View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by International Institute for Applied Systems Analysis (IIASA)

International Institute for
Applied Systems Analysis
IIASA
www.iiasa.ac.at

A Variable Metric Method of Centres for Nonsmooth Minimization

Kiwiel, K.

IIASA Collaborative Paper
June 1981

Kiwiel, K. (1981) A Variable Metric Method of Centres for Nonsmooth Minimization. IIASA Collaborative Paper.
Copyright © June 1981 by the author(s). http://pure.iiasa.ac.at/1780/ All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage. All copies must bear this notice and the full citation on the first page. For other purposes, to republish, to post on servers or to redistribute to lists, permission must be sought by contacting repository@iiasa.ac.at

NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

A VARIABLE METRIC METHOD OF CENTRES FOR NONSMOOTH MINIMIZATION

K.C. Kiwiel

June 1981
CP-81-23

Collaborative Papers report work which has not been performed solely at the International Institute for Applied Systems Analysis and which has received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

A VARIABLE METRIC METHOD OF CENTERS FOR NONSMOOTH MINIMIZATION

K.C. Kiwiel

INTRODUCTION

We consider the problem of minimizing f on $S = \{x \in \mathbb{R}^N : h(x) \leq 0\}$ where $f:S \to \mathbb{R}$ and $h:\mathbb{R}^N \to \mathbb{R}$ are locally Lipschitz continuous functions. We present an implementable modification of an algorithm constructed by Lemarechal (1978) and further extended by Mifflin (1979) and show that the algorithm's accumulation points are stationary if f_0 and h are weakly upper semismooth The method is a feasible point descent method which combines a generalization of the method of centers with quadratic approximation of some Lagrangian function in the spirit of (Lemarechal 1978). A simplified variant of the algorithm may be interpreted as an application of Shor's variable metric techniques (Shor 1979) to Wolfe's method of conjugate subgradients (Wolfe 1975). Our version differs from Lemarechal's and Mifflin's algorithms (Lemarechal 1978; Mifflin 1979), because of its rules for updating of the search direction finding subproblem. More specifically, our version does not require unlimited storage of gradient information, contrary to (Lemarechal 1978; Mifflin 1979). Instead, its storage requirements are flexible and may be controlled by a user. To this end we introduce rules for reduction or aggregation of gradient information, which necessitate new techniques of convergence analysis. We also give rules for variable metric updating. Preliminary numerical results seem to validate the approach presented in this paper.

The algorithm requires a feasible starting point, i.e., an $x^0 \in S$, but f need not be defined for $x \notin S$, which is important in some applications.

In sec. 2 we give definitions and preliminary results. The algorithm is defined in sec. 3 and in sec. 4 we discuss details of its implementations and how it compares with the methods of (Lemarechal 1978; Mifflin 1979). In sec. 5 we prove stationarity of its accumulation points. Numerical results are presented in sec. 6.

2. DEFINITIONS AND PRELIMINARY RESULTS

Throughout the paper we mostly adhere to the now standard notation in (Mifflin 1979; Clarke 1976; Clarke 1975). The scalar product of $u = (u_1,\dots,u_N)$ and $v = (v_1,\dots,v_N)$ in $\mathbb{R}^N$, defined by $\sum_{i=1}^{N} u_i v_i$ is denoted by $\langle u,v \rangle$ and the Euclidean norm of u, defined by $\langle u,u \rangle^{\frac{1}{2}}$, is denoted $|u|$. $B(x,\varepsilon) = \{x' \in \mathbb{R}^N : |x' - x| < \varepsilon\}$ is an open ball with centre x and radius ε. A convex hill of a set $W \subset \mathbb{R}^N$ is denoted conv(W). For any symmetric positive definite $N \times N$ matrix A, $\langle .,. \rangle_A$ denotes the scalar product induced by A, i.e., $\langle u,v \rangle_A = \langle Au,v \rangle$ for $u,v \in \mathbb{R}^N$, and $|u|_A = \langle u,u \rangle_A^{\frac{1}{2}}$. A^* denotes the adjoint of A and I the identity matrix. $\lambda_{min}(A)$ and $\lambda_{max}(A)$ denote the minimal and the maximal eigenvalues of A, respectively.

For any set $W \subset \mathbb{R}^N$ and a symmetric positive definite matrix A, there is a unique point w in the closure of conv (W) having minimum $|\cdot|_A$-norm; it will be denoted by $Nr_A W$. Algebraically, the point w is characterized by the relation

$$(2.1) \qquad \langle u,w \rangle_A \geq |w|_A^2 \text{ for all } u \in W.$$

Let $F : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function (Clarke 1976; Clarke 1975), i.e., for each bounded subset $B \subset \mathbb{R}^N$ there exists a constant L such that $|F(y) - F(z)| \leq L|y - z|$ for all $y,z \in B$.

The generalized gradient of F at x (Clarke 1976; Clarke 1975) $\partial F(x)$, is the convex hull of the set of limits of sequences of the form $\{\nabla F(x^k) : x^k \to x$ and F is differentiable at $x^k\}$. The point-to-set mapping ∂F is uppersemicontinuous and locally bounded (Clarke 1976; Clarke 1975).

As in (Mifflin 1979; Mifflin 1977), we say a point $\bar{x} \in S$ is stationary for f on S if $0 \in M(\bar{x})$ where

$$(2.2) \qquad M(x) = \begin{cases} \partial f(x) & \text{if } h(x) < 0 \\ \text{conv}\{\partial f(x) \cup \partial h(x)\} & \text{if } h(x) = 0 \\ \partial h(x) & \text{if } h(x) > 0 \end{cases},$$

because $0 \in M(\hat{x})$ is a necessary condition for $\hat{x} \in S$ to minimize f on S. The point-to-set mapping M is uppersemicontinuous and locally bounded (Mifflin 1979; Mifflin 1977).

In order to implement the algorithm, we suppose that we have subroutines that can evaluate functions $g_f(x) \in \partial f(x)$ for $x \in S$ and $g_h(x) \in \partial h(x)$ for $x \in \mathbb{R}^N$.

Associated with f, h, g_f and g_h let $\alpha : S \times \mathbb{R}^N \to \mathbb{R}_+$ be a non-negative-valued function

$$(2.3) \qquad \alpha(x,y) = \begin{cases} |f(x) - f(y) - \langle g(y), x-y \rangle + h_+(x)| & \text{if } g(y) = g_f(y) \\ |h_+(x) - h(y) - \langle g(y), x-y \rangle| & \text{if } g(y) = g_h(y) \end{cases}$$

where $h_+(x) = \max\{h(x), 0\}$. α is a measure of deviation from linearity. Note that it differs substantially from its counterpart introduced in (Mifflin 1979).

Associated with the sequence of points generated by the algorithm $\{x^k\}_{k=0}^{\infty}$ let

$$(2.4) \qquad \phi_k(x) = \max\{f(x) - f(x^k), h(x)\}$$

be the distance function of the modified method of centres (Pironneau and Polak 1972), and let

$$(2.5) \qquad M_k(x) = \begin{cases} \{g_f(x)\} & \text{if} \quad f(x) - f(x^k) = \phi_k(x) \\ \{g_h(x)\} & \text{if} \quad f(x) - f(x^k) < \phi_k(x) \end{cases}$$

be an algorithmic mapping.

3. THE ALGORITHM

Let m_a, m_c, m_L, m_k, ε_0, $\bar{\kappa}$ be fixed positive parameters satisfying $m_L < m_R < 1$. Let $\bar{M}_g$ denote the maximum number of gradients that the algorithm is allowed to store in a set G for direction finding; $\bar{M}_g \geq 1$. Let $\bar{M}_{up}$ denote the maximum number of variable metric updatings.

Suppose initially that $x^0 \in S$ and let A_0 be a positive definite $N \times N$ matrix. Let $p^0 = g_f(x^0)$ and $d^0 = -H_0 p^0$, where $H_0 = A_0^{-1}$. Note that $p^0 \in M_0(x^0)$. We suppose that $g_f(x^0) \neq 0$, and hence $v^0 = -|p^0|^2_{H_0} < 0$; otherwise x^0 would be stationary. Set $\kappa^0 = \bar{\kappa}$ and choose an estimated shift in x at the first iteration $s^0 > 0$. Set $G^0 = \emptyset$, $A^0 = \emptyset$, $a^0 = 0$ and $M_g^0 = 0$. Set $k = 0$ and proceed according to the instructions given below.

Step 1 (line search). By a line search procedure discussed below, find two stepsizes t_i^k and t_R^k such that $0 \leq t_L^k \leq t_R^k$ and such that the two corresponding points defined by

$$(3.1) \qquad y_L^k = x^k + t_L^k d^k \text{ and } y_R^k = x^k + t_R^k d^k$$

satisfy

$$(3.2a) \qquad |y_R^k - y_L^k| \leq \kappa^k |y_i^k - x^k| \quad \text{if } t_i^k > 0 \;,$$

$$(3.2b) \qquad |y_R^k - y_L^k| \leq \kappa^k s^k \qquad \text{if } t_L^k = 0$$

and

$$(3.3) \qquad \phi_k(y_L^k) \leq \phi_k(x^k) + m_L t_L^k v^k$$

and

$$(3.4) \qquad -\alpha(y_L^k, y_R^k) + \langle g(y_R^k), d^k \rangle \geq m_R v^k \quad \text{with } g(y_R^k) \in M_k(y_R^k) \quad .$$

Step 2. If $t_L^k = 0$, set $s^{k+1} = s^k$ and $\kappa^{k+1} = \kappa^k/2$; otherwise, i.e., if

$$t_L^k > 0, \text{ set } s^{k+1} = |y_L^k - x^k| \quad \text{and } \kappa^{k+1} = \bar{\kappa} \quad .$$

$$\text{Set } x^{k+1} = y_L^k \quad \text{and } a^{k+1} = a^k + t_R^k |d^k| \quad .$$

Step 3. (Bundle augmentation). If $M_g^k = \bar{M}_g$, delete one element $g(y_R^j)$ from G^k (e.g., the oldest one) and $\alpha(x^k, y_R^j)$ from A^k, replacing M_g^k by $M_g^k - 1$.

$$\text{Set } G^{k+1} = G^k \cup \{g(y_R^k\} \text{ and } A^{k+1} = A^k \cup \{\alpha(x^{k+1}, y_R^k)\} \text{ and}$$

$$M_g^{k+1} = M_g^k + 1 \quad .$$

Update the elements of A^{k+1} so that if $\alpha(x^k, y_R^j)$ corresponds to $g(y_R^j)$, then it is substituted by $\alpha(x^{k+1}, y_R^j)$ according to (2.3). Set α_p^{k+1} equal to the mean value of the elements of A^{k+1}.

Step 4. (Resetting tests). If either of the two tests given below

$$(3.5) \qquad f_0(x^k) - f_0(x^{k+1}) \geq m_c |v^k| \quad , \text{ or}$$

$$(3.6) \qquad a^{k+1} > m_a |v^k|$$

is satisfied, go to Step 8.

Step 5 (Direction finding). Solve for $(d,v) = (d^{k+1}, v^{k+1}) \in \mathbb{R}^{N+1}$ the k^{th} quadratic programming subproblem:

minimize

(3.7a) $$\tfrac{1}{2}|d|^2_{A_k} + v$$

subject to

(3.7b) $$-\alpha(x^{k+1}, y^j) + \langle g(y^j_R), d \rangle \leqslant v \text{ for } M_g^{k+1} \text{ elements } g(y^j_R) \in G^{k+1}$$

and

(3.7c) $$-\alpha_p^{k+1} + \langle p^k, d \rangle \leqslant v \quad .$$

Set λ^{k+1} equal to the dual variables of the subproblem (3.7) (see sec. 4) and

(3.8) $$p^{k+1} = -A_k d^{k+1} \quad .$$

Step 6 (Resetting tests). If either of the two tests given below

(3.9a) $$|p^{k+1}| \leqslant \varepsilon_0 \quad , \text{ or}$$

(3.9b) $$a^{k+1} \geqslant m_a |v^{k+1}|$$

is satisfied, go to Step 8.

Step 7. Update A_k, as discussed below, to get a positive definite A_{k+1}, set $H_{k+1} = A_{k+1}^{-1}$ and

(3.10) $$d^{k+1} = -H_{k+1}p^{k+1}$$

and

$$v^{k+1} = -|d^{k+1}|^2_{A_{k+1}} - \sum_{y(y_R^j)\in G^{k+1}} \lambda_j^{k+1}\alpha(x^{k+1},y_R^j) - \lambda_p^{k+1}\alpha_p^{k+1} .$$

Replace k by k+1 and go to Step 1.

Step 8 (Resetting). If the number of updatings of A_k since its last reinitialization exceeds $\bar{M}_{up}$, reinitialize the variable metric by setting $A_k = I$. Solve for (d^{k+1},v^{k+1}) the subproblem (3.6a) and (3.6b), and set $p^{k+1} = -A_k d^{k+1}$. If $|p^{k+1}| \leq \varepsilon_0$ and $G^{k+1} = M_{k+1}(x^{k+1})$, then stop. If $|p^{k+1}| > \varepsilon_0$, set $a^{k+1} = 0$ and update A_k to get a positive definite A_{k+1}. Compute d^{k+1} and v^{k+1} by (3.10), with $\lambda_p^{k+1} = 0$ (see sec. 4), replace k by k + 1 go to Step 1. If $|p^{k+1}| < \varepsilon_0$ and $M_y \geq 1$, then delete the oldest element of G^{k+1} and the corresponding element of A^{k+1} and replace M_g^{k+1} by $M_g^k - 1$. If $M_g^{k+1} = 0$, then set $G^{k+1} = M_{k+1}(x^{k+1})$ and $M_g^{k+1} = 1$. Repeat Step 8 from the beginning.

4. REMARKS ON THE ALGORITHM

A complete analysis of the direction finding subproblem (3.7) may be found in (Wierzbicki 1978). The k^{th} subproblem dual is to find values of the dual variables (multipliers) $\lambda \in \mathbb{R}^{M_g^{k+1}+1}$ to minimize

(4.1) $$\tfrac{1}{2}\Big|\sum_{g(y_R^j)\in G^{k+1}} \lambda_j g(y_R^j) + \lambda_p p^k\Big|^2_{H_k} + \sum_{g(y^j)\in G^{k+1}} \lambda_j \alpha(x^{k+1},y_R^j) + \lambda_p \alpha_p^{k+1}$$

subject to

$$\sum_{g(y_k^j)\in G^{k+1}} \lambda_j + \lambda_p = 1 \text{ and } \lambda_j \geq 0 \text{ for } g(y_R^j)\in G^{k+1},\ \lambda_p \geq 0 .$$

Let λ^{k+1} be some solution of (4.1). Then

$$v^{k+1} = -|d^{k+1}|^2_{A_k} - \sum_{g(y_R^j)\in G^{k+1}} \lambda_j^{k+1}\alpha(x^{k+1},y_R^j) - \lambda_p^{k+1}\alpha_p^{k+1} , \tag{4.2a}$$

$$p^{k+1} = -A_k d^{k+1} = \sum_{g(y_R^j)\in G^{k+1}} \lambda_j^{k+1} g(y_R^j) + \lambda_p^{k+1} p^k . \tag{4.2b}$$

For the subproblem (3.7a) and (3.7b) used at Step 8, we put $\lambda_p^{k+1} = 0$ in (4.1) and (4.2).

Note that the dual subproblem (4.1) has at most $\bar{M}_g + 1$ unknowns, where $\bar{M}_g$ is set up by a user, whereas in (Lemarechal 1978; Mifflin 1979) the size of the subproblem equals k and grows to infinity.

We shall now show that our algorithm is an extension of the modified method of centres ((Pironneau and Polak 1972), done in the spirit of (Lemarechal 1978). Suppose that

$$f(x) = \max_{i=1;n} f_i(x) \text{ and } h(x) = \max_{i=1;m} h_i(x) \tag{4.3}$$

where $f_i:\mathbb{R}^N \to \mathbb{R}$ and $h_i:\mathbb{R}^N \to \mathbb{R}$ are continuously differentiable. In (Kiwiel 1981), we have presented an extension of the method of centres to this case, in which the search direction d^k is computed by solving

$$\text{minimize } \tfrac{1}{2}|d|^2_{A_k} + v \tag{4.4a}$$

subject to

$$-[f(x^k) - f_i(x^k) + h_+(x^k)] + \langle \nabla f_i(x^k), d\rangle \leqslant v \tag{4.4b}$$

$$i \in I_0(x^k, \epsilon^k) \quad ,$$

$$(4.4c)\qquad -[h_+(x^k) - h_i(x^k)] + \langle \nabla h_i(x^k), d\rangle \leq v \quad i \in I_c(x^k,\varepsilon^k) \quad ,$$

where the activity sets I_0 and I_c are defined by

$$I_0(x,\varepsilon) = \{i : f_i(x) - f(x) \geq h_+(x) - \varepsilon \ , 1 \leq i \leq n\} \quad ,$$

$$I_c(x,\varepsilon) = \{i : h_i(x) \geq h_+(x) - \varepsilon \ , \ 1 \leq i \leq m\} \quad ,$$

and $\varepsilon^k \geq \underline{\varepsilon} > 0$ is an activity variable. The stepsize t^k is then computed by an Amijo-type rule so that $x^{k+1} = x^k + t^k d^k$ satisfies

$$(4.5)\qquad \phi_k(x^{k+1}) \leq \phi_k(x^k) + m_1 t^k v^k \quad .$$

Assuming that $\{A_k\}$ are uniformly positive definite and bounded, we prove in (Kiwiel 1981) that every accumulation point of the above algorithm is stationary and that, under additional regularity assumptions, the algorithm converges linearly. Moreover, we noted that by Wierzbicki's results (Wierzbicki 1978), (4.4) may be interpreted as a quadratic approximation direction finding subproblem for the function ϕ_k at x_k, which in turn approximates Ioffe's Lagrangian (Ioffe 1979),

$$(4.6)\qquad \phi(x) = \max\{f(x) - f(\hat{x}), h(x)\} \quad ,$$

where $\hat{x}$ minimizes f on s. Therefore, the results of (Wierzbicki 1978) suggest that in order to obtain faster convergence, the variable metric A_k should approximate the Hessian $L(\hat{x},\hat{\lambda})$ of the normal Lagrange function L for the problem of minimizing f on s, i.e.

$$(4.7)\qquad L(\hat{x},\hat{\lambda}) = \sum_{i=1}^{n} \hat{\lambda}_i f_i(\hat{x}) + \sum_{i=1}^{m} \hat{\lambda}_{i+n} h_i(\hat{x}) \quad ,$$

where $\hat{\lambda} \in \mathbb{R}^{m+n}$ is an optimal Lagrange multiplier [see (Clarke 1976)] satisfying

$$(4.8) \qquad \hat{\lambda} \geq 0 \text{ and } \sum_{i=1}^{m+n} \hat{\lambda}_i = 1 \quad .$$

To see the relevance of the above results for the algorithm presented in this paper, we start by showing that the subproblem (3.7) is an approximation of the subproblem (4.4). By (4.3) and (Clarke 1975, Theorem 2.1),

$$(4.9) \qquad \partial f(x) = \text{conv}\ \{\nabla f_i(x) : f_i(x) = f(x)\} \text{ and}$$

$$\partial h(x) = \text{conv}\ \{\nabla h_i(x) : h_i(x) = h(x)\} \quad .$$

If some h^j is close to x^k, linearization of f_i and h_i at y^j gives

$$(4.10) \qquad f_i(x^k) \simeq f_i(y^j) + \langle \nabla f_i(y^j), x^k - y^j \rangle \text{ and}$$

$$h_i(x^k) \simeq h_i(y^j) + \langle \nabla h_i(y^j), x^k - y^j \rangle \quad .$$

Now (4.9) implies that we may suppose that $g_f(y^j) = \nabla f_i(y^j)$ with $f_i(y^j) = f(y^j)$, or that $g_h(y^j) = \nabla h_i(y^j)$ with $h_i(y^j) = h(y^j)$. If we further assume that $\nabla f_i(x^k) \simeq \nabla f_i(y^j)$ or $\nabla h_i(x^k) \simeq \nabla h_i(y^j)$, which is justified when f_i, h_i are continuously differentiable and y^j is close to x^k, then collecting the above results we may write that

$$(4.11a) \qquad -[f(x^k) - f_i(x^k) + h_+(x^k)] + \langle \nabla f_i(x^k), d \rangle \simeq$$

$$-[f(x^k) - f(y^j) - \langle g_f(y^j), x^k - y^j \rangle + h_+(x^k)] + \langle g_f(y^j), d \rangle \quad ,$$

$$(4.11b) \qquad -[h_+(x^k) - h_i(x^k)] + \langle \nabla h_i(x^k), d \rangle \cong$$

$$[h_+(x^k) - h(y^j) - \langle g_h(y^j), x^k - y^j \rangle] + \langle g_h(y^j), d \rangle \quad .$$

Note that the bracketed terms on the left-hand side of (4.11) are nonnegative by (4.3). If we assume that their right-hand side counterparts are also nonnegative, then they are equal to $\alpha(x^k, y^j)$ defined by (2.3) and therefore (4.11) implies that the subproblem (3.7a) and (3.7b) is an approximation of the subproblem (4.4).

On the other hand, a closer inspection of the dual subproblem (4.1) shows that $g(y_R^j)$ - s with relatively smaller $\alpha(x^k, y^j)$ - s tend to contribute more to the direction d^{k+1}, since corresponding λ_j^{k+1} - s are larger (cf. (4.2b)). This fact provides another argument for using the absolute value in (2.3).

Although our algorithm is designed for functions of more general nature than that given by (4.3), we like to think of Lipschitz functions as if they were pointwise maxima of infinite collecitons of smooth functions. A straightforward extension of the above approach may be based on the observation that one may re-define the activity sets in (4.4) by putting $I_0(x,\varepsilon) = \{1,...n\} = \bar{I}_0$ and $I_c(x,\varepsilon) = \{1,..,m\} = \bar{I}_c$ without impairing the convergence of the algorithm in (Kiwiel 1981). Hence, in the general case, one may try to construct the activity sets by memorization, i.e., use all previously computed $g(y^j)$ - s and $f(y^j)$ - s or $h(y^j)$ - s for direction finding. This is done by Lemarechal (1978) and Mifflin (1979). We follow a different path, discarding the oldest information at Steps 3 and 8 and aggregating it by the use of the constraint (3.7c) in direction finding, since by (4.1) and (4.2b), a gradient deleted from G^{k+1} at Step 3 may still contribute to p^{k+1}, and hence to d^{k+1}, through its influence on p^k.

We shall now address the important question of the choice of the variable metric, using the results of the analysis of the "explicit" case (4.3). We start by noting that if λ_T^{k+1} denotes a Lagrange multiplier in the subproblem (4.4), then under regularity assumptions $\lambda_T^k \to \hat{\lambda}$, see (Kiwiel 1981). Therefore the use of some quasi-Newton formula which constructs A_{k+1} from A_k, $x^{k+1} - x^k$ and $L_x(x^{k+1}, \lambda_T^{k+1}) - L_x(x^k, \lambda_T^{k+1})$ is reasonable (Wierzbicki 1978). In the more general case, from (4.2b) and (4.7) we see that p^{k+1} approximates $L_x(x^{k+1}, \lambda_T^{k+1})$.

However, there is no quantity corresponding to $L_x(x^k, \lambda_T^{k+1})$, hence we consider using $p^{k+1} - p^k$ for variable metric updating.

On the other hand, since p^k may be interpreted as an element of the generalized gradient of a nonsmooth analogue of the Lagrangian (4.7), the use of Shor's famous variable metric (Shor 1979) based on the difference of two successive gradients, i.e., $p^{k+1} - p^k$ in our case, immediately suggests itself. Thus we take

$$(4.12) \qquad H_k = B_k B_k^*$$

where B_k is an $N \times N$ matrix updated in the following way. Introduce the operator of space dilation $R_\beta(\xi)$ in a direction $\xi \in \mathbb{R}^N, |\xi| = 1$ and a coefficient of space dilation $\beta \in [0,1]$ by

$$(4.13a) \qquad R_\beta(\xi)v = \beta \langle v,\xi \rangle \xi + (v - \langle v,\xi) = (\beta - 1)\langle v,\xi \rangle \xi + v \quad ,$$

or, in matrix form,

$$(4.13b) \qquad R_\beta(\xi) = I + (\beta - 1)\xi\xi^T \quad .$$

Then, following Shor (1979), we choose a fixed $\beta \in (0,1)$ and take

$$(4.14) \qquad B_{k+1} = B_k R_\beta(\xi_{k+1}) \text{ and } B_0 = I \quad ,$$

with the direction ξ_{k+1} satisfying

$$(4.15) \qquad \xi_{k+1} = \frac{B_k^*(p^{k+1} - p^k)}{|B_k^*(p^{k+1} - p^k)|} \quad .$$

It is quite easy to check that (4.12), (4.13) and (4.15) imply that

$$(4.16)\qquad \beta^2 |u|^2_{H_k} \leq |u|^2_{H_{k+1}} \leq |u|^2_{H_k} \text{ for any } u \in \mathbb{R}^N \quad .$$

Therefore we adopt the following strategy. During the run of the algorithm, the variable metric matrix is updated at most $\bar{M}_{up}$ - times, counting from its last reinitialization at Step 8. Therefore (4.16) implies that

$$(4.17)\qquad \beta^{2\bar{M}_{up}} \leq \lambda_{min}(H_k) \leq \lambda_{max}(H_k) \leq 1 \quad ,$$

hence $\{A_k\}$ and $\{B_k\}$ are uniformly positive definite and bounded.

Due to limited space, we shall not discuss details of possible line search procedures used at Step 1. It suffices to mention that Mifflin's procedures from (Mifflin 1979) or (Mifflin 1977) may be easily adapted to suit our needs expressed by (3.1) through (3.4). For example, take $\zeta < \min\{\frac{1}{2}, \kappa^k s^k/|d^k|\}$ in the procedure in (Mifflin 1979:9), substitute f by ϕ_k and delete h from its description. One may also check that the conditions for finite termination of that procedure do not change, i.e., f and h should be weakly upper semismooth [see (Mifflin 1979) or (Mifflin 1977) for the definition].

We shall now discuss the resetting tests which enable the algorithm to drop obsolete gradient information at Step 8. The test (3.5) allows resetting each time when there is sufficient decrease in the objective function value. The resetting tests (3.6) and (3.9b) force resetting when the bundle G^{k+1} is not local, i.e., a^{k+1} is large compared with $|v^k|$ or $|v^{k+1}|$, and hence the deletion of some old gradients is justified. The resetting test $|p^{k+1}| \leq \varepsilon_0$, ε_0 being of the order of machine zero, which appears at Steps 6 and 8, is used to force a resetting when p^{k+1} may be meaningless due to round-off errors. Its second purpose is to force $|x^{k+1} - x^k| \to 0$, as shown in the next section.

5. CONVERGENCE

Since $M(x)$ is a convex compact set for any $x \in \mathbb{R}^N$ (Mifflin 1977:Proposition 2.7), a point $\bar{x} \in S$ is stationary for f on S if and only if

(5.1) $$Nr_I M(\bar{x}) = 0 \quad .$$

We say that a point $\bar{x} \in S$ is ε_0-stationary for f on S if

(5.2) $$|Nr_I M(\bar{x})| \leq \varepsilon_0 \quad .$$

If the algorithm stops at Step 8, then by (4.1), (4.2b), the stopping rule and (2.5), we have $p^{k+1} \in \{g_f(x^{k+1})\} \cup \{g_h(x^{k+1})\}$ and $|p^{k+1}| \leq \varepsilon_0$, hence x^{k+1} is ε_0-stationary. Below we shall show that if $x^0 \in S$ then any $x^k \in S$. Summing up, we see that if the algorithm stops, then its last point is feasible and ε_0-stationary.

From now on we suppose that the algorithm does not terminate. Then we have the following convergence theorem.

THEOREM 5.1. Suppose that $\{g(y_R^k)\}_1^\infty$ is uniformly bounded. Then every accumulation point of $\{x^k\}$ is feasible and ε_0-stationary for f on S. The set of all accumulation points of $\{x^k\}$ is closed and connected and f is constant on this set.

Proof. To obtain contradiction, suppose that $\{x^k\}$ has some accumulation point $\bar{x}$ which is not ε_0-stationary, i.e., $x^k \to \bar{x}$ $k \in K_1$ and

(5.3) $$|Nr_I M(\bar{x})| > \varepsilon_0 \quad .$$

(A) We start by showing that the algorithm is regular, i.e., that

(5.4) $$|x^{k+1} - x^k| \to 0 \text{ as } k \to \infty \quad .$$

On entering Step 1, $|p^k| > \varepsilon_0$ by (3.9a) and the rules of Step 8. Since $|p^k|^2_{H_k} = |d^k|^2_{A_k}$ by (3.10), and (4.17) implies that $|p^k|^2_{H_k} \geqslant \beta^{2\bar{M}_{up}}|p^k|^2$, we conclude that on entering Step 1

$$(5.5) \qquad |p^k|^2_{H_k} = |d^k|^2_{A_k} \geqslant \beta^{2\bar{M}_{up}}\varepsilon_0^2 = \varepsilon_1 \quad .$$

Since $\alpha(x,y)$ is nonnegative by (2.3) and $\lambda^{k+1} \geqslant 0$, (3.10) and (5.5) imply that at Steps 1 and 7

$$(5.6) \qquad v^k \leqslant -|p^k|^2_{H_k} \leqslant -\varepsilon_1 \quad .$$

Suppose that $x^k \in S$. Then $\phi_k(x^k) = h_+(x^k) = 0$ and (3.3) with (5.6) imply that $h(x^{k+1}) = h(y_i^k) \leqslant 0$, since $t_i^k \geqslant 0$. Therefore $x^{k+1} \in S$ and if $x^0 \in S$, then

$$(5.7) \qquad \phi_k(x^k) = h_+(x^k) = 0 \quad \text{for all } k.$$

Now (2.4), (5.6) and (5.7) imply that

$$(5.8) \qquad f(x^{k+1}) - f(x^k) \leqslant m_i t_i^k v^k \leqslant 0 \quad \text{for all } k.$$

which together with (5.7) proves that every accumulation point of $\{x^k\}$ has the same f-value and is feasible. By (3.1) and (4.17) and $|p^k| > \varepsilon_0$

$$(5.9) \qquad t_i^k|p^k|^2_{H_k} \geqslant t_i^k\lambda_{min}(M_{k-1})|p^k|^2 \geqslant \beta^{2\bar{M}_{up}}\varepsilon_0|x^{k+1} - x^k| \quad .$$

Since $f(x^k) \to f(\bar{x})$ $k \in K_1$, (5.6), (5.8) and (5.9) imply (5.4).

(B) We now prove that

(5.10) $$|y_R^k - x^{k+1}| \to 0 \text{ as } k \to \infty \quad .$$

If the algorithm takes an infinite number of serious steps with $t_i^k > 0$, the rules of Step 2 and (5.4) show that the sequence of shifts $\{s^k\}$ converges to zero. On the other hand, if $t_i^k = 0$ for almost all k, then subsequent halving κ at Step 2 forces $\kappa^k \to 0$. Since $\kappa^k \leq \bar{\kappa}$ in both cases and (3.2) implies that

$$|y_R^k - x^{k+1}| \leq \kappa^k \max \{s^k, s^{k+1}\} \quad ,$$

the validity of (5.10) is established.

(C) We shall now consider asymptotic properties of the sets G^k and A^k. Define auxilliary variables

(5.11) $$\bar{a}^k = \max \{|y_R^j - x^{k+1}| : G(y_R^j) \in G^{k+1}\} \quad .$$

(5.12) $$\bar{\alpha}^k = \max \{\alpha(x^{k+1}, y_R^j) : g(y_R^j) \in G^{k+1}\},$$

$$\underline{\alpha}^k = \min \{\alpha(x^{k+1}, y_R^j) : g(y_R^j) \in G^{k+1}\} \quad ,$$

(5.13) $$\hat{\alpha}^k = \bar{\alpha}^k - \underline{\alpha}^k \quad .$$

Since G^k contains at most $\bar{M}_g$ elements, (5.4), (5.10) and (5.11) imply

(5.14) $$\bar{a}^k \to 0 \text{ as } k \to \infty \quad .$$

We shall now prove that, given two positive numbers $\tilde{\alpha}$ and ε and positive integers N_1 and N_2, there exists an integer $N_3 \geq N_1$ such that for $k = N_3, N_3 + 1, \ldots, N_3 + N_2$

$$(5.15a) \qquad x^k, y_R^k \in B(\bar{x}, \varepsilon)$$

$$(5.15b) \qquad \max\{\alpha^k, \bar{\alpha}^k, \hat{\alpha}^k\} \leq \tilde{\alpha}$$

$$(5.15c) \qquad g(y_R^j) = g_f(y_R^j) \text{ for } g(y_R^j) \in G^k \text{ if } h(\bar{x}) < 0 \quad .$$

(5.15a) follows from $x^k \to \bar{x}$ $k \in K_1$ and (5.4) with (5.10). (5.15b) follows from $x^k \to \bar{x}$ $k \in K_1$, (5.4) and (5.10), the assumed boundedness of $\{g(y^k)\}_{k=0}^{\infty}$, the definition of $\alpha(x,y)$, (2.5) and (3.4). In particular, if $h(\bar{x}) < 0$, then for sufficiently large $k \in K_1$ we must have $h(y_R^k) < h(\bar{x})/2$ by (5.10), hence (2.5) and (3.4) imply that $g(y_R^k) = g_f(g_R^k)$ and one need only consider the upper part of (2.3), which also proves (5.15c).

(D) We shall now analyze asymptotic properties of the subproblem (3.7). By (3.10) and (5.12)

$$(5.16) \qquad v^k \geq -(|p^k|_{H_k}^2 + \bar{\alpha}^k)$$

at Step 1, so (3.4) and (3.10) imply

$$(5.17) \qquad \langle g(y_R^k), -p^k \rangle_{H_k} \geq -m_R(|p^k|_{H_k}^2 + \bar{\alpha}^k) + \alpha(x^{k+1}, y_R^k) \quad .$$

Since $\alpha(x^{k+1}, y_R^k) \geq 0$ and $|p^k|_{H_k}^2 \geq \varepsilon_1$ by (5.5), if some constant m_R' satisfies $m_R' \in (m_R, 1)$ and

$$(5.18) \qquad \bar{\alpha}^k \leq (m_R' - m_R)\varepsilon_1/m_R \quad ,$$

then

$$(5.19) \qquad \langle g(y_R^k), -p^k \rangle_{H_k} \geq -m_R' |p^k|_{M_k}^2 \quad .$$

Introduce an auxilliary variable p_0^k by

(5.20) $$p_0^{k+1} = N_{r_{H_k}}[G^{k+1}\nu\{p^k\}] \quad .$$

Then (4.1) implies

(5.21) $$\tfrac{1}{2}|p^{k+1}|^2_{H_k} \leq \tfrac{1}{2}|p_0^{k+1}|^2_{H_k} + \hat{\alpha}^k \quad .$$

Since $\{g(y_R^k)\}$ is bounded by assumption, (4.17) implies the existence of a constant $C_g < +\infty$ satisfying

(5.22) $$|g(y_R^k)|_{H_k} \leq C_g \text{ for all } k.$$

Suppose now that $\bar{\alpha}^k$ and $\hat{\alpha}^k$ and a constant $\tilde{\alpha}$ satisfy

(5.23) $$\max\{\bar{\alpha}^k,\hat{\alpha}^k\} \leq \tilde{\alpha} \leq (m_R' - m_R)\varepsilon_1/m_R \quad .$$

Then (5.18) through (5.23) and [3, Lemma 4.4] imply that p^{k+1} solving (3.7) satisfies

(5.24) $$|p^{k+1}|^2_{H_k} \leq \phi(|p^k|^2_{H_k}) + 2\tilde{\alpha} \quad ,$$

where the function $\phi:\mathbb{R} \to \mathbb{R}$ is defined by

(5.25) $$\phi(t) = t\{1 - (1-m)^2 t/4C_g^2\} \quad .$$

One immediately checks that $\phi(t) < t$ for $t \in (0,C_g^2]$ and that $|p^k|^2_{H_k} \leq C_g^2$ by (4.26), (4.17) and (5.22).

For a given $\tilde{\alpha} > 0$, define a scalar $t(\tilde{\alpha})$ by $t(\alpha) = \phi(t(\tilde{\alpha})) + 2\tilde{\alpha}$. Then it is easy to show that $t(\tilde{\alpha}) \to 0$ as $\tilde{\alpha}\downarrow 0$ and that if $\tilde{\alpha} > 0$ is sufficiently small, then any sequence of scalars $\{t_i\} \subset \mathbb{R}_+$

$t_0 \leq c_g^2$ and $t_{i+1} \leq \phi(t_i) + 2\tilde{\alpha}$, converges to $t(\tilde{\alpha})$. Noting that by (4.16) $|p^{k+1}|^2_{H_{k+1}} \leq |p^{k+1}|^2_{H_k}$ and putting $t_k = |p^k|^2_{H_k}$, we come to the following conclusion.

Given any $\varepsilon_p > 0$, there exists an $\tilde{\alpha} > 0$ and a number $N_4(\varepsilon_p) > 0$ such that if (5.23) is satisfied for $N_4(\varepsilon_p)$ consecutive iterations without resettings, then at one of those iterations $|p^k|^2_{H_k} \leq \varepsilon_p$.

(E) From (5.3) and the properties of M it follows that there exists an $\varepsilon > 0$ such that [13, Lemma 2.1]

(5.26) $$|g| > \varepsilon_0 \text{ for all } g \in \text{conv } \{\bigcup_{x \in B(\bar{x},\varepsilon)} M(x)\} = M(\bar{x}, \varepsilon) \quad .$$

(F) Since (5.6) and $f(x^k) \to f(\bar{x})$ imply that the resetting test (3.5) may be fulfilled only finitely often, we may suppose this test is inactive for sufficiently large k.

(G) Reasoning as in part (C), it is easy to prove that given an $\varepsilon_1 > 0$ and N_1, $N_2 > 0$, there exists an integer N_3 such that if a resetting occurs at some $k_a \in \{N_3, N_3 + 1, \ldots, N_3 + N_2\}$ then

(5.27) $$a^k < m_a \varepsilon_1 \text{ for } k = k_a, k_a + 1, \ldots, N_3 + N_2,$$

and that for this N_3 the relations (5.15) hold. Since $|v^k| > \varepsilon_1$ by (5.6), (5.7) shows that the resetting tests (3.6) and (3.9b) remain inactive for $k = k_a, k_a + 1, \ldots, N_3 + N_2$.

(H) Using the results of part (D), take $\varepsilon_p = \varepsilon_1$ and the corresponding $\tilde{\alpha} > 0$ and $N_4(\varepsilon_p) = N_4(\varepsilon_1)$. Take $\varepsilon > 0$ introduced in (E) and N_1 sufficiently large for the resetting test (3.5) to be inactive by (F). Take $N_2 = 5[N_4(\varepsilon_1) + \bar{M}_y]$. Decrease $\tilde{\alpha}$, if necessary, to satisfy the right hand side of (5.23). Apply the results of part (C) to find $N_3 \geq N_1$, such that (5.15) and (5.27) hold for the quantities introduced in this part of the proof.

Suppose first that at some $\bar{k} \in \{N_3 + \bar{M}_g, \ldots, N_3 + M_g + 3N_4(\varepsilon_1)\}$ there is a resetting. By the rules of Step 8 and (5.15), we have

(5.28) $\quad G^{k+1} \subset M(\bar{x},\varepsilon)$ and $p^{k+1} \in M(\bar{x},\varepsilon)$,

for $k = \bar{k}$. Now (4.26) and (5.15) imply that (5.28) holds for k satisfying $\bar{k} \leq k \leq N_3 + N_2$. Then the results of parts (F) and (G) imply that the only resetting for those k may occur through (3.9a), i.e.,

(5.29) $\quad |p^{k+1}| \leq \varepsilon_0$,

which is impossible by (5.26) and (5.28). Thus for $\bar{k} \leq k \leq N_3 + N_2$, i.e., for more than $N_4(\varepsilon_1)$ iterations, there is no resetting. Since (5.23) is satisfied, part (C) of the proof indicates that for some k satisfying $\bar{k} \leq k \leq N_3 + N_2$, $|p^k|^2_{H_k} \leq \varepsilon_1$. By (4.17) and (5.5) this implies $|p^k|^2 \leq \varepsilon_1/\beta^{2\bar{M}_{up}} = \varepsilon_0^2$ and hence (5.29) holds, again leading to contradiction with (5.26) and (5.28).

It remains to consider the case when there is no resetting for k satisfying $N_3 + \bar{M}_g \leq k \leq N_3 + \bar{M}_g + 2N_4(\varepsilon_1)$, i.e., for at least 3 $N_4(\epsilon_1)$ iterations. Reasoning as above, we show that $|p^k| \leq \varepsilon_0$ for some such k, which forces a resetting by (3.9a). This contradiction ends the proof.

Remark 5.2. Suppose that the set $\{x \in \mathbb{R}^N : f(x) \leq f(x^0), x \in S\}$ is bounded. Then $\{x^k\}$ has at least one accumulation point. Due to the line search rules, we also have $\{y_R^k\}$ bounded and $\{g(y_R^k)\}$ is bounded by the local boundedness of generalized gradients.

Remark 5.3. One may also consider a variant of our algorithm in which Mifflin's line search (Mifflin 1979) is used. This involves a re-definition of ϕ_k and M_k, viz. taking $\phi_k(x) = f(x) - f(x^k)$ and $M_k(x) = M(x)$ and demanding that $h(y_i^k) \leq 0$. Thus one obtains an implementable version of Mifflin's method (Mifflin 1979), for which our convergence results are expressed by Theorem 5.1.

6. NUMERICAL RESULTS

In this section we present numerical results obtained with a simplified version of the algorithm. The simplification consists in taking $\alpha(x,y) = 0$ instead of using the definition (2.3). Note that our convergence results remain valid for this modification.

Taking $\alpha(x,y) = 0$ greatly simplifies the direction finding subproblem. Let us introduce a transformation at the k^{th} iteration by

$$\tilde{g}^k(y_R^j) = B_k^* g(y_R^j) \text{ for } g(y_R^j) \in G^{k+1} \quad , \tag{6.1}$$

$$\tilde{p}^k = B_k^* p^k \quad . \tag{6.2}$$

By (4.14), one may implement this transformation efficiently, since

$$\tilde{g}^{k+1}(y_R^j) = R_\beta(\xi_{k+1})\tilde{g}^k(y_R^j) \quad . \tag{6.3}$$

Problem (4.1) reduces to the following

minimize

$$\Big| \sum_{g(y_R^j)\in G^{k+1}} \lambda_j \tilde{g}^k(y_R^j) + \lambda_p \tilde{p}^k \Big|^2 \tag{6.4}$$

subject to

$$\sum_{g(y_R^j)\in G^{k+1}} \lambda_j + \lambda_p = 1 \text{ and } \lambda_j \geqslant 0 \text{ for } g(y_R^j) \in G^{k+1},\ \lambda_p \geqslant 0 \quad .$$

This problem is efficiently solved by Wolfe's algorithm (Wolfe 1976). The relations (4.2) now become

$$(6.5a)\qquad v^{k+1} = -|\tilde{p}^{k+1}|^2 \quad ,$$

$$(6.5b)\qquad \tilde{p}^{k+1} = -B_k^{-1} d^{k+1} = \sum_{g(y_R^j)\in G^{k+1}} \lambda_j^{k+1} \tilde{g}^k(y_R^j) + \lambda_p^{k+1} \tilde{p}^k \quad .$$

Then the direction d^{k+1} after a variable metric update is computed from

$$(6.6)\qquad d^{k+1} = -B_k R_{\beta^2}(\xi_{k+1}) \tilde{p}^{k+1} \quad .$$

In our implementation we also compute $g_f(x^{k+1})$ whenever $t_i^k > 0$ and append it to the bundle G^{k+1} at Step 3. Accordingly (6.4) and (6.5) undergo an obvious modification.

We shall now discuss the choice of parameters. We take $m_i = 0.5$ and $m_R = 0.6$, $\bar{\kappa} = \sqrt{2} - 1$ and $m_c = 10^{-10}$. This choice of m_c would force very frequent resettings, hence we reset by (3.5) only when there are L_R iterations since the last resetting with $L_R \simeq N/2 \div 2N$, or when there is a need for variable metric reinitialization. Since the variable metric is implementable by storing $\{\xi_k\}$ and using (6.1) through (6.6), the number of updatings depends on the amount of avaiable storage. For small problems we take $\bar{M}_{up} = 2N$, for $N \geq 10$ we take smaller $\bar{M}_{up}$.

The choice of m_a is guided by a stopping criterion. If one wants to attain final accuracy expressed by

$$(6.7)\qquad a^{k+1} \leq \varepsilon_a \text{ and } v^{k+1} \geq -\varepsilon_d^2$$

where $\varepsilon_a, \varepsilon_d > 0$ are set up by the user, then $m_a = \varepsilon_a | \varepsilon_d^2$ is taken. $\bar{M}_g$ is taken equal to N for small problems.

As we do not compute p^k in our implementation, we use a resetting test $|\tilde{p}^k| \leq \varepsilon_0$ with $\varepsilon_0 = 10^{-10}$. On the other hand, our implementation of Wolfe's algorithm (Wolfe 1976) has tests which discover when the numerical errors make $\tilde{p}^{k+1}$ meaningless.

The algorithm goes to Step 8 in this case to reduce the bundle G^{k+1}. This strategy was found to be reliable in practice.

We choose the coefficient of space dilation β equal to 1/3 when $N \leq 10$, and $\beta = 0.1$ for $N > 10$.

The line search procedure that we use is a modfication of Mifflin's procedure from (Mifflin 1977). In our implementation the number of gradient evaluations is equal to about half of the number of function evaluations.

The value of the parameter s^0 influences the number of function evaluations on the first iteration. We usually take $s^0 = \frac{1}{2}$.

We developed a FORTRAN subroutine and tested it on the ODRA 1325 computer both in single and double precision (11 and 20 significant digits, respectively).

The algorithm has been tested on about 30 nonsmooth problems. Details of the results of computations will appear elsewhere. Due to lack of space, we shall present here results for 3 standard nonsmooth unconstrained problems from (Lemarechal 1978).

Since the stopping test based on (6.7) proved to be unreliable for $N \geq 10$, most of the algorithm's runs were terminated by exceeding an allowable number of iterations and/or function evaluations.

The first problem MAXQUAD (Lemarechal 1978, Test Problem 1) is quite easy. It has 10 variables, i.e., $N = 10$. Accordingly we set $\beta = 1/3$, $\bar{M}_{up} = 15$, $L_R = 10$ and $\bar{M}_g = 10$. After 20 iterations and 84 function evaluations the vaue of $f(x^{19}) = -.84.1397$.

The second problem EQUIL (Lemarechal 1978, Test Problem 3) has $N = 8$. We took $\beta = 1/3$, $\bar{M}_{up} = 12$, $L_R = 8$ and $\bar{M}_g = 8$. After 30 iterations and 95 f-evaluations we got $f(x^{29}) = .4239.10^{-2}$.

The third problem SHELL DUAL (Lemarechal 1978, Test Problem 2) appears to be more interesting. Since $N = 15$, we take $\beta = 0.1$, $\bar{M}_{up} = 22$, $L_R = 22$ and $\bar{M}_g = 15$. Below we present a table illustrating the progress of the algorithm. N_f denotes the number of function evaluations.

N_f	218	222	235	249	262	271	284	325	419
f	35.1	34.7	34.4	34.1	33.9	33.8	33.6	33.2	32.86

Although our experience with the algorithm is still limited, we discovered that it is quite robust with respect to numerical errors. There are very small differences in its performance when it is run first in single and then in double-precision. The results presented above were obtained in single-precision.

RERERENCES

Chaney, R., and A. Goldstein. 1978. An extension of the method of subgradients. Nonsmooth Optimization, edited by C. Lemarechal and R. Mifflin. Proceedings of a IIASA Workshop, Laxenburg, Austria, March 28-April 8, 1977. Oxford: Pergamon Press.

Clarke, F.H. 1975. Generalized gradients and applications. Transactions of the American Mathematical Society 205:247-262.

Clarke, F.H. 1976. A new approach to Lagrange multipliers. Mathematics of Operations Research 1:165-174.

Ioffe, A.D. 1979. Necessary and sufficient conditions for a local minimum. A reduction theorem and first order conditions. SIAM J. Control and Optimization 17:245-250.

Kiwiel, K.C. 1981. A Globally Convergent Quadratic Approximation Algorithm for Inequality Constrained Minimax Problems. CP-81-9, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Lemarechal, C. 1978. Nonsmooth Optimization and Descent Methods. RR-78-04, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Lemarechal, C. 1978. A set of nonsmooth optimization test problems. Nonsmooth Optimization, edited by C. Lemarechal and R. Mifflin. Proceedings of a IIASA Workshop, Laxenburg, Austria, March 28-April 8, 1977. Oxford: Pergamon Press.

Lemarechal, C. 1978. Nonlinear Programming and Nonsmooth Optimization: A Unification. Rapport de Recherche No. 332, Institut de Recherche d'Informatique et d'Automatique, Le Chesnay, France.

Mifflin R. 1977. An algorithm for constrained optimization with semismooth functions. Mathematics of Operations Research 2:959-972.

Mifflin, R. 1979. A Modification and Extension of Lemarechal's Algorithm for Nonsmooth Minimization. TR-80-1, Washington State University, Department of Mathematics, Washington, USA.

Pironneau, O., and E. Polak. 1972. On the rate of convergence of certain methods of centers. Mathematical Programming 2:230-257.

Shor, N.Z. 1979. Methods for Minimizing Nondifferentiable Functions and their Applications. Kiev: Naukova Dumka (in Russian).

Wierzbicki, A.P. 1978. Lagrangian Functions and Nondifferentiable Optimization. WP-78-63, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Wolfe, P. 1975. A method of conjugate subgradients for minimizing nondifferentiable functions. Pages 145-173, Nondifferentiable Optimization, Mathematical Programming Study 3, edited by M.L. Balinski and P. Wolfe. Amsterdam: North Holland.

Wolfe, P. 1976. Finding the nearest point in a polytope. Mathematical Programming 11:128-149.