Extensions of Subgradient Projection Algorithms

V. P. SREEDHARAN

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

Communicated by E. W. Cheney

Received September 4, 1984; revised January 30, 1985

The validity of the subgradient projections algorithms (V. P. SREEDHARAN, *J. Approx. Theory* **41** (1984), 217–243; **35** (1982), 111–126) is extended by removing part of the previous assumptions on the objective function. The appropriate modifications in the algorithms and their proofs of convergence are then given. © 1986 Academic Press, Inc.

1. INTRODUCTION

The purpose of this note is to extend the range of applicability of the subgradient projection algorithms for nonsmooth optimization [5, 4], with virtually no change in the algorithms or computations during their implementations. We use the notation and terminology employed in [5]. For the sake of brevity, we do not restate in full the algorithms from [5, 4], but refer to [5, 4] for their statements and all details.

In [5] we were given a nonempty, open, convex subset Ω in $\mathbb{R}^d$ along with the convex, differentiable functions f, g_i, v_j: $\Omega \to \mathbb{R}$; $i = 1,..., m$; $j = 1,..., r$. We considered the problem of minimizing $f(x) + v(x)$, subject to the constraints $g_i(x) \leqslant 0$, $i = 1,..., m$, where

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

and f was assumed to be strictly convex. We labeled this problem (P). The aim of adding the strictly convex f to v was to ensure that the objective function to be optimized was strictly convex, albeit nonsmooth. We indicate now how this assumption of strict convexity can be deleted. It turns out that we can take f to be identically zero, with trivial changes in the algorithms of [5, 4]. This includes the case, where f is convex and differentiable, but not strictly convex, for in this case we simply redefine v_j to be $v_j + f$ for every j.

2. Main Result

The algorithms of [5, 4] are applicable, with very minor changes, to their respective problems in [5, 4], when the objective function v is the pointwise upper envelope of a finite collection of differentiable convex functions as in [5], or that of a finite collection of affine functions as in [4], or a mixture of these, as outlined in Algorithm 7.1 of [5]. In other words, the objective function henceforth will be v, which is nonsmooth and convex, but not necessarily strictly convex. This is the raison d'etre of this note. In passing, we also observe that we can weaken the earlier assumption on the feasibility set

$$X = \{x \in \Omega \mid g_i(x) \leqslant 0, i = 1, ..., m\}.$$

Instead of requiring X to be compact, we can stipulate that the sublevel set

$$S_0 = \{x \in X \mid v(x) \leqslant v(x_0)\}$$

is bounded, where x_0 is the starting point for Algorithm 4.1 of [5]. This is equivalent to the assumption that v is coercive on X. Recall that one says $v: X \to \mathbb{R}$ is *coercive* on X iff $x_k \in X$, $|x_k| \to \infty \Rightarrow v(x_k) \to \infty$. We now state as a theorem a quick verification of the equivalence of these two notions.

2.1. Theorem. *Let v be a real, lower semi-continuous and convex function on the nonempty, closed, convex set $X \subset \mathbb{R}^d$ and let x_0 be any point in X. Then v is coercive on X iff the sublevel set*

$$S_0 = \{x \in X \mid v(x) \leqslant v(x_0)\}$$

is bounded.

Proof. We first prove the "if" part. By [6, Lemma 4.1.14] S_0 is bounded iff

$$S_n = \{x \in X \mid v(x) \leqslant v(x_0) + n\}$$

is bounded for every integer $n \geqslant 1$. If $x_k \in X$ is such that $|x_k| \to \infty$, then given integer n, there exists k_0 such that $x_k \notin S_n$, $\forall k \geqslant k_0$. This shows that $v(x_k) > v(x_0) + n$, $\forall k \geqslant k_0$. Since the choice of n was arbitrary, $v(x_k) \to \infty$.

To prove the "only if" part, assume that S_0 is unbounded. Then there exists $x_k \in S_0$ such that $|x_k| \to \infty$. On the other hand, since $v(x_k) \leqslant v(x_0)$, $\forall k$, $v(x_k) \nrightarrow \infty$, precluding the coercivity of v on X.

From the above theorem we see that the following corollary holds.

2.2 Corollary. *Let v be a real, lower semi-continuous, convex and coer-*

cive function on a nonempty, closed, convex subset X of $\mathbb{R}^d$. *Then the set of minimizers of v on X forms a nonempty, compact, convex subset of X.*

Proof. Let $x_0 \in X$ be arbitrary. By Theorem 2.1 the set $S_0 = \{x \in X \mid v(x) \leqslant v(x_0)\}$ is compact. So v restricted to S_0 has a minimizer $\bar{x} \in S_0 \subset X$. Clearly, $\bar{x}$ is a minimizer of v on X. Again by Theorem 2.1, the set $X^* = \{x \in X \mid v(x) \leqslant v(\bar{x})\}$ is nonempty, compact and convex, proving the corollary.

The minor alterations to Algorithm 4.1 of [5], which would allow us to handle the present more general problem, will now be given below, as Algorithm 2.3. The step numbers of Algorithm 2.3 correspond to the like numbered steps in Algorithm 4.1 of [5].

For the appropriate changes and results for the problem in [4], see the end of Section 4.

2.3 ALGORITHM.

Steps 1 *and* 2: Deleted. Start by setting $f = 0$, throughout.

Steps 3 *through* 7: Unchanged.

Step 8: Let

$$\Delta_k = \max_{1 \leqslant j \leqslant r} |\nabla v_j(x_k)|.$$

Determine u_k, as in Step 8 of Algorithm 4.1 of [5]. At the end of this step, we will have

$$M_k = \Delta_k |u_k|.$$

Note that, if we are at the stage of executing this step, then M_k is positive; the case $M_k = 0$ would occur only if $y_0 = 0$ in Step 4, in which event the algorithm would halt at the end of Step 4.

Steps 9 *through* 12: Unchanged.

Henceforth, when we refer to Algorithm 4.1 of [5], we will understand that the modifications indicated in Algorithm 2.3 have been incorporated into Algorithm 4.1 of [5]. The basic result of [5], Theorem 6.10 of [5] will be modified as follows.

2.4 THEOREM. *Algorithm 4.1 of* [5] *(i.e., Algorithm 2.3 above) generates either a terminating sequence whose last term is a minimizer of problem* (P), *or else an infinite sequence such that every cluster point of this sequence is a minimizer of problem* (P).

The proof of Theorem 6.10 of [5] may be repeated, essentially word for word, to prove Theorem 2.4 the main result of this paper. The only changes

needed are: put $f = 0$ throughout, and due to Lemma 3.4 below, assume that zero is not a cluster point of the sequence (Δ_k). Note that the proof in [5] already starts out by assuming that zero is not a cluster point of the sequence (s_k) and that the sequence (ε_k) does not converge to zero. Also the sequence (x_k) generated by the algorithm, appearing in Theorem 6.10 of [5], is bounded in view of Lemma 3.1 below. The assumption that the objective function was strictly convex was never used in the proof of the theorem. The strict convexity asumption entered only through the Lemmas 6.1, 6.2, 6.3, 6.6. and 6.8 of [5]. This assumption figured significantly in the proof of Lemma 6.8 of [5], which in turn played a key role in proving Theorem 6.10 of [5]. We shall now make the appropriate restatements of these lemmas and give proofs wherever necessary. It turns out that, under the current hypotheses, only Lemma 6.8 of [5] needs a new elaborate proof. Henceforth, we shall follow the convention that once a lemma of [5] has been restated and (or) reproved, then all uses of the original lemma in [5] will be replaced by the newer version. We emphasize that the newer version uses the weaker (and more natural) hypotheses, that the objective function of problem (P) is $v = \max\{v_j \mid 1 \leqslant j \leqslant r\}$, with v coercive on X and with each v_j convex and differentiable on a convex neighborhood of X. Note that convexity plus differentiability of v_j in a neighborhood of X implies that each v_j is continuously differentiable on X.

3. THE LEMMAS

The Lemmas 6.1, 6.2, 6.3, 6.4, 6.6, 6.8 and 6.9 of [5] will be replaced respectively by Lemmas 3.1, 3.2, 3.3, 3.5, 3.6, 3.7 and 3.8 of this section.

3.1. LEMMA. *The sequence (x_k) is bounded, and if some cluster point of (x_k) is a minimizer of F, then every cluster point of (x_k) is a minimizer of F.*

Proof. Recall that $F = v + \chi$, where χ is the indicator function of X. By Corollary 5.23 of [5], the sequence $(F(x_k))$ is monotone decreasing. Since $F = v$ on X, F is coercive on X, and so (x_k) is bounded. Also, since the sequence $(F(x_k))$ is monotone decreasing, all its subsequences converge to the same limit $L = \lim F(x_k)$. So for every cluster point x of (x_k), we have $F(x) = L$, which implies the lemma.

3.2 LEMMA. *Let zero be a cluster point of the sequence (s_k). Then every cluster point $\bar{x}$ of (x_k) is a minimizer of F.*

Proof. Using Lemma 3.1, we pass to a subsequence (k'), such that $s_{k'} \to 0$ and $x_{k'} \to \hat{x} \in X$. The proof of Lemma 6.2 of [5] now shows that $\hat{x}$ is a minimizer of F. Since $\hat{x}$ and $\bar{x}$ are both cluster points of (x_k), by

Lemma 3.1 we conclude that $\bar{x}$ is a minimizer of F, completing the proof of the lemma.

3.3 LEMMA. *If the sequence* (ε_k) *defined in Algorithm* 4.1 *of* [5] *converges to zero, then every cluster point* $\bar{x}$ *of* (x_k) *is a minimizer of* F.

Proof. This proof is the same as the proof of Lemma 6.3 of [5], except that, having found $y_{\varepsilon_{k'}}$ as in [5], we replace the occurances of $(s_{k'})$ in the proof of Lemma 3.2 by $y_{\varepsilon_{k'}}$, and repeat the proof of Lemma 3.2 to complete the proof.

The next lemma pertains to Step 8 of Algorithm 2.3.

3.4 LEMMA. *If zero is a cluster point of the sequence* $(\varDelta_k)$, *then every cluster point* $\bar{x}$ *of* (x_k) *is a minimizer of* F.

Proof. Using Lemma 3.1, we pass to a subsequence (k'), such that $\varDelta_{k'} \to 0$ and $x_{k'} \to x \in X$. We see that $\nabla v_j(x) = 0$, $\forall j$. As in Lemma 5.10 of [5], this implies that x is a minimizer of F. Then by Lemma 3.1 every cluster point $\bar{x}$ of (x_k) is a minimizer of F, completing the proof of the lemma.

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

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

$$|s_k| = |N[K_{\varepsilon_k}(x_k) + C_{\varepsilon_k}(x_k)]|$$

$$\leqslant |N[K_0(x_k)]|$$

$$\leqslant \max\{|\nabla v_j(x_k)| \mid j \in J_0(x_k)\}$$

$$\leqslant \max\{|\nabla v_j(x_k)| \mid 1 \leqslant j \leqslant r\}$$

$$\leqslant \max_{x \in X_0} \max_{1 \leqslant j \leqslant r} |\nabla v_j(x)|$$

where X_0 is the closure of the set $\{x_0, x_1, x_2, \dots\}$. The right hand side of the above inequality is finite, since each v_j is of class C^1 on X, and X_0 is compact by Lemma 3.1.

3.6. LEMMA. *Let the sequences* $(|s_k|)$ *and* $(\varDelta_k)$ *be bounded away from zero. Then the sequences* (t_k) *and* (α_k) *are both bounded. Moreover,* (t_k) *is bounded away from zero.*

Proof. Let

$$\delta = \inf\{\varDelta_k \mid k \geqslant 0\}.$$

Then $\delta > 0$ and by Step 8 of Algorithm 2.3, we see that

$$M_k = \Delta_k \, |u_k| \geqslant \delta \, |u_k|.$$

Inequality (6.6.1) of [5] now follows, and as in [5] we see that (t_k) is a bounded sequence. Following the proof of Lemma 6.6 of [5] we see that (t_k) is bounded away from zero.

Now let X_0 denote the closure of the set $\{x_0, x_1, x_2,...\}$. We have

$$\alpha_k \, |t_k| = |x_{k+1} - x_k| \leqslant \mathrm{dia}(X_0),$$

where $\mathrm{dia}(X_0)$ is the diameter of the set X_0. By Lemma 3.1, $\mathrm{dia}(X_0)$ is finite. Since (t_k) is bounded away from zero, we conclude that the positive sequence (α_k) is also bounded, completing the proof of the lemma.

The next lemma, which corresponds to Lemma 6.8 of [5], has its assertion unchanged but its proof is drastically different.

3.7 LEMMA. *Let* (ε_k), *etc., be as in Algorithm 4.1 of* [5]. *Suppose that the sequences* $(|s_k|)$ *and* (Δ_k) *are both bounded away from zero and that there exists* $\varepsilon > 0$ *and* k_0 *such that* $\varepsilon_k = \varepsilon$, $\forall k \geqslant k_0$. *Then the sequence* (α_k) *converges to zero.*

Proof. Assume that (α_k) does not converge to zero. We shall derive a contradiction. By Lemma 3.6, there exists a subsequence $(\alpha_{k'})$ of (α_k) such that $\alpha_{k'} \to \alpha > 0$. Due to Lemma 3.1 and the boundedness of (s_k) (Lemma 3.5), we can pass to a further subsequence of (k'), again denoted by (k'), such that $s_{k'} \to s \neq 0$, $x_{k'} \to x \in X$, and such that there exist index sets I and J for which

$$I_\varepsilon(x_{k'}) = I \quad \text{and} \quad J_\varepsilon(x_{k'}) = J, \quad \forall k'. \tag{3.7.1}$$

As in Theorem 6.10 of [5], let us define the sets $K(x_{k'})$, $C(x_{k'})$, K^* and C^* by the following equations:

$$K(x_{k'}) = \mathrm{conv}\{\nabla v_j(x_{k'}) \mid j \in J\}, \tag{3.7.2}$$

$$C(x_{k'}) = \mathrm{cone}\{\nabla g_i(x_{k'}) \mid i \in I\}, \tag{3.7.3}$$

$$K^* = \mathrm{conv}\{\nabla v_j(x) \mid j \in J\}, \tag{3.7.4}$$

and

$$C^* = \mathrm{cone}\{\nabla g_i(x) \mid i \in I\}. \tag{3.7.5}$$

By the definition of $s_{k'}$,

$$s_{k'} = N[K(x_{k'}) + C(x_{k'})]. \tag{3.7.6}$$

Due to the nearest point inequality [5, Eq. (3.7)] and (3.7.6) above, we see that

$$s_{k'}\left(\sum_{j \in J} \lambda_j \nabla v_j(x_{k'}) + \sum_{i \in I} \mu_i \nabla g_i(x_{k'})\right) \geq |s_{k'}|^2, \qquad (3.7.7)$$

for all λ_j, $\mu_i \geq 0$ with $\sum_{j \in J} \lambda_j = 1$. For fixed (λ_j) and (μ_i), we allow $k' \to \infty$ in (3.7.7) to get the inequality

$$s\left(\sum_{j \in J} \lambda_j \nabla v_j(x) + \sum_{i \in I} \mu_i \nabla g_i(x)\right) \geq |s|^2. \qquad (3.7.8)$$

Inequality (3.7.8) is exactly the inequality derived after equation (6.10.10) in [5]. In [5] we stated that this inequality yielded the assertion

$$s = N[K^* + C^*], \qquad (3.7.9)$$

which in turn was invoked to derive inequalities (6.10.12) and (6.10.15) in [5]. As in [5], what we need here also are the inequalities corresponding to (6.10.12) and (6.10.15) of [5]. We now make the following observation, which obviates (3.7.9).

By Lemma 5.4 of [5], $J_0(x) \subset J$ and $I_0(x) \subset I$, so that $K_0(x) \subset K^*$ and $C_0(x) \subset C^*$. Setting all μ_i equal to zero in (3.7.8), we get the inequality $sy \geq |s|^2$, when $y = \sum_{j \in J} \lambda_j \nabla v_j(x)$, $\lambda_j \geq 0$, $\sum_{j \in J} \lambda_j = 1$. This yields the inequality

$$sy \geq |s|^2, \qquad \forall y \in K_0(x). \qquad (3.7.10)$$

We also have the inequality

$$s\nabla g_i(x) \geq 0, \qquad \forall i \in I_0(x). \qquad (3.7.11)$$

For, if there exists some index $p \in I_0(x)$ such that $s\nabla g_p(x) < 0$, we fix all the λ_j's, set $\mu_i = 0$, $\forall i \in I \setminus \{p\}$, and allow $\mu_p \to \infty$ in (3.7.8) to see a contradiction. So we have verified that (3.7.11) prevails. The above observation shows that reference to (3.7.9) here (and in [5] the reference to Eq. (6.10.11) there) can be avoided to derive Eqs. (3.7.10) and (3.7.11).

By (5.8.1) of [5] we have,

$$F'(x; -s) = -\min\{ys \mid y \in K_0(x)\}$$
$$\leq -|s|^2, \qquad \text{by (3.7.10).} \qquad (3.7.12)$$

We now distinguish two cases.

Case 1. Let us consider first the simpler situation where $I_{\varepsilon_{k'}}(x_{k'})$ is empty for an infinity of indices in the subsequence (k'). Passing to a further

subsequence of (k'), again denoted by (k'), we can require that $I_{\varepsilon_k}(x_{k'})$ be empty for every k'. In this case, due to Step 8 of Algorithm 4.1 of [5], $u_{k'} = 0$, $\forall k'$, so that $t_{k'} = s_{k'}$. Hence

$$x_{k'+1} = x_{k'} - \alpha_k s_{k'} \to x - \alpha s. \tag{3.7.13}$$

This shows that $x - \alpha s$ and x are both cluster points of the sequence (x_k); but $(F(x_k))$ is monotone decreasing, and so we get

$$F(x - \alpha s) = F(x). \tag{3.7.14}$$

Since $I_{\varepsilon_k}(x_{k'})$ is empty for every k', I is empty and so $I_0(x)$ is also empty. Hence $-s$ is a feasible direction at x. Due to (3.7.12) we also see that $-s$ is a direction of strict descent for F at x. The remainder of the proof of this simpler case will be merged shortly with that of Case 2 below.

Case 2.   We now consider the case where the subsequence (k') is such that $I_{\varepsilon_k}(x_{k'}) = I_\varepsilon(x_{k'})$ are nonempty, for all sufficiently large k'. Lemma 6.7 of [5] now applies, and so by Eqs. (6.7.1) and (6.7.2) of [5], we can pass to a further subsequence of (k'), which we again denote by (k') so that $u_{k'} \to u$, $|u| \geq 1$. By Step 8 of Algorithm 2.3,

$$M_{k'} = (\max_{1 \leq j \leq r} |\nabla v_j(x_{k'})|) |u_{k'}|. \tag{3.7.15}$$

Since each v_j is of class C^1, with (Δ_k) bounded way from zero, we see that $M_k \to M$, where

$$M = (\max_{1 \leq j \leq r} |\nabla v_j(x)|) |u| > 0. \tag{3.7.16}$$

By Step 9 of the algorithm,

$$\lambda_{k'} = |s_{k'}|^2/(2M_{k'} + 1) \to |s|^2/(2M + 1) = \lambda. \tag{3.7.17}$$

By Lemma 3.6, (t_k) is bounded away from zero. Hence

$$t_{k'} = s_{k'} + \lambda_{k'} u_{k'} \to s + \lambda u = t \neq 0. \tag{3.7.18}$$

and

$$x_{k+1} = x_{k'} - \alpha_{k'} t_{k'} \to x - \alpha t. \tag{3.7.19}$$

Due to (3.7.19), $x - \alpha t$ and x are cluster points of (x_k), and so, as in Case 1, we get

$$F(x - \alpha t) = F(x). \tag{3.7.20}$$

We now show that $-t$ is a feasible direction of strict descent for F at x. By (5.17.2) of [5], for every $i \in I_0(x)$ we have

$$\nabla g_i(x_{k'}) u_{k'} \geqslant |\nabla g_i(x_{k'})|,$$

and so in the limit

$$\nabla g_i(x) u \geqslant |\nabla g_i(x)|,$$

$$> 0, \qquad \text{by Lemma 5.16 of [5].} \qquad (3.7.21)$$

Inequalities (3.7.11) and (3.7.21), by virtue of Lemma 5.18 of [5], show that $-t$ is a feasible direction of x. Due to (3.7.16)

$$(\max_{j \in J_0(x)} |\nabla v_j(x)|) |u| \leqslant M, \qquad (3.7.22)$$

whereas by (5.8.1) of [5] we have

$$F'(x; -u) = \max_{j \in J_0(x)} \{ -\nabla v_j(x) u \}. \qquad (3.7.23)$$

Combining (3.7.22) and (3.7.23) we see that

$$F'(x; -u) \leqslant M. \qquad (3.7.24)$$

Due to (3.7.18), the sublinearity of the function $z \mapsto F'(x; z)$, and the fact that $\lambda > 0$, we have

$$F'(x; -t) \leqslant F'(x; -s) + \lambda F'(x; -u), \qquad (3.7.25)$$

$$\leqslant -|s|^2 + |s|^2 M/(2M + 1), \qquad (3.7.26)$$

$$< -|s|^2/2 < 0. \qquad (3.7.27)$$

In arriving at (3.7.26) from (3.7.25), we used (3.7.12), (3.7.17) and (3.7.24). Inequality (3.7.27) shows that $-t$ is a direction of strict descent for F at x, completing our verification of the assertion that $-t$ is a feasible direction of strict descent at x. Parenthetically, we note that we have now reproved the inequalities (6.10.12), (6.10.14) and (6.10.15) of [5].

Recall that in Case 1, $t = s$ and so (3.7.14) is the same as (3.7.20). We have, therefore, shown that (3.7.20) holds, and that $-t$ is a feasible direction of strict descent at x, irrespective of whether Case 1 or Case 2 prevails. So, there exists $\delta > 0$ such that

$$x - \theta t \in X \qquad \text{and} \qquad F(x - \theta t) < F(x), \qquad \forall \theta \in (0, \delta]. \quad (3.7.28)$$

By Step 11 of Algorithm 4.1 and Lemma 5.22 of [5],

$$F(x_{k'} - \alpha_{k'} t_{k'}) \leqslant F(x_{k'} - \lambda t_{k'}), \qquad \forall \lambda \in [0, \bar{\alpha}_k]. \tag{3.7.29}$$

Since $\alpha_{k'} \to \alpha > 0$, $\bar{\alpha}_{k'} \geqslant \alpha_{k'} \geqslant \alpha/2 > 0$, $\forall k'$ sufficiently large. So we can find a fixed λ such that $0 < \lambda \leqslant \delta$ and $0 < \lambda \leqslant \bar{\alpha}_{k'}$ $\forall k'$ sufficiently large. With this choice of λ, allowing $k' \to \infty$ in (3.7.29) yields the inequality

$$F(x - \alpha t) \leqslant F(x - \lambda t), \tag{3.7.30}$$

whereas, because of (3.7.28), we also have

$$F(x - \lambda t) < F(x). \tag{3.7.31}$$

Combining (3.7.30) and (3.7.31) we get the inequality

$$F(x - \alpha t) < F(x), \tag{3.7.32}$$

contradicting (3.7.20); and the proof of the lemma is complete.

3.8. LEMMA. *Let (ε_k), (s_k), etc., be as in the algorithm. Suppose that there exists $\varepsilon > 0$ such that $\varepsilon_k = \varepsilon > 0$ eventually and such that the sequences $(|s_k|)$ and (Δ_k) are both bounded away from zero. Let the subsequence $(x_{k'})$ be such that $x_{k'} \to x \in X$. Then there is a subsequence of (k'), again denoted by (k'), such that $I_0(x_{k'}) = I_0(x)$ for all k'.*

Proof. This is Lemma 6.9 of [5], and the proof given there carries over verbatim, if we redefine M occurring in the proof of that lemma by

$$M = \sup_k |\nabla g_i(x_k)|.$$

By Lemma 3.1 the sequence (x_k) is bounded, and hence M is finite. Combining this with Lemma 5.16 of [5], we see that $0 < M < \infty$. The changes required in the proof of this lemma are now complete.

4. OTHER MINOR CHANGES

4.1. Note also the minor change in Lemma 5.22 of [5]. We cannot any longer assert that α_k is unique. A similar remark applies to Lemma 5.13 of [4] also. The more elaborate reasoning used in Lemma 3.7 may be used to reprove Lemma 5.19 of [4], with f set identically zero there. The required alterations in proof are similar, but clearly simpler. One can then modify the statement of Theorem 5.23 of [4] in the same manner as Theorem 2.4 of Section 2. More explicitly: Let (P) be the problem of minimizing a

piecewise affine, convex function v, subject to a finite collection of affine constraints. Assume that v is coercive on the feasibility set X. We have the following theorem.

4.2 THEOREM. *The Algorithm 4 in* [4] *generates a sequence which either terminates at a minimizer of problem* (P) *or else clusters only at minimizers of problem* (P).

Finally, we can combine Theorems 2.4 and 4.2 into a single theorem, and thus generalize Algorithm 2.3 in the spirit of Algorithm 7.1 of [5] to the case of mixed constraints. Let $g_1,..., g_p$ all be nonaffine, convex and differentiable on Ω and let $g_{p+1},..., g_m$ be affine. The algorithm corresponding to Algorithm 2.3 is contained in the following theorem.

4.3. THEOREM. *Let* v *be as in Section* 2, $g_1,..., g_p$ *nonaffine, convex and differentiable, and* $g_{p+1},..., g_m$ *affine. Suppose that generalized Slater's constraint qualification* (GSQ) *(as explained in Section* 7.1 *of* [5]) *holds. Define the index set* I *in Steps* 8 *and* 10 *of Algorithm* 2.3 *as in the corresponding steps of Algorithm* 7.1 *of* [5]. *Then the generated sequence* (x_k) *either terminates at a global minimizer, or is such that every cluster point of* (x_k) *is a global minimizer of the problem* (P).

The minimizers in the Theorems 2.4, 4.2 and 4.3 need not be unique.

5. NUMERICAL RESULTS

We had suspected all along that the algorithms in [5] and [4] produce cluster points all of which are solutions of problem (P), even when the objective functions are not strictly convex. In fact, many of the numerical examples tested by both Rubin [3] and Owens [2] have objective functions that are not strictly convex. Rubin, after successfully solving two examples of Wolfe [7] (which had $f = 0$) using the algorithm in [4], remarked in [3, p. 326] that the algorithm in [4] is not guaranteed to converge since the objective function is not strictly convex. Owens [2] applied the algorithm in [5] to Wolfe's [7] examples and Dem'yanov and Malozemov's [1] "jamming" example, all of which had objective functions that are not strictly convex, and found that the algorithm in [5] converged, i.e., cluster points are solutions of problem (P). See [2] for more details.

REFERENCES

1. V. F. DEM'YANOV AND V. N. MALOZEMOV, "Introduction to Minimax," Wiley, New York, 1974.
2. R. W. OWENS, Implementation of subgradient projection algorithm II, *Internat. J. Comput. Math.* **16** (1984), 57–69.
3. P. RUBIN, Implementation of a subgradient projection algorithm, *Internat. J. Comput. Math.* **12** (1983), 321–328.
4. V. P. SREEDHARAN, A subgradient projection algorithm, *J. Approx. Theory* **35** (1982), 111–126.
5. V. P. SREEDHARAN, Subgradient projection algorithm II, *J. Approx. Theory* **41** (1984), 217–243.
6. J. STOER AND C. WITZGALL, "Convexity and Optimization in Finite Dimensions, I," Springer-Verlag, New York, 1970.
7. P. WOLFE, A method of conjugate subgradients for minimizing nondifferentiable functions, *Math. Programming Study* **3** (1975), 145–173.