence whose last term is a minimizer of problem* (*P*), *or an infinite sequence such that every cluster point of this sequence is a minimizer of problem* (P).

Proof. In view of Lemma 6.1 we need only consider the case in which Algorithm 4.1 generates an infinite sequence (x_k). In this case $s_k \neq 0$ for every k. We intend to show that 0 is a cluster point of (s_k), so that by Lemma 7.3 the proof of the theorem would then be complete. With a view of arriving at a contradiction let us assume that there exists $\eta > 0$ such that $|s_k| \geqslant \eta$ for every k. In view of Lemma 7.4, we can also assume that $\varepsilon_k = \varepsilon$, for every k. The sequences (x_k) and (s_k) are bounded by virtue of Lemmas 6.8 and 7.5, respectively. So we may pass to corresponding convergent subsequences such that $x_{k'} \to x \in X$ and $s_{k'} \to s \neq 0$. By Lemma 7.6, $(\alpha_{k'})$ converges to zero and since $x_{k'+1} = x_{k'} - \alpha_{k'} s_{k'}$, we find that $x_{k'+1} \to x$. Passing to a subsequence of (k'), again denoted by (k'), we may suppose that there exist index sets I, J, and J' such that

$$I_\varepsilon(x_{k'}) = I, \qquad J_\varepsilon(x_{k'}) = J, \qquad J_0(x_{k'+1}) = J' \qquad (7.11.1)$$

for all k'. By Lemmas 5.1 and 5.4 we see that $J_0(x_{k'+1}) \subset J_0(x) \subset J_\varepsilon(x_{k'})$, for sufficiently large k'. In view of (7.11.1), we therefore have $J' \subset J$.

Now using Lemma 7.10 twice, since $x_{k'} \to x$ and $x_{k'+1} \to x$, we can find yet another subsequence, as usual denoted again by (k'), such that

$$I_0(x_{k'}) = I_0(x) = I_0(x_{k'+1}), \tag{7.11.2}$$

for all k'. From (7.11.2) we deduce that $\alpha_{k'} < \bar{\alpha}_{k'}$ for every k'; for if $\alpha_{k'} = \bar{\alpha}_{k'}$, some nonbinding constraint at $x_{k'}$ becomes binding 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 7 of the algorithm exists, i.e.,

$$z_{k'} \in K_0(x_{k'+1}) \qquad \text{and} \qquad z_{k'} s_{k'} = 0. \tag{7.11.3}$$

Now

$$\begin{aligned} z_{k'} \in K_0(x_{k'+1}) &= \operatorname{conv}\{\nabla f_j(x_{k'+1}) \mid j \in J'\}, \\ &\subset \operatorname{conv}\{\nabla f_j(x_{k'+1}) \mid j \in J\}. \end{aligned} \tag{7.11.4}$$

So there exist $\lambda_{j,k'} \geqslant 0$, $\sum_j \lambda_{j,k'} = 1$, $j \in J$ such that

$$z_{k'} = \sum_{j \in J} \lambda_{j,k'} \nabla f_j(x_{k'+1}). \tag{7.11.5}$$

By passing to yet another subsequence, denoted again by (k'), we can require $\lambda_{j,k'} \to \lambda_j$ for every $j \in J$. Let us define z and $\hat{z}_{k'}$ by

$$z = \sum_{j \in J} \lambda_j \nabla f_j(x), \tag{7.11.6}$$

and

$$\hat{z}_{k'} = \sum_{j \in J} \lambda_j \nabla f_j(x_{k'}). \tag{7.11.7}$$

Since $(x_{k'})$ and $(x_{k'+1})$ converge to x, by the continuity of ∇f_j at x, we see that both the sequences $(z_{k'})$ and $(\hat{z}_{k'})$ converge to z. Observe that

$$\hat{z}_{k'} \in K_\varepsilon(x_{k'}) \subset K_\varepsilon(x_{k'}) + C_\varepsilon(x_{k'}) \tag{7.11.8}$$

and by Step 3 of Algorithm 4.1

$$s_{k'} = N[K_\varepsilon(x_{k'}) + C_\varepsilon(x_{k'})]. \tag{7.11.9}$$

So, by (3.7), we get the inequality

$$\hat{z}_{k'} s_{k'} \geqslant |s_{k'}|^2, \qquad \forall k'. \tag{7.11.10}$$

Allowing $k' \to \infty$, we see that

$$zs \geqslant |s|^2 \geqslant \eta^2 > 0. \tag{7.11.11}$$

Finally, we allow $k' \to \infty$ in the second half of (7.11.3) to get the statement $zs = 0$, contradicting (7.11.11). So we conclude that 0 is a cluster point of (s_k), thus completing the proof of the theorem.

7.12 COROLLARY. *Suppose that each f_j is strictly pseudoconvex. Then in the non-terminating case the whole sequence (x_k) converges to $\bar{x}$ the minimizer of problem* (P).

Proof. Every cluster point $\bar{x}$ of (x_k) is a minimizer of f on X. But due to the strict pseudoconvexity of the f_j's, $\bar{x}$ is unique. So the sequence (x_k) has a unique cluster point $\bar{x}$ in X_0, the closure of $\{x_0, x_1, \ldots\}$, which is compact by Lemma 6.8. Hence (x_k) converges to $\bar{x}$.

ACKNOWLEDGMENT

Thanks are due to the referee for his helpful remarks.

REFERENCES

1. V. F. DEM'YANOV AND V. N. MALOZEMOV, "Introduction to Minimax," Wiley, New York, 1974.
2. C. L. LAWSON AND R. J. HANSON, "Solving Least Squares Problems," Prentice-Hall, Englewood Cliffs, N. Y., 1974.
3. C. LEMARECHAL, An extension of Davidon methods to nondifferentiable problems, *Math. Programming Stud.* **3** (1975), 95–109.
4. O. L. MANGASARIAN, "Nonlinear programming," McGraw-Hill, New York, 1969.
5. J. MYHRE AND V. P. SREEDHARAN, A nondifferentiable optimization algorithm for a logistic model, under preparation.
6. R. W. OWENS, Implementation of subgradient projection algorithm II, *Internat. J. of Comput. Math.* **16** (1984), 57–69.
7. E. POLAK, "Computational Methods in Optimization," Academic Press, New York, 1971.
8. J. PONSTEIN, Seven kinds of convexity, *SIAM Rev.* **9** (1967), 115–119.
9. J. B. ROSEN, The gradient projection method for nonlinear programming, Part I: Linear constraints, *J. SIAM* **8** (1960), 181–217.
10. P. RUBIN, Implementation of a subgradient projection algorithm, *Internat. J. Comput. Math.* **12** (1982), 321–328.
11. V. P. SREEDHARAN, A subgradient projection algorithm, *J. Approx. Theory* **35** (1982), 111–126.
12. V. P. SREEDHARAN, Subgradient projection algorithm II, *J. Approx. Theory* **41** (1984), 217–243.
13. V. P. SREEDHARAN, Extensions of subgradient projection algorithms, *J. Approx. Theory* **47** (1986), 228–239.
14. P. WOLFE, A method of conjugate subgradients for minimizing nondifferentiable functions, *Math. Programming Stud.* **3** (1975), 145–173.