JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 55, No. 1, OCTOBER 1987

Penalty-Proximal Methods in Convex Programming

A. AUSLENDER,[1] J. P. CROUZEIX,[2] AND P. FEDIT[3]

Communicated by O. L. Mangasarian

Abstract. An implementable algorithm for constrained nonsmooth convex programs is given. This algorithm combines exterior penalty methods with the proximal method. In the case of a linear program, the convergence is finite.

Key Words. Constrained nonsmooth convex programming, penalty methods, proximal methods, linear programming.

1. Introduction

Let f be a real-valued convex function defined on $\mathbb{R}^N$, not necessarily differentiable; and let C be a nonempty closed convex subset of $\mathbb{R}^N$. We shall consider the convex program

$$(\mathrm{P}) \quad \alpha = \inf(f(x) | x \in C). \tag{1}$$

Methods given initially by Lemarechal (Refs. 1-2) and Wolfe (Ref. 3) for solving unconstrained problems have been extended to the constrained case by several authors [Mifflin (Ref. 4), Lemarechal, Strodiot, and Bihain (Ref. 5), and Kiwiel (Ref. 6)], especially when C is defined by

$$C = \{x : f_i(x) \leq 0, i = 1, \ldots, m\}, \tag{2}$$

where f_i are convex real-valued functions defined on $\mathbb{R}^N$.

The method that we propose here is related with a method given by Auslender in Ref. 7, where the author proposes a new class of algorithms for (P). We refer to Ref. 7 for a discussion about the relations of these algorithms with the classical methods in convex optimization and with the proximal method studied by several authors [Martinet (Ref. 8), Rockafellar (Refs. 9-10), Fukushima (Ref. 11)].

[1] Professor, Department of Applied Mathematics, University of Clermont II, Aubière, France.
[2] Professor, Department of Applied Mathematics, University of Clermont II, Aubière, France.
[3] Research Student, Department of Applied Mathematics, University of Clermont II, Aubière, France.

Let $\bar{x}$ be a fixed vector in $\mathbb{R}^N$, $\varepsilon > 0$, and let h be a real-valued convex function defined on $\mathbb{R}^N$; consider the following optimization problem:

$(Q(\epsilon, h, \bar{x}))$ Find $\tilde{x} \in \mathbb{R}^N$:

$$h(\tilde{x}) + \tfrac{1}{2}\|\tilde{x} - \bar{x}\|^2 \leq h(x) + \tfrac{1}{2}\|x - \bar{x}\|^2 + \epsilon, \qquad \forall x \in \mathbb{R}^N. \tag{3}$$

The result that we shall use is the following: there exists an implementable algorithm (Algorithm A) which computes in a *finite number* of steps such a point $\tilde{x}$; such an algorithm is described in Ref. 7. Let us denote this point by

$$\tilde{x} = A(\epsilon, \bar{x}, h). \tag{4}$$

Then, for solving Problem (P), we shall combine Algorithm A with other methods. This idea was used already in Ref. 7 with the method of centers (Ref. 12). In this case, C is given by (2), and the minimizing sequence is obtained by the induction formula

$$x_{n+1} = A(\epsilon_n, x_n, \max(f(\cdot) - f(x_n) - 2\epsilon_n, f_i(\cdot) | i = 1, \dots, m),$$

where the sequence $\{\epsilon_n\}$ satisfies

$$\epsilon_n > 0, \tag{5}$$

$$\sum_{n=0}^{\infty} \epsilon_n < +\infty, \qquad \epsilon_{n+1} \leq \tfrac{1}{2}\epsilon_n. \tag{6}$$

The starting point x_0 is taken in $\operatorname{int}(C)$, the interior of C; then, the whole sequence $\{x_n\}$ is in $\operatorname{int}(C)$.

In this paper, we combine Algorithm A with exterior penalty methods, and we do this under weaker convergence assumptions. In particular, (6) is replaced by

$$\lim_{n \to \infty} \epsilon_n = 0. \tag{7}$$

As in Ref. 7, we assume that f is inf-compact on C; that is, for each λ, the set $\{x \in C : f(x) \leq \lambda\}$ is compact, but now we require no assumptions about the existence of $\operatorname{int}(C)$, and it is not necessary to have C defined by (2). We try to *give minimal assumptions*; in particular, the penalty functions Φ_n are functions satisfying the following conditions:

(H1) Φ_n is a real-valued convex function on $\mathbb{R}^N$, for all $n \in \mathbb{N}$;
(H2) $0 \leq \Phi_n(x) \leq \Phi_{n+1}(x)$, $\forall x \in \mathbb{R}^N$, $\forall n \in \mathbb{N}$;
(H3) $\Phi_n(x) = 0$, if $x \in C$; $\lim_{n \to \infty} \Phi_n(x) = +\infty$, $\forall x \notin C$.

Then, the proposed method consists in computing by induction a sequence $\{x_n\}$ such that

$$f(x_{n+1})+\Phi_n(x_{n+1})+\tfrac{1}{2}\|x_{n+1}-x_n\|^2$$
$$\leq f(x)+\Phi_n(x)+\tfrac{1}{2}\|x-x_n\|^2+\epsilon_n, \qquad \forall x\in\mathbb{R}^N, \tag{8}$$

starting from an arbitrary point $x_0\in\mathbb{R}^N$.

The computation of x_{n+1} can be performed by Algorithm A, and in this case we have

$$x_{n+1}=A(\epsilon_n, x_n, f+\Phi_n). \tag{9}$$

We must remark that such a kind of method was already given by Kaplan (Ref. 13). But his penalty functions are different from ours and not exterior; they have to satisfy the following conditions:

(H_1^*) $\lim_{n\to\infty}\Phi_n(x)=\begin{cases}0, & \text{if } x\in \operatorname{int} C,\\ +\infty, & \text{if } x\notin C;\end{cases}$

(H_2^*) $\Phi_n(x)\geq c>0$, for $x\in\partial C=C\backslash\operatorname{int} C$;

(H_3^*) $\Phi_{n+1}(x)\leq\Phi_n(x)$, $\forall x\in C$.

In fact, Kaplan has proved that, for n sufficiently large, $x_n\in\operatorname{int} C$. When C is given by (2), it is easy to see that the classical exterior penalty function

$$\Phi_n(x)=k_n\sum_{i=1}^{m}[f_i^+(x)]^2, \tag{10}$$

with $a^+=\max(a,0)$, and the exact penalty function

$$\Phi_n(x)=k_n\sum_{i=1}^{m}f_i^+(x), \tag{11}$$

with

$$0<k_n\leq k_{n+1}, \qquad \lim_{n\to\infty}k_n=+\infty, \tag{12}$$

do not satisfy assumptions (H_i^*), but satisfy assumptions (H_i).

Furthermore, Kaplan supposes that $\operatorname{int} C\neq\emptyset$, C is compact, f is differentiable, $\sum_{n=0}^{\infty}\epsilon_n<+\infty$, whereas we shall suppose that f is inf-compact and that $\{\epsilon_n\}$ satisfies only (5) and (7). Then, of course, the convergence tools are completely different from those used by Kaplan, but are also different from those introduced in Ref. 7 for the method of centers.

We shall also give additional results under supplementary assumptions. More precisely, suppose additionally that:

(i) $\sum_{n=0}^{\infty}\sqrt{\epsilon_n}<+\infty;$ (13)

(ii) a technical assumption (H), exposed in the next section (in particular, Slater's condition implies this assumption for the classical penalty method), holds.

Then, we shall prove that the sequence $\{x_n\}$ converges to a *single* optimal point.

Penalty methods are considered as ill conditioned, because the penalty parameter k_n tends to infinity. In the exact penalty case, we shall overcome this default by limiting the parameter growth.

Finally, we shall consider the case where (P) is a linear program; i.e., $f(x)=(c, x)$, and C is defined by

$$C=\{x: (a_i, x)\leq b_i, i=1,2,\ldots, m; (a_i, x)=b_i, i=m+1,\ldots, r\}. \quad (14)$$

Here, (x, y) denotes the usual inner product in $\mathbb{R}^N$ of x and y. Under Slater's condition, we shall prove in that case that the method converges in a finite number of steps when:

(i) $\epsilon_n=0, \forall n$; (15)

(ii) Φ_n is the exact penalty function given by

$$\Phi_n(x)=n\left(\sum_{i=1}^{m}[(a_i, x)-b_i]^+ + \sum_{i=m+1}^{r}|(a_i, x)-b_i|\right). \quad (16)$$

This can be considered as a promising result. Indeed, in this case, at each step n the problem to be solved can be transformed into a quadratic programming problem by introducing some artificial variables. Then, if one uses a finite quadratic method, the new algorithm will solve a linear program in a finite number of steps. The efficiency of the whole method will be dependent of course on the quadratic method employed.

2. Theoretical Results

Throughout the section, f is a real-valued convex function on $\mathbb{R}^N$, C is a closed nonempty convex subset of $\mathbb{R}^N$, (P) is the convex program defined by

$$(P) \quad \alpha=\inf(f(x)|x\in C).$$

We assume that S, the set of optimal solutions of (P), is compact and nonempty. This last assumption (Theorem 27.1, Ref. 14) is equivalent to requiring the inf-compactness of the function f with respect to C, that is, the compactness of the set $\{x\in C|f(x)\leq\lambda\}$, for all $\lambda\in\mathbb{R}$.

Also, let $\{\epsilon_n\}$ be a sequence of nonnegative reals which converges to 0, and let $\{\Phi_n\}$ be a sequence of penalty functions satisfying assumptions

(H1), (H2), (H3). For the sake of simplicity, we set $f_n = f + \Phi_n$. Then, we have the following lemma.

Lemma 2.1. There exists $\bar{n}$ such that, for each $n \geq \bar{n}$, f_n is inf-compact.

Proof. For all n, define

$$S_n = \{x \in \mathbb{R}^N | f_n(x) \leq \alpha\}.$$

From Theorem 27.1 of Ref. 14, f_n is inf-compact, if S_n is a nonempty compact set.

Assumptions (H1) to (H3) imply that S_n is convex and closed,

$$S \subset S_{n+1} \subset S_n, \qquad \text{for all } n.$$

Besides,

$$S = \bigcap_n S_n.$$

By assumption, S is nonempty, so that it is enough to prove that S_n is bounded for n large enough.

Assume for contradiction that S_n is not bounded for all n. Let $\bar{x}$ be fixed in S and define, for all n,

$$K_n = \{d \in \mathbb{R}^N | \, \|d\| = 1, \bar{x} + td \in S_n, \text{ for all } t > 0\}.$$

Then, K_n is the set of points having norm one which belong to the recession cone of S_n; K_n is compact and nonempty, because S_n is unbounded. On the other hand,

$$K_{n+1} \subset K_n, \qquad \text{for all } n.$$

This implies the existence of some $d \in \mathbb{R}^N$ such that $d \neq 0$ and $d \in \bigcap_n K_n$. By definition of K_n and by virtue of the relation

$$S = \bigcap_n S_n,$$

$\bar{x} + td$ belongs to S for all positive t. This contradicts the boundedness of S. □

For the following, we shall denote as in Ref. 14 by $\text{prox}(y|f)$ the point that minimizes on $\mathbb{R}^N$ the mapping

$$x \to f(x) + \tfrac{1}{2}\|x - y\|^2.$$

Consider now a sequence $\{x_n\}$ satisfying relation (8), and set

$$y_n = \text{prox}(x_n | f_n).$$

Then, we have the following lemma.

Lemma 2.2

$$\|y_p - x_{p+1}\| \leq \sqrt{2\epsilon_p}, \qquad \text{for all } p > 0. \tag{17}$$

Proof. Set

$$\Psi_p(y) = f_p(y) + \tfrac{1}{2}\|y - x_p\|^2.$$

Since Ψ_p is uniformly strongly convex, the optimality conditions at y_p, Proposition 6 of Ref. 10, and relation (8) imply that

$$\tfrac{1}{2}\|x_{p+1} - y_p\|^2 + \Psi_p(y_p) \leqslant \Psi_p(x_{p+1}) \leqslant \Psi_p(y_p) + \epsilon_p,$$

from which (17) follows. □

For all $n \geqslant \bar{n}$ and $\delta > 0$, define

$$S_\delta^n = \{x \in \mathbb{R}^N \mid f_n(x) \leqslant \alpha + \delta\}, \qquad S_\delta^\infty = \{x \in C \mid f(x) \leqslant \alpha + \delta\}.$$

Then, S_δ^n is convex and compact (Lemma 2.1) and, from assumptions (H1) to (H3),

$$S \subset S_\delta^n, \qquad S_\delta^\infty = \bigcap_n S_\delta^n.$$

Lemma 2.3. There exist reals M^n satisfying

$$M_\delta^n \geqslant M_\delta^{\bar{n}} > 0, \qquad \text{for all } n \geqslant \bar{n},$$

such that

$$(x^*, x - y) \geqslant M_\delta^n \|x - y\|, \tag{18}$$

for all $x \notin S_\delta^n$, $y \in S$, $x^* \in \partial f_p(x)$, $p \geqslant n$.

Proof. Let z be the point where the line segment between x and y intersects the boundary of S_δ^n. Then,

$$f_p(z) = f(z) + \Phi_p(z) \geqslant f(z) + \Phi_n(z) = \alpha + \delta.$$

We consider the restriction of f to the line passing through x and y. It results from the differentiability properties of convex functions of one real variable that

$$\begin{aligned}\delta \leqslant f_p(z) - f_p(y) \leqslant f_p'(z, z - y) &\leqslant f_p'(x, z - y)\\ &= f_p'(x, x - y)[\|z - y\| / \|x - y\|],\end{aligned}$$

where $f_p'(a, d)$ denotes the one-sided directional derivative of f_p at a with respect to d. Hence,

$$f_p'(x, x - y) \geqslant M_\delta^n \|x - y\|, \tag{19}$$

where

$$M_\delta^n = \inf(\delta / \|y' - z'\| \mid y' \in S, z' \in bd(S_\delta^n)),$$

where $bd(S_\delta^n)$ denotes the boundary of S_δ^n.

Since $S_\delta^n \subset S_\delta^{\bar{n}}$, and since S is strictly included in $S_\delta^{\bar{n}}$, then $M_\delta^n \geq M_\delta^{\bar{n}} > 0$. Now, let y be fixed in S; and let E_p be the set of points where f_p is differentiable. We have proved that (18) holds for all $x \in E_p$. Let $x \notin E_p$ and $x^* \in \partial f_p(x)$. From Ref. 14, Theorem 25.6, there exist m, $\lambda_i > 0$, x_i^*, $x_i^j \in E_p$, $i = 1, 2, \ldots, m$, such that

$$x^* = \sum_{i=1}^m \lambda_i x_i^*, \; x_i^* = \lim_{j\to\infty} \nabla f(x_i^j),$$

$$x = \lim_{j\to\infty} x_i^j, \qquad i = 1, \ldots m, \qquad \sum_{i=1}^m \lambda_i = 1.$$

For j large enough, since S_δ^n is closed, $x_i^j \notin S_\delta^n$. Hence, from (18),

$$(\nabla f(x_i^j), x_i^j - y) \geq M_\delta^n \|x_i^j - y\|;$$

and, passing to the limit, we obtain

$$(x^*, x - y) \geq M_\delta^n \|x - y\|.$$ □

Theorem 2.1. The sequence $\{x^n\}$ is bounded and all its limit points belong to S.

Proof. (i) Set, for $n \geq \bar{n}$, $\delta > 0$,

$$r_n(\delta) = \sup_{y \in S_\delta^n} \inf_{x \in S} \|x - y\|,$$

$$r_\infty(\delta) = \sup_{y \in S_\delta^\infty} \inf_{x \in S} \|x - y\|.$$

Observe that $r_n(\delta)$ [resp., $r_\infty(\delta)$] is the Hausdorff distance between S and S_δ^n [resp., S_δ^∞]. Since

$$S_\delta^\infty = \bigcap_l S_\delta^l, \qquad S = \bigcap_{\delta>0} S_\delta^\infty,$$

it follows that

$$r_\infty(\delta) = \lim_{l\to\infty} r_l(\delta), \qquad \lim_{\delta\to 0^+} r_\infty(\delta) = 0. \tag{20}$$

Then, let us define

$$T_\delta^n = \{x \in \mathbb{R}^N : f(x) + \Phi_n(x) \leq \alpha + \tfrac{1}{2} r_n(\delta)^2\},$$

$$T_\delta^\infty = \{x \in C : f(x) \leq \alpha + \tfrac{1}{2} r_\infty(\delta)^2\},$$

$$A_\delta^n = S_\delta^n \cup T_\delta^n, \qquad A_\delta^\infty = S_\delta^\infty \cup T_\delta^\infty;$$

and let $q_n(\delta)$ [resp., $q_\infty(\delta)$] denote the Hausdorff distance between S and A_δ^n[resp., A_δ^∞],

$$q_n(\delta) = \sup_{x \in A_\delta^n} \inf_{x \in S} \|x - y\|,$$

$$q_\infty(\delta) = \sup_{y \in A_\delta^\infty} \inf_{x \in S} \|x - y\|.$$

From (20), it follows that

$$T_\delta^\infty = \bigcap_l T_\delta^l,$$

so that we have

$$q_\infty(\delta) = \lim_{l \to \infty} q_l(\delta), \ \lim_{\delta \to 0^+} q_\infty(\delta) = 0. \tag{21}$$

Let $p(n) \geq n$ be such that

$$\epsilon_{p(n)} = \max_{j > n} \epsilon_j,$$

and set

$$A_\delta^{n,p} = A_\delta^n + B(0, \sqrt{(2\epsilon_p)}),$$

where $B(0, r)$ is the closed ball centered at 0 with radius r. Let $q_{n,p}(\delta)$ be the Hausdorff distance between the compact sets $A_\delta^{n,p}$ and S; then,

$$q_{n,p}(\delta) = \sup_{x \in A_\delta^{n,p}} \inf_{y \in S} \|x - y\|;$$

and, since

$$\lim_{l \to \infty} \epsilon_{p(l)} = 0,$$

we have from (21) that

$$q_\infty(\delta) = \lim_{l \to \infty} q_{l,p(l)}(\delta). \tag{22}$$

Finally, set

$$W_{\delta,n} = S + B(0, q_{n,p(n)}(\delta)), \qquad W_{\delta,\infty} = S + B(0, q_\infty(\delta)).$$

We shall prove that x_p belongs to the compact set $W_{\delta,n}$ for p large enough. Then, letting first $n \to +\infty$ and next $\delta \to 0$, we shall obtain Theorem 2.1 as a consequence of (22) and (21).

(ii) We shall now prove that x_p belongs to $W_{\delta,n}$ for p large enough. Since

$$M_\delta^n \geq M_\delta^{\bar{n}} \quad \text{and} \ \lim_{p \to \infty} \epsilon_p = 0,$$

there exists $\bar{p} \geq p(n)$ such that

$$\sqrt{(2\epsilon_p)} - M_\delta^n < -\tfrac{1}{2} M_\delta^{\bar{n}}, \qquad \forall p \geq \bar{p}. \tag{23}$$

For the following, let $p \geq \bar{p}$.

(a) If $x_p \in S_\delta^n$, then, by definition of y_p,

$$f(y_p)+\Phi_p(y_p)+\tfrac{1}{2}\|y_p-x_p\|^2 \leqslant f(x)+\tfrac{1}{2}\|x-x_p\|^2, \qquad \text{for all } x\in S;$$

consequently, since $p \geqslant n$,

$$f(y_p)+\Phi_n(y_p)\leqslant f(y_p)+\Phi_p(y_p)+\tfrac{1}{2}\|y_p-x_p\|^2\leqslant \alpha+\tfrac{1}{2}r_n^2(\delta).$$

By definition of T_δ^n, y_p belongs to T_δ^n; and, by Lemma 2.2, it follows that

$$x_{p+1}\in T_\delta^n+B(0,\sqrt{(2\epsilon_p)}). \tag{24}$$

(b) If $x_p \notin S_\delta^n$ and $y_p\in S_\delta^n$, apply Lemma 2.2. Then,

$$x_{p+1}\in S_\delta^n+B(0,\sqrt{(2\epsilon_p)}). \tag{25}$$

(c) If $x_p\notin S_\delta^n$ and $y_p\notin S_\delta^n$, by definition of y_p one has

$$x_p-y_p\in\partial f_p(y_p).$$

Apply Lemma 2.3. Then,

$$(x_p-y_p, y_p-y)\geqslant M_\delta^n\|y_p-y\|, \qquad \text{for all } y\in S.$$

It follows straightforwardly that, for all $y\in S$,

$$(x_p-y, y_p-y)\geqslant M_\delta^n\|y_p-y\|+\|y_p-y\|^2,$$

$$\|x_p-y\|\,\|y_p-y\|\geqslant M_\delta^n\|y_p-y\|+\|y_p-y\|^2,$$

$$\|x_p-y\|\geqslant M_\delta^n+\|y_p-y\|.$$

Apply Lemma 2.2 and (23). Then,

$$\|x_{p+1}-y\|\leqslant\|x_p-y\|+\sqrt{(2\epsilon_p)}-M_\delta^n, \qquad \text{for all } y\in S,$$

$$\|x_{p+1}-y\|<\|x_p-y\|-\tfrac{1}{2}M_\delta^{\bar{n}}, \qquad \text{for all } y\in S. \tag{26}$$

(d) We have already seen that the theorem will be established if we prove that x_p belongs to $W_{\delta,n}$ for p large enough. For this, we shall prove the following:

(d1) the existence of some integer p such that $x_p\in W_{\delta,n}$;
(d2) $x_p\in W_{\delta,n}\Rightarrow x_{p+1}\in W_{\delta,n}$.

Let us introduce

$$K=\{p\geqslant\bar{p}: x_p\notin S_\delta^n \text{ and } y_p\notin S_\delta^n\},$$

$$K^c=\{p\geqslant\bar{p}: p\notin K\}.$$

Observe that

$$A_\delta^{n,p(n)}\subset W_{\delta,n}.$$

From the definition of K^c and relations (24) and (25), we have

$$x_{p+1}\in A_\delta^{n,p}\subset A_\delta^{n,p(n)}\subset W_{\delta,n}, \qquad \forall p\in K^c. \tag{27}$$

Furthermore, it follows from (26) that K^c is infinite; hence, (d1) is proved. On the other hand, we see from (27) that (d2) is valid for $p \in K^c$; for $p \in K$, this is a consequence of (26), since

$$d(x_{p+1}, S) < d(x_p, S) \leq q_{n,p(n)}(\delta),$$

where $d(x, S)$ denotes the distance of x to S. □

Remark 2.1. It can be seen easily that Theorem 2.1 remains true when replacing in the formula (8) the regularizing function $\frac{1}{2}\|x - x_n\|^2$ by $(1/2c_n)\|x - x_n\|^2$, with

$$0 < c_n \nearrow c_\infty \leq +\infty,$$

subject to

$$\lim_{n\to\infty} c_n \epsilon_n = 0,$$

which is satisfied in particular when $c_\infty < +\infty$.

Now, we shall give two supplementary conditions in order to obtain the convergence of the whole sequence $\{x_n\}$ to an unique optimal point. The first condition is

$$\sum_{n=0}^{\infty} \sqrt{(\epsilon_n)} < +\infty. \tag{28}$$

A counterexample given by Rockafellar in (Ref. 10) for an unconstrained problem shows that this condition is necessary for obtaining the convergence of $\{x_n\}$.

The second condition concerns the choice of the penalty functions Φ_n. We assume that, for each convex program (P) defined as in the introduction, there exists a positive constant L and a sequence $\{k_n\}$ of positive reals (depending only on f, C, Φ_n), such that

$$\sum_{n=0}^{\infty} 1/\sqrt{(k_n)} < +\infty, \tag{29}$$

$$\alpha \leq f(y_n^*) + L/k_n, \tag{30}$$

where y_n^* minimizes $f + \Phi_n$ on $\mathbb{R}^N$, that is,

$$f(y_n^*) + \Phi_n(y_n^*) \leq f(y) + \Phi_n(y), \qquad \forall y \in \mathbb{R}^N. \tag{31}$$

This assumption is satisfied, for example, for the classical penalty function given by (10) when C is given by (2) and when the Slater's condition is satisfied. This is a consequence of formula (8.11) of Theorem 8.3 of Ref. 15, $L = (5/4)\|\bar{u}\|^2$, where $\bar{u}$ is a Kuhn-Tucker multiplier, $k_n = n^{2+\theta}$, with $\theta > 0$.

Theorem 2.2. Under the above assumptions, the sequence $\{x_n\}$ converges to an optimal solution.

Proof. (a) Let $\tilde{y} \in \mathbb{R}^N$, and consider the following optimization problem:

$$(P(\tilde{y})) \quad \tilde{\alpha} = \min(F(x) | x \in C),$$

where

$$F(x) = f(x) + \tfrac{1}{2}\|x - \tilde{y}\|^2, \qquad \forall x \in \mathbb{R}^N. \tag{32}$$

Let $y_n(\tilde{y})$ be the point that minimizes $F + \Phi_n$ on $\mathbb{R}^N$; and let $y^*(\tilde{y})$ be the point that minimizes F on C. From (30), there exists a constant $L(\tilde{y})$ and a sequence $\{k_n(\tilde{y})\}$ such that

$$F(y^*(\tilde{y})) \leq F(y_n(\tilde{y})) + L(\tilde{y})/k_n(\tilde{y}). \tag{33}$$

Now, since $\Phi_n(y^*(\tilde{y})) = 0$ and since F is strongly convex, the necessary optimality conditions for minimizing $F + \Phi_n$ on $\mathbb{R}^N$ give

$$F(y^*(\tilde{y}) \geq F(y_n(\tilde{y})) + \tfrac{1}{2}\| y^*(\tilde{y}) - y_n(\tilde{y})\|^2.$$

Combining this inequality with (33), we obtain

$$\| y^*(\tilde{y}) - y_n(\tilde{y})\| \leq \sqrt{[2L(\tilde{y})/k_n(\tilde{y})]}. \tag{34}$$

(b) Now, we choose $\tilde{y}$ in the optimal set S. Then, observing that in this case $\tilde{y}$ minimizes F on C, we conclude that $y^*(\tilde{y}) = \tilde{y}$, so that (34) becomes

$$\| \tilde{y} - y_n(\tilde{y})\| \leq \sqrt{[2L(\tilde{y})/k_n(\tilde{y})]}. \tag{35}$$

Furthermore, taking the notation given before Lemma 2.2, we observe that

$$y_n = \text{prox}(x_n | f_n), \; y_n(\tilde{y}) = \text{prox}(\tilde{y} | f_n);$$

and then, since the mapping $\text{prox}(\cdot | f_n)$ is nonexpansive (page 340, Ref. 14), we obtain

$$\| y_n - y_n(\tilde{y})\| \leq \|\tilde{y} - x_n\|. \tag{36}$$

Finally, we have

$$\|x_{n+1} - \tilde{y}\| \leq \|x_{n+1} - y_n\| + \| y_n - y_n(\tilde{y})\| + \| y_n(\tilde{y}) - \tilde{y}\|.$$

Using Lemma 2.2, (28), (29), (35), and (36), this gives

$$\|x_{n+1} - \tilde{y}\| \leq \|x_n - \tilde{y}\| + \theta_n, \qquad \text{with } \theta_n > 0, \; \sum_{n=1}^{\infty} \theta_n < +\infty. \tag{37}$$

(c) The end of the proof is now classical (see, for example, the proof of Theorem 1 of Ref. 10). Relations (37) imply, for each $\tilde{y} \in S$, the existence of

$$\lim_{n \to \infty} \|x_n - \tilde{y}\| = \mu(\tilde{y}) < +\infty.$$

We have then to show that there cannot be more than one cluster point of $\{x_n\}$. Suppose that there are two, $y_1 \neq y_2$. Then, from Theorem 2.1, y_1, y_2 belong to S and

$$\|x_n - y_1\|^2 - \|x_n - y_2\|^2 = -2(x_n, y_1 - y_2) + \|y_1\|^2 - \|y_2\|^2.$$

Passing to the limit, we obtain, respectively,

$$a = \mu(y_1)^2 - \mu(y_2)^2 = -2(y_1, y_1 - y_2) + \|y_1\|^2 - \|y_2\|^2 = -\|y_1 - y_2\|^2$$
$$= -2(y_2, y_1 - y_2) + \|y_1\|^2 - \|y_2\|^2 = +\|y_1 - y_2\|^2,$$

so that $y_1 = y_2$. □

We now give some additional properties for exact penalty functions. Suppose that C is defined by

$$C = \{x: f_i(x) \leq 0,\ i = 1, 2, \ldots, m;\ (a_i, x) = b_i,\ i = m+1, \ldots, r\}.$$

Theorem 2.3. Suppose that the following regularity assumptions hold:

(a) the vectors a_i, $i = m+1, \ldots, r$, are linearly independent;
(b) there exists $\tilde{x}$ such that

$$(a_i, \tilde{x}) = b_i, \quad \forall i = m+1, \ldots, r, \qquad f_i(\tilde{x}) < 0, \quad \forall i = 1, 2, \ldots, m.$$

Consider the classical exact penalty function $\Phi_n(x) = k_n \Phi(x)$, with

$$\Phi(x) = \sum_{i=1}^{m} f_i^+(x) + \sum_{i=m+1}^{r} |(a_i, x) - b_i|,$$

$$0 \leq k_n \leq k_{n+1}, \qquad \lim_{n\to\infty} k_n = +\infty.$$

Suppose that

$$\sum_{n=0}^{\infty} \sqrt{(\epsilon_n)} < +\infty.$$

Then,

(i) there exists n^* such that, for $n \geq n^*$, we have $y_n \in C$, and the sequences $\{y_n\}, \{x_n\}$ converge to an unique optimal solution;
(ii) if $f(x) = (c, x)$, $f_i(x) = (a_i, x) - b_i$, and if $\epsilon_n = 0$, for each n, then the algorithm gives an optimal solution in a finite number of steps.

Proof. (i) Denote by $\langle p, q\rangle$ the set of integers included in $[p, q]$, and set

$$I(x) = \{i \in \langle 1, m\rangle: f_i(x) = 0\} \cup \{i \in \langle m+1, r\rangle: (a_i, x) = b_i\},$$

$$I_-(x) = \{i \in \langle 1, m\rangle: f_i(x) < 0\}.$$

(a) Let us prove that $y_n \in C$, for n large enough. In the contrary case, from Lemma 2.2 and Theorem 2.1, there would exist $y^* \in S$, $x^* \in S$, a non-empty constant set $I \subset I(y^*)$, and subsequences $\{y_{n_l}\}$, $\{x_{n_l}\}$, such that:

$$y^* = \lim_{l\to\infty} y_{n_l}, \qquad x^* = \lim_{l\to\infty} x_{n_l};$$

$$f_i(y_n) > 0, \quad i \in \langle 1, m\rangle \cap I; \qquad (a_i, y_{n_l}) - b_i \neq 0, \quad i \in \langle m+1, r\rangle \cap I;$$

$$f_i(y_n) \leq 0, \quad i \in \langle 1, m\rangle \cap I^c; \qquad (a_i, y_n) - b_i = 0, \quad i \in \langle m+1, r\rangle \cap I^c;$$

here, I^c is the complementary set of I in $\langle 1, r\rangle$. Set

$$I_1 = I \cap \langle 1, m\rangle, \qquad I_2 = I \cap \langle m+1, r\rangle,$$

$$I_3 = I^c \cap I_-(y^*), \qquad I_4 = I^c \cap I(y^*) \cap \langle 1, m\rangle, \qquad I_5 = I^c \cap \langle m+1, r\rangle.$$

For l large enough, observe first that

$$f_i(y_{n_l}) < 0, \qquad \text{for } i \in I_-(y^*);$$

then, the necessary optimality conditions at y_{n_l} imply the existence of reals

$$\lambda_i^{n_l} \in [0, 1], \qquad \mu_i^{n_l} \in [-1, 1], \qquad \xi_i \in \{1, -1\},$$

and vectors

$$d^{n_l} \in \partial f(y_{n_l}), \; c_i^{n_l} \in \partial f_i(y_{n_l}),$$

such that

$$0 = d^{n_l} + (y_{n_l} - x_{n_l}) + k_{n_l}\left[\sum_{i\in I_1} c_i^{n_l} + \sum_{i\in I_2} a_i\xi_i + \sum_{i\in I_4} \lambda_i^{n_l} c_i^{n_l} + \sum_{i\in I_5} \mu_i^{n_l} a_i\right]. \tag{38}$$

Referring to local boundedness and upper semicontinuity of subdifferentials, without loss of generality we can suppose that there exist

$$\lambda_i \in [0, 1], \qquad \mu_i \in [-1, 1], \qquad d \in \partial f(y^*), \qquad c_i \in \partial f_i(y^*),$$

such that

$$\lambda_i = \lim_{l\to\infty} \lambda_i^{n_l}, \qquad \mu_i = \lim_{l\to\infty} \mu_i^{n_l}, \qquad d = \lim_{l\to\infty} d^{n_l}, \qquad c_i = \lim_{l\to\infty} c_i^{n_l}.$$

Then, if we divide both sides of (38) by k_{n_l} and take the limit, we obtain

$$0 = \sum_{i\in I_1} c_i + \sum_{i\in I_2} a_i\xi_i + \sum_{i\in I_4} \lambda_i c_i + \sum_{i\in I_5} \mu_i a_i. \tag{39}$$

Let us prove now that this equality yields a contradiction with the regularity assumptions. Since $\tilde{x}$ and y^* belong to C, $I \subset I(y^*)$, and since f_i is convex, we obtain, when taking the inner product of both members of (39) with $\tilde{x}-y$,

$$0=\sum_{i\in I_1}(c_i, \tilde{x}-y^*)+\sum_{i\in I_4}\lambda_i(c_i, \tilde{x}-y^*)\leq \sum_{i\in I_1} f_i(\tilde{x})+\sum_{i\in I_4}\lambda_i f_i(\tilde{x}).$$

Recall now that $I_1 \cup I_2$ is nonempty; there are two possibilities:

(a1) $I_1 \neq \emptyset$ or $I_1 = \emptyset$ and $\lambda_i > 0$, for some $i \in I_4$; then, the second member of the last inequality is strictly negative, which is impossible;
(a2) $I_1 = \emptyset$ and $\lambda_i = 0$, $\forall i \in I_4$; then, $I_2 \neq \emptyset$ and

$$\sum_{i\in I_2} a_i \xi_i + \sum_{i\in I_5} \mu_i a_i = 0,$$

which is impossible, since a_i are linearly independent and $\xi_i \neq 0$.
(b) Let $\delta(\cdot|C)$ the indicator function of C, that is,

$$\delta(x|C)=0, \qquad \text{if } x\in C,$$

$$\delta(x|C)=+\infty, \qquad \text{if } x\notin C.$$

Then, from part (a), it follows that

$$y_n = \operatorname{prox}(x_n|f+\delta(\cdot|C)). \tag{40}$$

Since by Lemma 2.2

$$\|y_n - x_{n+1}\| \leq \sqrt{(2\epsilon_n)}, \tag{41}$$

the convergence of the whole sequence $\{y_n\}$ is a consequence of Theorem 1 of Ref. 10, and the sequence $\{x_n\}$ converges to the same point by (41).
(ii) Suppose now that $\epsilon_n = 0$, for each n, and that f, f_i are affine functions. Then, from (40), we have

$$x_{n+1} = \operatorname{prox}(x_n|f+\delta(\cdot|C)),$$

and the announced result is a consequence of Proposition 8 of Ref. 10.

□

Remark 2.2. As said in Remark 2.1 for Theorem 2.1, the regularizing function $(1/2)\|x-x_n\|^2$ can be multiplied by a controlling parameter $1/c_n$; the condition

$$\sum_{n=1}^{\infty} \sqrt{(\epsilon_n c_n)} < +\infty$$

ensures the validity of Theorem 2.3.

From a practical point of view, penalty methods such as (10) and (11) tend to be ill conditioned when the penalty parameter k_n tends to infinity. However, in the case of the exact penalty method, this can be remedied. In all that follows, ϕ denotes the penalty function of Theorem 2.3. Let us begin with some preliminary remarks.

Remark 2.3. For $k \geqslant 0$, let $A(k)$ be the optimal set of the optimization problem

$$(P(k)) \quad \inf(f(x) + k\phi(x) \colon x \in R^N).$$

Let $\{\epsilon_n\}$ be a sequence of positive numbers verifying

$$\sum_{n=1}^{\infty} \sqrt{(\epsilon_n)} < +\infty.$$

Starting from an arbitrary point x_0, let

$$y_n = \mathrm{prox}(x_n | f + k\phi),$$

and let x_{n+1} be such that

$$f(x_{n+1}) + k\phi(x_{n+1}) + \tfrac{1}{2}\|x_{n+1} - x_n\|^2 \leqslant f(x) + k\phi(x) + \tfrac{1}{2}\|x - x_n\|^2 + \epsilon_n.$$

As in Lemma 2.2,

$$\|x_{n+1} - y_n\|^2 \leqslant 2\epsilon_n.$$

Applying Theorem 1 of Ref. 10, it follows that:

(a) the sequence $\{x_n\}$ is bounded iff $A(k) \neq \varnothing$;
(b) if $A(k) \neq \varnothing$, the sequence $\{x_n\}$ converges to a point of $A(k)$.

Lemma 2.1 implies the existence of some $\bar{k} \geqslant 0$ such that, for all $k \geqslant \bar{k}$, $f + k\phi$ is inf-compact and $A(k)$ is nonempty. On the other hand, there exists (Ref. 16) $\hat{k} \geqslant \bar{k}$ such that

$$A(k) = S, \qquad \text{for all } k \geqslant \hat{k}.$$

It follows that, if one replaces, in Theorem 2.3, k_n by $\min(k_n, \hat{k})$, then the sequence $\{x_n\}$ remains convergent and converges to a point of S.

Unfortunately, $\hat{k}$ is not known. For this reason, we shall design a new algorithm where the penalty parameter remains constant after a finite number of steps. For this algorithm, we suppose that either C is compact, and we take $\bar{k} = 0$, or $\bar{k}$ is a value such that $f + \bar{k}\phi$ is inf-compact.

Algorithm A1. Let $\gamma \in \,]0, \tfrac{1}{2}[$ be fixed. Starting from $k_0 = \bar{k}$, the sequence $\{k_n\}$ is generated as follows:

$$k_{n+1} = 2k_n + 1, \qquad \text{if } \phi(x_n) > \epsilon_{n-1}^{(1/2)-\gamma},$$

$$k_{n+1} = k_n, \qquad \text{otherwise.}$$

Theorem 2.4. Suppose that the regularity assumptions (a) and (b) of Theorem 2.3 hold and the sequence $\{\epsilon_n\}$ verifies

$$\sum_{n=1}^{\infty} \sqrt{(\epsilon_n)} < \infty.$$

Then:

(a) there exists k^* and n_0 such that $k^* = k_n \forall n \geq n_0$;
(b) the sequence $\{x_n\}$ converges to a point of S.

Proof. (a) Suppose for contradiction that (a) is false. Then, $k_n \to \infty$. By Theorem 2.3, $y_n \in C$ for n large enough, and the sequences $\{x_n\}$ and $\{y_n\}$ converge to the same optimal solution $x^* \in C$. Since ϕ is convex, ϕ is locally Lipschitzian and there is a neighborhood V of x^* and $L > 0$ such that

$$|\phi(u) - \phi(u')| \leq L\|u - u'\|, \qquad \forall u, u' \in V.$$

On the other hand, there is n^* such that $x_n \in V$ and $y_{n-1} \in V \cap C$ for all $n \geq n^*$. But then, $\phi(y_{n-1}) = 0$; and, from Lemma 2.1, it follows that

$$0 \leq \phi(x_n) \leq L\|y_{n-1} - x_n\| \leq L\sqrt{(2\epsilon_{n-1})}.$$

Since $\epsilon_n \to 0$,

$$\phi(x_n) > \epsilon_{n-1}^{1/2-\gamma},$$

for n large enough, which leads to a contradiction.

(b) Now, for n large enough, $k_n = k^*$, and $\phi(x_n) \to 0$. We shall distinguish two cases.

(b1) $C = \{x / \phi(x) \leq 0\}$ is compact. Since ϕ is convex, it is inf-compact. Hence, $\{x_n\}$ is bounded and, from Remark 2.3, converges to some $x^* \in A(k^*)$. Since $\phi(x_n) \to 0$, then $\phi(x^*) = 0$. It follows that $x^* \in S$.

(b2) The function $f + \bar{k}\phi$ is inf-compact. Since $k^* \geq \bar{k}$, then $f + k^*\phi$ is also inf-compact. It follows from Remark 2.3 that $\{x_n\}$ converges to some $x^* \in A(k^*)$. Proceed as above. □

Remark 2.4. It remains to determine $\bar{k}$ when C is not compact. This is trivial when f is inf-compact; in this case, $\bar{k}$ can be taken equal to 0. Another easy case is when f is bounded from below; then, any $\bar{k} > 0$ is suitable. Indeed, there is $\tilde{k} > 0$ such that $f + \tilde{k}\phi$ is inf-compact. Let $\{z_n\}$ be a sequence such that $\|z_n\| \to +\infty$; then,

$$f(z_n) + \tilde{k}\phi(z_n) \to \infty.$$

Since f is bounded from below, then

either $f(z_n) \to +\infty$ or $\phi(z_n) \to +\infty$.

In both cases,

$$f(z_n)+k\phi(z_n)\to\infty, \qquad \text{for all } k>0.$$

Thus, $f+k\phi$ is inf-compact for all $k>0$.

Remark 2.5. The idea of controlling the growth of the penalty parameter is not new; see, for example, Kiwiel (Refs. 17 and 18). One can try to replace the test

$$\phi(x_n)>\epsilon_{n-1}^{(1/2)-\gamma}$$

by

$$\phi(x_n)>\delta_n(x_n),$$

where $\delta_n(x_n)$ is a value related to Algorithm A. Such a kind of test has already been used in another framework by Auslender (Ref. 7) for the method of centers and by Kiwiel (Refs. 17 and 18) for the penalty method. Notice also that, in Kiwiel, the penalty parameter stays constant after a finite number of steps.

Remark 2.6. Under the Slater condition, the solution set B of the dual problem of (P) is known to be compact. Consider, for simplicity, a problem with only inequality constraints, and let

$$k_B=\min[\|u\|_\infty \colon u\in B].$$

Then, Proposition 1 of Ref. 16 says that

$$A(k)=S, \qquad \text{for } k>k_B.$$

If k_B is not too large, it is reasonable to think that k^* in Theorem 2.4 is close to k_B, so that the function $f+k^*\phi$ is not ill conditioned.

3. Computational Efficiency

In this discussion, we consider only the exact penalty method (Theorems 2.3 and 2.4), for which we can expect a good computational efficiency. For simplicity, we suppose that we have only inequality constraints.

3.1. In Theorem 2.4, we have given a rule for controlling the parameter k_n, and there is a large class of problems for which one can expect a good computational behavior. Another reason for thinking that the penalty aspect would not trouble the behavior of the method is suggested by an algorithm given by Mangasarian (Ref. 19) for linear programs.

Consider the linear program

$$(P) \quad \min[(c, x): Ax \geq b].$$

Mangasarian has pointed out that there exists $\bar{\epsilon} > 0$ such that, for $\epsilon \in]0, \bar{\epsilon}[$, the perturbed quadratic program

$$(P(\epsilon)) \quad \min[(\epsilon/2)\|x\|^2 + (c, x): Ax \geq b]$$

has a unique solution $\bar{x}$, not dependent on ϵ, which solves (P).

Taking the dual of (P), he obtains

$$(Q(\epsilon)) \quad \max[(b, u) - (1/2\epsilon)\|A^t u - c\|^2: u \geq 0].$$

Then, he notices that $Q(\epsilon)$ is exactly the exterior penalty problem associated to the dual linear program (Q).

$$(Q) \quad \max[(b, u): A^t u = c, u \geq 0],$$

with the penalty parameter $\alpha = 1/\epsilon$. Usual results on exterior penalty methods require that $\alpha \to +\infty$. Sharper results taking advantage of the linearity of the problem require merely that

$$\alpha \geq \bar{\alpha}, \qquad \text{for some } \bar{\alpha} > 0.$$

Mangasarian gives an algorithm for solving $(Q(\epsilon))$ which, under Slater's condition, generates a sequence $\{u_n(\epsilon)\}$ having cluster points. Let $u(\epsilon)$ be such a point; then, $u(\epsilon)$ is an optimal solution of $(Q(\epsilon))$ and

$$\bar{x} = (1/\epsilon)[A^t u(\epsilon) - c].$$

Notice that the sequences $\{u_n(\epsilon)\}$ and $u(\epsilon)$ depend on $\epsilon \in]0, \bar{\epsilon}[$, but $\bar{\epsilon}$ is unknown.

A common feature appears in Mangasarian's method and ours. Mangasarian uses the Tikonov method to regularize the objective function, whereas we use the proximal method. But, in Theorem 2.4, we are able to determine k^*, whereas this is not done for $\bar{\epsilon}$. It seems that the conditioning of both methods is equivalent. Since the Mangasarian method has a good computational behavior, we have then confidence that our method is efficient as well.

3.2. Augmented Lagrangian methods have been introduced to avoid the numerical instability of the classical penalty method. The usual augmented Lagrangian is defined by

$$L(x, y, c) = f(x) + \begin{cases} \sum y_i f_i(x) + (c/2) f_i(x)^2, & \text{if } y_i + cf_i(x) \geq 0, \\ -(1/2c) y_i^2, & \text{if } y_i + cf_i(x) \leq 0. \end{cases}$$

Multiplier methods compute sequences $\{x^k\}$, $\{y^k\}$ via the formulas

$$x^{k+1} = \arg\min L(x, y^k, c_k), \tag{42}$$

$$y_i^{k+1} = \max(0, y_i^k + c_k f_i(x^{k+1})), \qquad i = 1, 2, \ldots, m, \tag{43}$$

$$0 < c_k \nearrow c_\infty \leq +\infty. \tag{44}$$

The efficiency of such methods has been improved by Rockafellar, who regularizes the function L by introducing the proximal method (Refs. 9–20). The function L in (42) is replaced by F_k, where

$$F_k(x, y^k, c_k) = L(x, y^k, c_k) + (\mu^2/2c_k)\|x - x^k\|^2.$$

In Ref. 9, $\mu = 1$; the multiplier μ was introduced in Ref. 20 to restore flexibility. Indeed, numerical experiments have disclosed that, for $\mu = 1$, the proximal method moves rather slowly in the initial steps by comparison with the usual multiplier method, despite its ultimate convergence properties. Rockafellar explains in Ref. 19 that, "when c_k is too low, the quadratic term in F_k dominates and does not allow the Lagrangian term to have a strong enough effect in the selection of x^{k+1}. On the other hand, when c_k is too high, the penalty aspects of the augmented Lagrangian are too strong and the prime advantage over penalty methods gets lost." We do not need to introduce the parameter μ to restore flexibility, since the sequences $\{k_n\}$ and $\{c_n\}$ are independent. A penalty aspect exists in Theorem 2.4, but there is no reason to believe that this aspect would be worse than for the augmented Lagrangian, so that, if k_B defined in Remark 2.6 is small, one can hope that our method is competitive.

Stopping rules for the proximal method need the computation of $\mathrm{dist}(0, \partial F_k(x_{k+1}))$. Such a computation is easy for differentiable functions; actually, Rockafellar assumes that functions f_i are of this kind. Such an assumption is not needed for Algorithm A (Ref. 7). Besides, no result of finite convergence for linear programs is given for the proximal point algorithms (Refs. 9–20). Such a result exists only for the multiplier method and the primal proximal minimization problem (Ref. 9). Finally, notice that, if Rockafellar obtains, under the assumptions of Theorem 2.4, the convergence both in primal and dual problems, in counterpart, in our method, k_n becomes constant after a finite number of steps.

3.3. Finally, the efficiency of the method would depend on Algorithm A and on the size of the parameter k^*. For Problems (P), in which k^* is not too high, the method would not be ill-conditioned. Concerning the efficiency of Algorithm A, for the general case we think that the algorithm described in Ref. 7 is efficient [this algorithm is close in spirit to the classical methods given as a subroutine for obtaining a descent direction in nonsmooth optimization (bundle, aggregate methods)]. Nevertheless, it can be improved and replaced by other ones. In the linear case, we have at each iteration to minimize on $\mathbb{R}^N$ a function

$$x \to (c, x) + \tfrac{1}{2}\|x - x_n\|^2 + \cdots$$
$$+ k_n \left[\sum_{i=1}^{m} [(a_i, x) - b_i]^+ + \sum_{i=m+1}^{r} |(a_i, x) - b_i| \right].$$

Efficiency finite algorithms could give a finite algorithm competitive with the simplex method or with the method given by Mangasarian (Ref. 19), in particular when k^* is small.

References

1. LEMARECHAL, C., *An Extension of Davidon Methods to Nondifferentiable Problems*, Mathematical Programming Study, Vol. 3, pp. 93–109, 1975.
2. LEMARECHAL, C., *Bundle Methods in Nonsmooth Optimization*, Nonsmooth Optimization, Edited by C. Lemarechal and R. Mifflin, Pergamon Press, Oxford, England, 1978.
3. WOLFE, P., *A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions*, Mathematical Programming Study, Vol. 3, pp. 145–173, 1975.
4. MIFFLIN, R., *A Superlinearly Convergent Algorithm for One-Dimensional Constrained Minimization Problems with Convex Functions*, Mathematics of Operations Research, Vol. 8, pp. 185–195, 1983.
5. LEMARECHAL, C., STRODIOT, J. J., and BIHAIN, A., *On a Bundle Algorithm for Nonsmooth Optimization*, Nonlinear Programming 4, Edited by O. Mangasarian, R. R. Meyer, and S. M. Robinson, Academic Press, New York, New York, pp. 246–282, 1981.
6. KIWIEL, K. C., *An Algorithm for Linearly Constrained Convex Nondifferentiable Minimization Problems*, Journal of Mathematical Analysis and Applications, Vol. 105, pp. 446–452, 1985.
7. AUSLENDER, A., *Numerical Methods for Nondifferentiable Convex Optimization*, Mathematical Programming Study, Vol. 30, pp. 102–127, 1987.
8. MARTINET, B., *Regularization d'Inégalités Variationnelles par Approximations Successives*, Revue Française de Recherche Operationnelle, Vol. 3, pp. 154–159, 1970.
9. ROCKAFELLAR, R. T., *Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming*, Mathematics of Operations Research, Vol. 4, pp. 97–116, 1976.
10. ROCKAFELLAR, R. T., *Monotone Operators and the Proximal Point Algorithm*, SIAM Journal on Control and Optimization, Vol. 14, pp. 877–898, 1976.
11. FUKUSHIMA, M., *A Descent Algorithm for Nonsmooth Convex Programming*, Mathematical Programming, Vol. 30, pp. 163–175, 1984.
12. HUARD, P., *Programmation Mathematique Convexe*, Revue d'Informatique et de Recherche Operationnelle, Vol. 7, pp. 43–59, 1968.

13. KAPLAN, A. A., *On Convex Programming with Internal Regularization*, Soviet Mathematics, Doklady Akademii Nauk, Vol. 19, pp. 795–799, 1975.
14. ROCKAFELLAR, R. T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
15. PCHENITCHNY, B., and DANILINE, Y., *Méthodes Numériques dans les Problèmes d'Extremums*, MIR, Moscow, USSR, 1977.
16. BERTSEKAS, D. P., *Necessary and Sufficient Conditions for a Penalty Method to Be Exact*, Mathematical Programming, Vol. 9, pp. 87–100, 1975.
17. KIWIEL, K. C., *Descent Methods for Constrained Nonsmooth Optimization*, Abstracts of the IIASA Meeting on Nondifferentiable Optimization, Sopron, Hungary, 1984.
18. KIWIEL, K. C., *An Exact Penalty Function Algorithm for Nonsmooth Convex Minimization Problems*, INA Journal of Numerical Analysis, Vol. 5, pp. 111–119, 1985.
19. MANGASARIAN, O. L., *Iterative Solution of Linear Programs*, SIAM Journal on Numerical Analysis, Vol. 18, pp. 606–614, 1981.
20. ROCKAFELLAR, R. T., *Monotone Operators and Augmented Lagrangian Methods*, Nonlinear Programming 3, Edited by O. Mangasarian, R. R. Meyer, and S. M. Robinson, Academic Press, New York, New York, pp. 1–15, 1978.