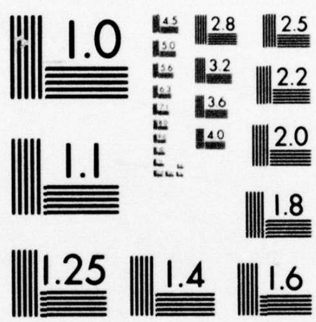
AD-A050 166 FLORIDA UNIV GAINESVILLE DEPT OF INDUSTRIAL AND SYS--ETC F/G 12/1
A SUBGRADIENT ALGORITHM FOR CERTAIN MINIMAX AND MINISUM PROBLEM--ETC(U)
JAN 77 J CHATELON, D HEARN, T J LOWE DAHC04-75-G-0150
UNCLASSIFIED RR-77-1 ARO-12640.22-M NL

| OF |

AD
A050166

END
DATE
FILMED
3-78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

A SUBGRADIENT ALGORITHM FOR CERTAIN MINIMAX AND MINISUM PROBLEMS

Research Report No. 77-1

by

Jacques Chatelon
Donald Hearn
Timothy J. Lowe

January, 1977

Department of Industrial and Systems Engineering
University of Florida
Gainesville, Florida 32611

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

This research was supported in part by the Army Research Office, Triangle Park, NC, under contract number DAHC04-75-G-0150, and by NSF grants ENG 74-22159 and ENG 75-10225.

THE FINDINGS OF THIS REPORT ARE NOT TO BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, UNLESS SO DESIGNATED BY OTHER AUTHORIZED DOCUMENTS.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 77-1 18 ARO 19 12640.22-M	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 9 Research rept.,
4. TITLE (and Subtitle) 6 A Subgradient Algorithm for Certain Minimax and Minisum Problems.		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER 14 RR-77-1
7. AUTHOR(s) 10 Jacques/Chatelon, Donald/Hearn, Timothy J./Lowe		8. CONTRACT OR GRANT NUMBER(s) 15 DAHC04-75-G-0150, NSF-ENG74-22159, ENG 74-22159, ENG 75-10225
9. PERFORMING ORGANIZATION NAME AND ADDRESS Industrial and Systems Engineering, University of Florida, Gainesville, Florida 32611		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 16 20061102A14D Rsch in & Appl of Applied Math.
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office, P.O. Box 12211, Triangle Park, NC 27709; NSF, Washington, DC 20550		12. REPORT DATE 11 Jan 1977
		13. NUMBER OF PAGES 39 12 40p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

N/A

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Subgradient methods
Nonlinear programming
Location theory
Linear approximation theory

an element of R to the n power

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper, we present a subgradient algorithm for the problem Min $\{F(x);\ x \in R^n\}$ where $F(x) = \text{Max}\ \{f_i(x);\ i = 1, 2, \ldots, m\}$ and where $f_i(x) = \sum_j f_{ij}(x)$. Each f_{ij} is assumed to be a proper convex function, and the number of different subgradient sets associated with nondifferentiable

sub i

sub ij

sub i

sum over j of

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

404 399

JEK

20.

points of f_{ij} is assumed to be finite on any bounded set. Problems belonging to this class include those where the f_{ij} are ℓ_p - norms ($1 \leq p \leq \infty$); for example, the linear approximation problem and both the minimax and minisum problems of location theory. The algorithm is an extension of the work of Dem'yanov and Malozemov [6].

sub ij sub P < or = infinity

We prove convergence of the algorithm to an ε - optimal solution is proved and demonstrate its effectiveness by solving a number of problems from location theory and linear approximation theory. Our computational results are compared with other solution methods.

Table of Contents

Abstract

In this paper, we present a subgradient algorithm for the problem Min $\{F(x);\ x \in R^n\}$ where $F(x) = \text{Max}\ \{f_i(x);\ i = 1, 2, \ldots, m\}$ and where $f_i(x) = \sum_j f_{ij}(x)$. Each f_{ij} is assumed to be a proper convex function, and the number of different subgradient sets associated with nondifferentiable points of f_{ij} is assumed to be finite on any bounded set. Problems belonging to this class include those where the f_{ij} are ℓ_p - norms $(1 \leq p \leq \infty)$; for example, the linear approximation problem and both the minimax and minisum problems of location theory. The algorithm is an extension of the work of Dem'yanov and Malozemov [6].

We prove convergence of the algorithm to an ε - optimal solution and demonstrate its effectiveness by solving a number of problems from location theory and linear approximation theory. Our computational results are compared with other solution methods.

1. Introduction

The minimax problem that we consider

$$\text{(P)} \qquad \min_{x \in R^n} F(x), \quad F(x) = \max\ \{f_i(x);\ i = 1, 2, \ldots, m\},$$

$$f_i = \sum_{j=1}^{\ell} f_{ij},\ f_{ij} \text{ convex not necessarily differentiable,}$$

is a special case of an unconstrained nondifferentiable convex programming problem. For $m = 1$, we minimize a sum of convex nondifferentiable functions and we call this problem a minisum problem.

To solve such nondifferentiable convex problems, it seems straightforward to extend the steepest descent method, using the fact that the steepest descent direction is given by the opposite of the element of minimum norm in the subgradient set. Unfortunately, it is well-known that this extended procedure is in general not convergent [6, 19]. As is indicated in [4], it is necessary to consider a larger set than the subgradient set itself in order to guarantee convergence.

The methods proposed in the literature to accomplish this can be divided into two main families.

The first family enlarges the subgradient set at any point and determines a descent direction from this enlarged set. For example, Dem'yanov and Malozemov [6] solve minimax problems with continuously differentiable functions by considering also 'near binding' functions in the following way: to the extreme points of the subgradient set (formed by the gradients of the binding functions) they add the gradients of those functions which are almost binding. In another study, Bertsekas and Mitter [2] used the ε-subgradient set to calculate a descent direction. All of these methods require the knowledge of the complete subgradient set.

By contrast, the second family requires only partial local information, in general only one element of the subgradient set. At each step, a subgradient is added to a 'bundle' of previous subgradients and a descent direction is obtained from this bundle. At some steps, the bundle is reinitialized so that it always consists of a limited number of subgradients. This idea has been proposed initially by Lemarechal [14] and Wolfe [19]. A nice feature of these methods is that they are conjugate gradient methods [15] when applied to smooth problems. However, it is well-known that conjugate gradient methods require very accurate line searches [11, 15], thus considerable time has to be spent on the line search part of these algorithms. Further extensions of these methods have been developed by Feuer [9] and recently by Mifflin [16]. In these two works, the idea of bundling is intimately related to the generalized gradient of Clarke [3].

Our approach is more related to the first class of methods in that we require the same type of local information. However, we introduce the concept of considering subgradients at neighboring points of the current iterate. Thus in some sense, we anticipate nondifferentiability [4, p. 43].

As an overview of what follows, in Section 2 we introduce notation and definitions. In Section 3, we discuss the outcome of the direction finding optimization problem and show that we either find a direction of descent, or we have attained a near optimal point. Then, after presenting the algorithm in Section 4, we prove in Section 5 that it is convergent. In this section, we combine the methods used in [4] and [6] but is is also necessary to rely on the properties of the functions involved. Section 6 contains details for a modified version of the algorithm, while Section 7 gives some computational results. The following notations and basic definitions will be used in the remaining sections.

We denote n-dimensional Euclidean space as R^n and $||x||_p$ as the ℓ_p-norm of $x \in R^n$, where $||x||$ is the ℓ_2-norm. Given a point $x \in R^n$, we denote the Euclidean ball about x of radius η as $N(x, \eta)$. In the case where $x = 0$ and $\eta = 1$, $B \equiv N(0, 1)$; that is, B is the Euclidean unit ball.

Given a function F, defined on R^n, the subgradient set of F at x is denoted as $\partial F(x)$, and the directional derivative of F at x in the direction y is $F'(x, y)$.

Given a subset $S \subset R^n$, we let Conv(S) be the convex hull of S and Nr(S) be the element of minimum Euclidean norm in S when S is closed and convex. Given any function F defined on S, we let $\partial F(S) \equiv \cup \{\partial F(x); x \in S\}$.

2. Definitions and Notations

We shall assume that the functions f_{ij} involved in the definition of

$$F(x) = \text{Max } \{f_i(x) = \sum_{j=1}^{\ell} f_{ij}(x);\ i = 1, 2, \ldots, m\}$$

are convex and finite on R^n. For notational convenience, we write each f_i as the sum of exactly ℓ functions f_{ij}, assuming, if necessary, that some of the f_{ij} are identically zero on R^n. Then it is clear that each f_i is a continuous convex function and thus, F is a finite convex, continuous function on R^n.

Suppose $\varepsilon \geq 0$ is given. At each point x, we consider the set of indices

$$R(x, \varepsilon) \equiv \{i \in \{1, 2, \ldots, m\};\ f_i(x) \geq F(x) - \varepsilon\}. \tag{2.1}$$

In particular

$$R(x, 0) = \{i \in \{1, 2, \ldots, m\};\ f_i(x) = F(x)\}, \text{ and} \tag{2.2}$$

$$R(x, \varepsilon_1) \subseteq R(x, \varepsilon_2) \quad \text{for} \quad 0 \leq \varepsilon_1 \leq \varepsilon_2. \tag{2.3}$$

We shall make use of the following property [17]:

$$\partial F(x) = \text{Conv } (\cup \{\partial f_i(x);\ i \in R(x, 0)\}). \tag{2.4}$$

We restrict our attention to the functions f_{ij} which belong to the class of functions defined below.

<u>Definition 2.1</u> A finite convex function is <u>LFS</u> if, in any closed bounded Euclidean ball, the number of different subgradient sets, corresponding to the points of nondifferentiability of this function, is finite. (LFS is an abbreviation for the phrase 'locally finitely subdifferentiable').

<u>Assumption A.1</u> The functions f_{ij} are LFS.

The concept of LFS functions originated from the study of minimization problems in location theory, where functions involving norms are frequently encountered, such as $||x - a||_p$, $1 \leq p \leq \infty$. The function $||x||_1$ (for $x \in R^2$) is not differentiable along either axis, but the total number of different subgradient sets associated with all points of nondifferentiability is five.

As an example, we note that piecewise linear functions are also LFS.

The representation of each function f_i as a sum of LFS functions may be useful, because the sum of LFS functions is not always LFS. As an example, consider $x_1^2 + x_2^2 + ||x||_1$, $x = (x_1, x_2) \in R^2$, which viewed as a single function is not LFS. But $f_1 \equiv x_1^2 + x_2^2$ and $f_2 \equiv ||x||_1$ are each LFS.

In the next definition, we introduce the points of nondifferentiability of a function f_{ij}.

<u>Definition 2.2</u>

$$G_{ij}(x, \eta) \equiv \{x\} \cup \{y;\ y \in N(x, \eta),\ f_{ij} \text{ not differentiable at } y\}.$$

We then define the set S, which is our enlarged subgradient set, obtained by considering also neighboring points of nondifferentiability.

<u>Definition 2.3</u>

$$S_i(x, \eta) \equiv \sum_{j=1}^{\ell} \partial f_{ij}\ (G_{ij}\ (x, \eta)),\ i = 1, 2, \ldots, m, \tag{2.5}$$

$$S(x, \varepsilon, \eta) \equiv \text{Conv}\ (\cup \{S_i\ (x, \eta);\ i \in R\ (x, \varepsilon)\}). \tag{2.6}$$

The following example illustrates Definition 2.3.

<u>Example 1</u>. Consider the problem

$$\min_{x \in R^2}\ \text{Max}\ \{f_1(x), f_2(x)\},\ f_1(x) = ||x||_1 = |x_1| + |x_2|,$$

$$f_2(x) = ||x - (1, -1)\ ||_1 = |x_1 - 1| + |x_2 + 1|.$$

Let $\ell = 1$, $\varepsilon = .2$ and $\eta = 1$. At $x_0 = (.9, 1)$ we have

$$f_1(x_0) = 1.9,\ f_2(x_0) = 2.1 = F(x_0),\ R(x_0, \varepsilon) = \{1, 2\},$$

$$\partial f_1(x_0) = \{(1,1)\},\quad \partial f_2(x_0) = \{(-1, 1)\}.$$

Furthermore, in $N(x_0, \eta)$, there are points of nondifferentiability

(a) for f_1, along the axis $x_1 = 0$ and at $(.9, 0)$, and

(b) for f_2, along the line $x_1 = 1$.

These supplementary points yield two more different sets to include in S,

$$\text{Conv}\ (\{(1, -1),\quad (1, 1)\}),\quad \text{Conv}\ (\{(-1, 1),\quad (1, 1)\}).$$

Hence, $S(x_0, \varepsilon, \eta)$ is the triangle

$$\mathrm{Conv}\ (\{(-1, 1),\ (1, 1),\ (1, -1)\}).$$

On the other hand, suppose we choose $\ell = 2$ with $f_{11}(x) = |x_1|$, $f_{12}(x) = |x_2|$, $f_{21}(x) = |x_1 - 1|$, $f_{22}(x) = |x_2 + 1|$. In this case, it is easy to check that

$$S_1(x_0, \eta) = \mathrm{Conv}\ (\{(1, 1), (-1, 1), (-1, -1), (1, -1)\}),$$

$$S_2(x_0, \eta) = \mathrm{Conv}\ (\{(1, 1), (-1, 1)\}).$$

Then $S(x_0, \varepsilon, \eta)$ is the square

$$S(x_0, \varepsilon, \eta) = \mathrm{Conv}\ (\{(1, 1), (-1, 1), (-1, -1), (1, -1)\}).$$

Remark. From Definitions 2.2 and 2.3, it is clear that $\partial f_{ij}(x) \subseteq \partial f_{ij}(G_{ij}(x, \eta))$, $j = 1, 2, \ldots, \ell$, $i = 1, 2, \ldots, m$,

and,

$$\partial f_i(x) = \sum_{j=1}^{\ell} \partial f_{ij}(x) \subseteq \sum_{j=1}^{\ell} \partial f_{ij}(G_{ij}(x, \eta)) = S_i(x, \eta),$$

$$i = 1, 2, \ldots, m. \tag{2.7}$$

Since, by (2.3),

$$R(x, 0) \subseteq R(x, \varepsilon),$$

we get from (2.7)

$$\cup\ \{\partial f_i(x);\ i \in R(x, 0)\} \subseteq \cup\ \{S_i(x, \eta);\ i \in R(x, 0)\}$$

$$\subseteq \cup\ \{S_i(x, \eta);\ i \in R(x, \varepsilon)\},$$

and consequently,

$$\mathrm{Conv}\ (\ \cup\ \{\partial f_i(x);\ i \in R(x, 0)\}) \subseteq \mathrm{Conv}\ (\ \cup\ \{S_i(x, \eta);\ i \in R(x, \varepsilon)\}). \tag{2.8}$$

Then (2.4) and (2.8) imply

$$\partial F(x) \subseteq S(x, \varepsilon, \eta). \tag{2.9}$$

From Definition 2.3, it is easy to establish the following property.

Property 2.1. $S(x, \varepsilon, \eta)$ is a nonempty compact convex subset of R^n.

Proof. From Assumption A.1, each set

$$\partial f_{ij}(G_{ij}(x, \eta)), \quad j = 1, 2, \ldots, \ell, \quad i = 1, 2, \ldots, m,$$

is a finite union of nonempty compact sets, so that $\partial f_{ij}(G_{ij}(x, \eta))$ is itself a compact set. Hence, the finite sum

$$\sum_{j=1}^{\ell} \partial f_{ij}(G_{ij}(x, \eta)) = S_i(x, \eta), \quad i = 1, 2, \ldots, m,$$

is also a compact set and the set

$$\cup \{S_i(x, \eta);\ i \in R(x, \varepsilon)\}$$

is a finite union of nonempty compact sets. Thus, this set is also a nonempty compact set and Property 2.1 is a direct consequence of Theorem 17.2 of [18]. ||

$S(x, \varepsilon, \eta)$, which contains the subgradient set but also pertinent local information, now replaces this subgradient set for the determination of a descent direction and the corresponding directional derivative, provided that they exist. These will be given by the next definition.

Definition 2.4. At any point x, let

$$\psi(x, \varepsilon, \eta) \equiv \mathrm{Min}\ \{\mathrm{Max}\ \{(g, d);\ d \in S(x, \varepsilon, \eta)\ ;\ ||g|| \leq 1\}. \qquad (2.10)$$

By compactness of $S(x, \varepsilon, \eta)$ and of the Euclidean unit ball B, this quantity is always well-defined. Using a minimax theorem (Corollary 37.3.2,[18]) since the two sets $S(x, \varepsilon, \eta)$ and B are both non-empty compact convex sets, we obtain

Property 2.2

$$\psi(x, \varepsilon, \eta) = -||Nr(S(x, \varepsilon, \eta))||. \qquad (2.11)$$

Proof. $\psi(x, \varepsilon, \eta) = \mathrm{Min}\ \{\mathrm{Max}\ \{(g, d);\ d \in S(x, \varepsilon, \eta)\};\ g \in B\}$

$$= \mathrm{Max}\ \{\mathrm{Min}\ \{(g, d);\ g \in B\};\ d \in S(x, \varepsilon, \eta)\}$$

$$= \mathrm{Max}\ \{(d, -d/||d||);\ d \in S(x, \varepsilon, \eta)\}$$

$$= -\ \mathrm{Min}\ \{||d||;\ d \in S(x, \varepsilon, \eta)\}$$

$$= -||Nr(S(x, \varepsilon, \eta))||.$$ ||

As a consequence of Property 2.2, we have $\psi = 0$ or $\psi < 0$. These are two cases that we consider in the next section.

3. Study of the Alternative $\psi(x, \varepsilon, \eta) = 0$ or $\psi(x, \varepsilon, \eta) < 0$

As shown in Section 2, $\psi(x, \varepsilon, \eta)$ is similar to a directional derivative. Hence it is reasonable to expect that its value will give some clue about the optimality of the point x. Before considering the two cases $\psi(x, \varepsilon, \eta) = 0$ and $\psi(x, \varepsilon, \eta) < 0$, we prove two lemmas which will be useful in the sequel.

Lemma 3.1 Let X be a non-empty compact subset of R^n. Then the subgradients of the functions f_{ij} are uniformly bounded on the set $X' = \{y;\ y \in N(x, \eta),\ x \in X\}$. That is, $\exists M,\ 0 < M < \infty$ such that

given any $y \in X'$ and any $s \in \partial f_{ij}(y)$,

$||s|| \leq M, \quad \forall\ i = 1, 2, \ldots, m, \quad \forall j = 1, 2, \ldots, \ell.$

Proof. Since X is compact, X' is also compact. Then, by Theorem 24.7 of [18], $\partial f_{ij}(X')$ is a non-empty compact subset of R^n and consequently, the number

$$M_{ij} = \text{Sup } \{||s||;\ s \in \partial f_{ij}(X')\}$$

is finite. Then

$$M = \text{Max } \{M_{ij};\ i = 1, 2, \ldots, m;\ \ j = 1, 2, \ldots, \ell\}$$

satisfies the requirements of the Lemma. ||

In Lemma 3.2, we shall consider the following piecewise linear function:

$$V(z) = \text{Max } \{(w_k, z) - v_k;\ k = 1, 2, \ldots q\},$$

in which the w_k are q given vectors of R^n and the v_k are q real numbers.

Lemma 3.2 A necessary and sufficient condition that the infimum v^* of V(z) be attained is that there exists a convex combination of the w_k equal to the null vector, that is, $\exists \lambda_k$, $k = 1, 2, \ldots, q$, $\lambda_k \geq 0$ such that

$$\sum_k \lambda_k = 1 \text{ and } \sum_k \lambda_k w_k = 0.$$

Furthermore, $v^* \geq - \sum_k \lambda_k v_k$ and if the λ_k are unique,

$$v^* = - \sum_k \lambda_k v_k.$$

Proof. The infimum of $V(z)$ is attained if and only if the linear program

$$\begin{aligned} &\text{Min} \quad v \\ &\text{s.t.} \quad (w_k, z) - v_k - v \leq 0 \qquad k = 1, 2, \ldots, q \end{aligned}$$

has an optimal solution. Its dual may be written

$$\begin{aligned} &\text{Min} \ \sum_k v_k u_k \\ &\text{s.t.} \ \sum_k u_k w_k = 0 \\ &\qquad \sum_k u_k = 1 \\ &\qquad u_k \geq 0, \quad k = 1, 2, \ldots, q. \end{aligned}$$

By duality theory, if the dual is not feasible, $v^* = -\infty$. Otherwise, v^* is finite and for any dual feasible solution λ_k, $k = 1, 2, \ldots, q$,

$$v^* \geq - \sum_k \lambda_k v_k \, .$$

If there is only one feasible solution, then

$$v^* = - \sum_k \lambda_k v_k. \qquad ||$$

We now consider $\psi(x, \varepsilon, \eta) = 0$. We call such a point x, at which $\psi(x, \varepsilon, \eta) = 0$, a stationary point. Let $F^* = F(x^*)$ be the function value at an optimal point x^*, provided such a point exists. From the two previous lemmas, we obtain the following theorem which provides a bound on the difference between F^* and the function value at a stationary point.

Theorem 3.3 For any stationary point x of a compact set X, we have

$$F(x) - \varepsilon - 2\ell\eta M \leq F^* \leq F(x). \tag{3.1}$$

Proof. Stationarity at x is equivalent to $0 \in S(x, \varepsilon, \eta)$.

With $S(x, \varepsilon, \eta)$ as the convex hull of $\cup \{S_i(x, \eta);\ i \in R(x, \varepsilon)\}$, $0 \in S(x, \varepsilon, \eta)$ if and only if there exists λ_k, w_k such that

$$\lambda_k \geq 0 \qquad , \quad k = 1, \ldots, q \leq n + 1, \tag{3.2}$$

$$w_k \in \cup \{S_i(x, \eta);\ i \in R(x, \varepsilon)\} \tag{3.3}$$

where,

$$\sum_{k=1}^{q} \lambda_k = 1 \tag{3.4}$$

and

$$\sum_{k=1}^{q} \lambda_k w_k = 0. \tag{3.5}$$

For arbitrary k, $1 \leq k \leq q$, we have that $w_k \in S_h(x, \eta)$ for some $h \in R(x, \varepsilon)$. But then from the definition of $S_h(x, \eta)$, we have that $w_k = \sum_{r=1}^{\ell} w_{kr}$ where

$$w_{kr} \in \partial f_{hr}(G_{hr}(x, \eta)). \tag{3.6}$$

Let $y_{kr} \in G_{hr}(x, \eta)$, where $w_{kr} \in \partial f_{hr}(y_{kr})$ for $r = 1, 2, \ldots, \ell$.

From the subgradient inequality, $\forall r = 1, 2, \ldots, \ell$,

$$f_{hr}(z) \geq f_{hr}(y_{kr}) + (w_{kr}, z - y_{kr}), \ \forall z \in R^n. \tag{3.7}$$

Furthermore for arbitrary $s_{hr} \in \partial f_{hr}(x)$,

$$f_{hr}(y_{kr}) \geq f_{hr}(x) + (s_{hr}, y_{kr} - x). \tag{3.8}$$

Adding (3.7) and (3.8) yields

$$f_{hr}(z) \geq f_{hr}(x) + (w_{kr}, z - y_{kr}) + (s_{hr}, y_{kr} - x). \tag{3.9}$$

With $x \in X$, a compact set, from Lemma 3.1, there exists M such that $M \geq ||s_{hr}||$ and thus,

$$(s_{hr}, y_{kr} - x) \geq -\eta M. \tag{3.10}$$

Using (3.10) and summing (3.9) over all $r = 1, \ldots, \ell$,

$$\sum_{r=1}^{\ell} f_{hr}(z) \geq \sum_{r=1}^{\ell} f_{hr}(x) - \ell\eta M + \sum_{r=1}^{\ell} (w_{kr}, z - y_{kr}),$$

or

$$f_h(z) \geq f_h(x) - \ell\eta M + \sum_{r=1}^{\ell} (w_{kr}, z - y_{kr}). \qquad (3.11)$$

Now, since $h \in R(x, \varepsilon)$, $f_h(x) \geq F(x) - \varepsilon$ and then from (3.11)

$$F(z) \geq f_h(z) \geq F(x) - \varepsilon - \ell\eta M + \sum_{r=1}^{\ell} (w_{kr}, z - y_{kr}),$$

or

$$F(z) \geq F(x) - \varepsilon - \ell\eta M + \sum_{r=1}^{\ell} (w_{kr}, z - y_{kr}). \qquad (3.12)$$

Since k is arbitrary, we obtain

$$F(z) \geq F(x) - \varepsilon - \ell\eta M + \text{Max}\,\{\sum_{r=1}^{\ell} (w_{kr}, z - y_{kr});\ k = 1, \ldots, q\}. \qquad (3.13)$$

Letting $v_k = \sum_{r=1}^{\ell} (w_{kr}, y_{kr})$ in (3.13), noting (3.2), (3.4) and (3.5) and taking the minimum over $z \in R^n$ on both sides of (3.13) we have by Lemma 3.2,

$$F^* \equiv \operatorname*{Min}_z F(z) \geq F(x) - \varepsilon - \ell\eta M - \sum_{k=1}^{q} \lambda_k v_k. \qquad (3.14)$$

Decomposing $v_k = \sum_{r=1}^{\ell} (w_{kr}, y_{kr})$ into

$$\sum_{r=1}^{\ell} (w_{kr}, y_{kr} - x) + \sum_{r=1}^{\ell} (w_{kr}, x)$$

and using the inequality

$$\eta M \geq (w_{kr}, y_{kr} - x), \ r = 1, \ldots, \ell,$$

we get

$$-v_k \geq -\ell\eta M - \sum_{r=1}^{\ell} (w_{kr}, x) = -\ell\eta M - (w_k, x).$$

But then,

$$-\sum_{k=1}^{q} \lambda_k v_k \geq -\sum_{k=1}^{q} \lambda_k \ell\eta M - \sum_{k=1}^{q} (\lambda_k w_k, x). \tag{3.15}$$

Using (3.4), (3.5) in (3.15) we get

$$-\sum_{k=1}^{q} \lambda_k v_k \geq -\ell\eta M. \tag{3.16}$$

Therefore (3.14) and (3.16) give

$$F^* \geq F(x) - \varepsilon - 2\ell\eta M. \tag{3.17}$$

||

We now present an example which illustrates Theorem 3.3.

Example 2. Consider the very simple problem

$$\operatorname{Min}_{x \in R^2} F(x), \quad F(x) = ||x||_1 = |x_1| + |x_2|.$$

With $\varepsilon = 0$ and $\eta = 1$, we look at two different formulations of this problem.

(a) $F(x) = ||x||_1$ with $m = 1$, $\ell = 1$ and clearly $M = \sqrt{2}$. The point $x_0 = (1, 1)$ is stationary since $S(x_0, \varepsilon, \eta)$ is the convex hull of the three points

$$(-1, 1), (1, 1), (1, -1).$$

At x_0, we have $F(x_0) = 2$ and since $F^* = 0$, we verify

$$-0.82 = 2 - 2 \times \sqrt{2} \leq 0 \leq 2.$$

(b) $F(x) = |x_1| + |x_2|$ with $m = 1$, $\ell = 2$ and $M = 1$. x_0 is also stationary because $S(x_0, \varepsilon, \eta)$ is the square

$$\operatorname{Conv}(\{(1, 1), (-1, 1), (-1, -1), (1, -1)\}).$$

Then, Theorem 3.3 yields

$$-2 = 2 - (2)(2) \leq 0 \leq 2.$$

This example illustrates a general result. When each function f_i is itself LFS, choosing $\ell = 1$ will give better bounds in the inequalities

$$F(x) - \varepsilon - 2\ell\eta M \leq F^* \leq F(x).$$

To see this result, write f_i as $\sum_{r=1}^{\ell} f_{ir}$ and decompose any subgradient s_i of f_i at x into the sum

$$s_i = \sum_{r=1}^{\ell} s_{ir}, \; s_{ir} \in \partial f_{ir}(x), \; r = 1, \ldots, \ell. \tag{3.18}$$

If we call M the bound on the norm of any subgradient of any function f_{ir} and M_1 the bound for the subgradients of the functions f_i, $i = 1, \ldots, m$, from (3.18) we obtain

$$||s_i|| = ||\sum_{r=1}^{\ell} s_{ir}|| \leq \ell M$$

and consequently

$$M_1 \leq \ell M. \tag{3.19}$$

Writing now the inequalities given by Theorem 3.3 in both cases, $\ell > 1$ and $\ell = 1$, we have

$$F(x) - \varepsilon - 2\ell\eta M \leq F^* \leq F(x), \tag{3.20}$$

$$F(x) - \varepsilon - 2\eta M_1 \leq F^* \leq F(x). \tag{3.21}$$

But inequality (3.19) indicates that

$$F(x) - \varepsilon - 2\ell\eta M \leq F(x) - \varepsilon - 2\eta M_1 \leq F^*$$

so that in general (3.21) gives better bounds.

On the other hand, considering each f_i as a sum of $\ell > 1$ LFS functions f_{ij}, could ease the task of determining the sets $S_i(x, \eta)$. Hence, there is a trade-off between accuracy and implementation.

The set of stationary points does not seem to have any obvious property, such as convexity. With the problem of Example 1, and choosing $\ell = 1$

$$\underset{(x_1, x_2)}{\text{Min}} \text{ Max } \{|x_1| + |x_2|, \; |x_1 - 1| + |x_2 + 1|\},$$

the set of optimal points is the line-segment joining the two points (0, -1) and (1, 0). But, with $\varepsilon = .2$ and $\eta = 1$, the set of stationary points is as sketched in Figure 1.

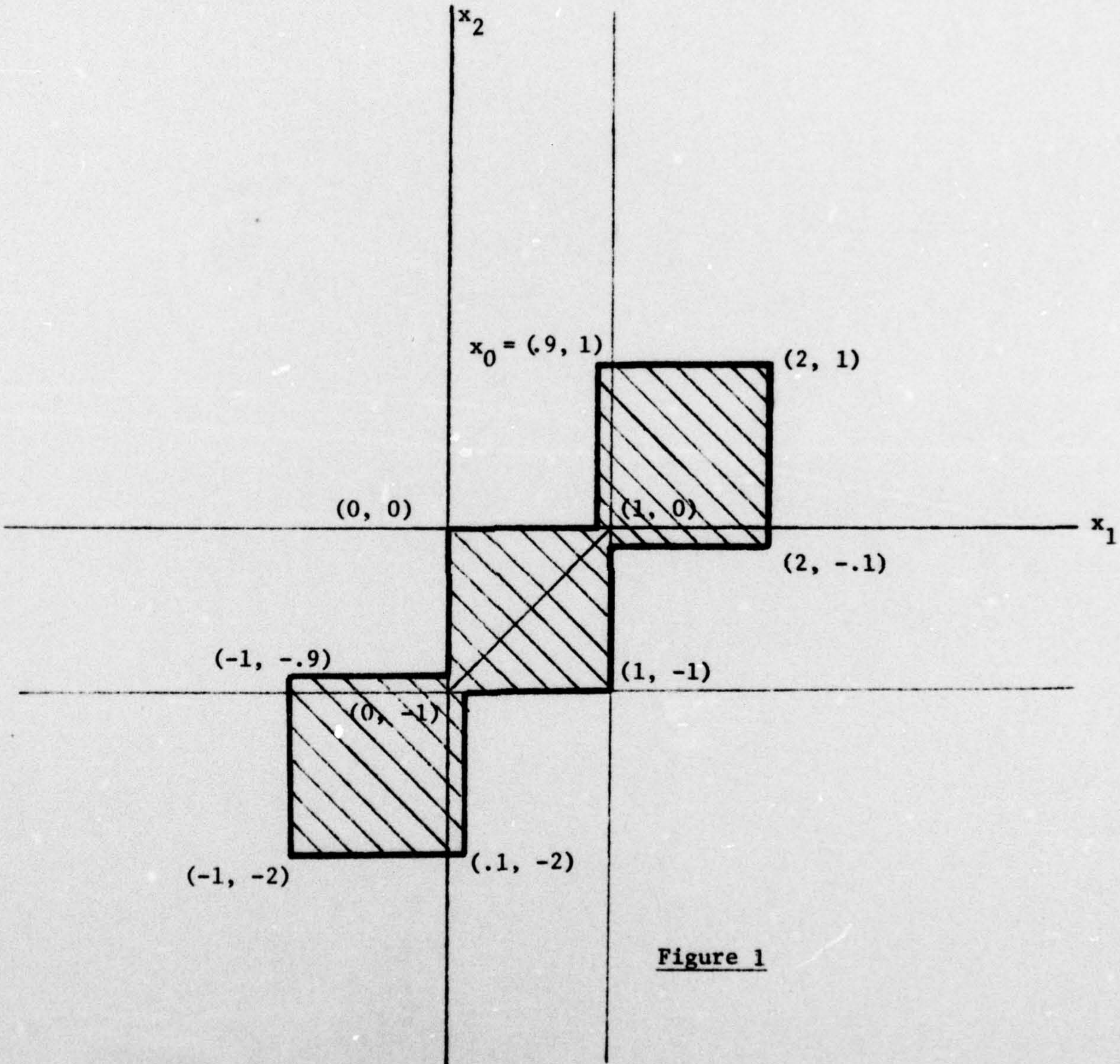


Figure 1

We now consider the remaining case, $\psi(x, \varepsilon, \eta) < 0$.

Since x is not stationary, we are able to find a descent direction.

Theorem 3.4 If $\psi(x, \varepsilon, \eta) < 0$, then there exists a nontrivial descent direction for F at x.

Proof. $\psi(x, \varepsilon, \eta) < 0$ implies the existence of a direction g_0 for which

$$
\begin{aligned}
0 > \psi(x, \varepsilon, \eta) &= -||Nr(S(x, \varepsilon, \eta))|| \\
&= \text{Max } \{(d, g_0);\ d \in S(x, \varepsilon, \eta)\} \qquad (3.22) \\
&= -||d_0|| \\
&= (d_0, g_0).
\end{aligned}
$$

Thus, the property

$$F'(x, g_0) = \text{Max } \{(d, g_0);\ d \in \partial F(x)\}$$

and (2.9), (3.22) imply

$$F'(x, g_0) < 0. \qquad ||$$

4. Description of the Algorithm

Using the results of Section 3, particularly Theorems 3.3 and 3.4, it is clear how to construct a descent method for problem (P).

Choose $\varepsilon > 0$, $\eta > 0$ and x_0 a starting point. Set $k = 0$ and go to Step 1.

Step 1

At x_k, find $F(x_k)$ and $R(x_k, \varepsilon)$. Calculate $S(x_k, \varepsilon, \eta)$ and $\psi(x_k, \varepsilon, \eta)$. Go to Step 2.

Step 2

If $\psi(x_k, \varepsilon, \eta) = 0$, stop: x_k is a stationary point. If $\psi(x_k, \varepsilon, \eta) < 0$, there exists g_k for which

$$\psi(x_k, \varepsilon, \eta) = \mathrm{Max}\ \{(d, g_k);\ d \in S(x_k, \varepsilon, \eta)\}.$$

Line search: find t_k such that

$$F(x_k + t_k g_k) = \min_{t \geq 0} F(x_k + t g_k).$$

Set $x_{k+1} = x_k + t_k g_k$, $k = k + 1$ and return to Step 1.

The next section establishes that this algorithm converges to a stationary point for the class of problems studied in this paper.

5. Proof of Convergence

For the proof of convergence, we assume there exists some x_0 such that the level set $X = \{x; F(x) \leq F(x_0)\}$ is bounded. Since F is continuous, we deduce that the minimum value, F^*, of F is attained at some $x^* \in X$.

Now suppose that the starting point is x_0. The algorithm generates a sequence of points $\{x_k\} \subseteq X$, since $F(x_{k+1}) < F(x_k) < F(x_0)$ by Theorem 3.4. If the algorithm terminates at some iteration k, by Theorem 3.3, we have an estimate of F^*. Otherwise, the infinite sequence $\{x_k\}$ must have a limit point $x_* \in X$. Let $K \subseteq \{0, 1, 2, \ldots\}$ be the set of indices such that $x_k \to x_*$ for $k \to \infty$ and $k \in K$.

We now show (by contradiction) that x_* is a stationary point, that is, $\psi(x_*, \varepsilon, \eta) = 0$.

Assume that $\psi(x_*, \varepsilon, \eta) = -b < 0$. First, we show that for k large enough, $\psi(x_k, \varepsilon, \eta)$ is uniformly negative, i.e., $\psi(x_k, \varepsilon, \eta) \leq c < 0$. This follows because in a neighborhood of x_*, the sets $S(x_k, \varepsilon, \eta)$ approximate $S(x_*, \varepsilon, \eta)$ [4, p. 43].

Lemma 5.1 There exists a number N_1 such that $k \in K$ and $k > N_1$ implies $R(x_k, \varepsilon) \subseteq R(x_*, \varepsilon)$.

Proof. By Lemma 7.1 of [6, p. 92], $\exists \delta > 0$ such that $||x_* - x|| < \delta$ implies $R(x, \varepsilon) \subseteq R(x_*, \varepsilon)$. Since $x_k \to x_*$ for $k \in K$ and $k \to \infty$, we can find N_1 for which $k > N_1$ implies $||x_k - x_*|| < \delta$ and consequently $R(x_k, \varepsilon) \subseteq R(x_*, \varepsilon)$. ||

Lemma 5.2 Let $g: R^n \to R$ be an LFS function and suppose $\{x_k; k \in K\} \to x_*$ is a convergent sequence. Then, for any $\gamma > 0$, there exists N such that

$$\partial g(G(x_k, \eta)) \subseteq \partial g(G(x_*, \eta)) + \gamma B, \ \forall k \geq N, \ k \in K. \tag{5.1}$$

Proof. For any $x \in R^n$ we define $G'(x, \eta) \equiv G(x, \eta) / x$ so that $G(x, \eta) = G'(x, \eta) \cup \{x\}$.

From Corollary 24.5.1 of [18] there exists N' such that $\forall k \geq N'$, $k \in K$ we have

$$\partial g(x_k) \subset \partial g(x_*) + \gamma B. \qquad (5.2)$$

We now consider $G'(x_k, \eta)$. Define the set

$$H(s) \equiv \cup \{\partial g(G'(x_k, \eta)); k \geq s, k \in K\}, s = 1, 2, \ldots. \qquad (5.3)$$

From this definition, it is clear that

$$H(s) = \partial g(\cup\{G'(x_k, \eta); k \geq s, k \in K\}) \qquad (5.4)$$

and

$$H(s_1) \subseteq H(s_2) \text{ for } s_2 < s_1. \qquad (5.5)$$

We suppose $H(p) \neq \phi$, $\forall p \geq N'$, $p \in K$ for otherwise by choosing N large enough, $G(x_k, \eta) = \{x_k\}$, $\forall k > N'$ and thus from (5.2) we are done. Since the sequence $\{x_k\}$ is convergent, the set

$$\cup\{G'(x_k, \eta); k \geq p, k \in K\}$$

is contained in a closed bounded Euclidean ball and thus with g an LFS function, H(p) is the union of a finite number of distinct subgradient sets for all p, $p = 1, 2, \ldots$.

The sets of the sequence H(p) are nonempty for all p. Therefore, there must be a finite number of subgradient sets, say, S^r, $r = 1, 2, \ldots, q$, each of which occurs infinitely often in this sequence. In other words, there exists some N" such that

$$H(p) \subseteq \cup\{S^r; r = 1, \ldots, q\}, \forall p \geq N".$$

Consequently, we can find q index sets K_r, $r = 1, \ldots, q$, where each K_r is an infinite subset of K and $S^r = \partial g(y_k)$ for some $y_k \in G'(x_k, \eta)$, $\forall k \in K_r$.

We now show that $H(p) \subseteq \partial g(G'(x_*, \eta))$. For fixed r, since the sequence $\{x_k; k \in K\}$ is convergent, the sequence $\{y_k; k \in K_r\}$ is in a compact set and thus has a limit point $\bar{y}_r$. Hence there exists a subsequence $K'_r \subseteq K_r$ such that $y_k \to \bar{y}_r$ for $k \to \infty$, $k \in K'_r$.

Since $x_k \to x_*$, it easily follows that

$$\bar{y}_r \in N(x_*, \eta). \tag{5.6}$$

Furthermore, by Corollary 24.5.1 of [18] given $\varepsilon' > 0$, $\exists N_r \geq N''$ such that

$$\partial g(y_k) \subseteq \partial g(\bar{y}_r) + \varepsilon' B, \ \forall k \in K'_r, \ k \geq N_r. \tag{5.7}$$

But $S^r = \partial g(y_k)$ $\forall k \in K'_r$ and thus (5.7) clearly implies that

$$S^r \subseteq \partial g(\bar{y}_r). \tag{5.8}$$

Since S^r is not a singleton, from (5.8) we have that g is not differentiable at $\bar{y}_r$, and thus using (5.6)

$$\bar{y}_r \in G'(x_*, \eta). \tag{5.9}$$

From (5.8) and (5.9) it follows that $S^r \subseteq \partial g(G'(x_*, \eta))$ and hence, for $p \geq N''' \equiv \mathrm{Max}\ \{N_r;\ r = 1, \ldots, q\}$,

$$H(p) \subseteq \cup\{S^r;\ r = 1, \ldots, q\} \subseteq \partial g(G'(x_*, \eta)). \tag{5.10}$$

Now, since $\partial g(G'(x_k, \eta)) \subseteq H(p)$ $\forall k \in K$, $k \geq p$, defining $N \equiv \mathrm{Max}\ \{N', N'''\}$, we have from (5.2) and (5.10)

$$\partial g(G'(x_k, \eta)) \cup \partial g(x_k) \subseteq \partial g(G(x_*, \eta)) + \gamma B, \quad k \in K, \ k \geq N,$$

or

$$\partial g(G(x_k, \eta)) \subseteq \partial g(G(x_*, \eta)) + \gamma B, \ \forall k \in K, \ k \geq N. \qquad ||$$

Lemma 5.3 For any $\varepsilon' > 0$, there exists a number N_2 such that $k \in K$ and $k > N_2$ imply

$$S(x_k, \varepsilon, \eta) \subseteq S(x_*, \varepsilon, \eta) + \varepsilon' B.$$

Proof. By Lemma 5.1, for $k > N_1$, we have

$$\begin{aligned} S(x_k, \varepsilon, \eta) &= \mathrm{Conv}\ (\cup\{S_i(x_k, \eta);\ i \in R(x_k, \varepsilon)\}) \\ &\subseteq \mathrm{Conv}\ (\cup\{S_i(x_k, \eta);\ i \in R(x_*, \varepsilon)\}). \end{aligned} \tag{5.11}$$

For each $i \in R(x_*, \varepsilon)$, $S_i(x_k, \eta)$ was defined to be

$$S_i(x_k, \eta) = \sum_{j=1}^{\ell} \partial f_{ij}(G_{ij}(x_k, \eta)). \tag{5.12}$$

From Lemma 5.2, we know that there is some number N_{2i} such that for $k \in K$, $k \geq N_{2i}$ and for all $j = 1, \ldots, \ell$,

$$\partial f_{ij}(G_{ij}(x_k, \eta)) \subseteq \partial f_{ij}(G_{ij}(x_*, \eta)) + \varepsilon' B/\ell. \tag{5.13}$$

Thus (5.12) and (5.13) yield

$$S_i(x_k, \eta) \subseteq S_i(x_*, \eta) + \varepsilon' B, \; k \in K, \; k \geq N_{2i}, \tag{5.14}$$

for all $i \in R(x_*, \varepsilon)$.

Let $N_2' \equiv \mathrm{Max}\,\{N_{2i};\ i \in R(x_*, \varepsilon)\}$ so that $k \in K$ and $k \geq N_2'$ imply

$$\cup\{S_i(x_k, \eta);\ i \in R(x_*, \varepsilon)\} \subseteq \cup\{S_i(x_*, \eta);\ i \in R(x_*, \varepsilon)\} + \varepsilon' B. \tag{5.15}$$

Taking the convex hull of both sides in (5.15), from (5.11) and for k in K, $k > N_2 \equiv \mathrm{Max}\,\{N_1, N_2'\}$

$$S(x_k, \varepsilon, \eta) \subseteq S(x_*, \varepsilon, \eta) + \varepsilon' B. \qquad ||$$

Now, from Lemmas 5.1, 5.2 and 5.3, the uniform negativity of the quantity $\psi(x_k, \varepsilon, \eta)$ is easy to establish.

Theorem 5.4. There exists a number N_2 such that for $k \in K$ and $k > N_2$, the numbers $\psi(x_k, \varepsilon, \eta)$ are uniformly negative and bounded away from 0, that is,

$$\psi(x_k, \varepsilon, \eta) \leq -\bar{b}/2 < 0.$$

Proof. In the result of Lemma 5.3, choose ε' less than or equal to $b/2$. Then, for any g with $||g|| \leq 1$ and $k > N_2$,

$$\begin{aligned} \mathrm{Max}\,\{(d, g);\ d \in S(x_k, \varepsilon, \eta)\} &\leq \mathrm{Max}\,\{(d, g);\ d \in S(x_*, \varepsilon, \eta) + \varepsilon' B\} \\ &\leq \mathrm{Max}\,\{(d, g);\ d \in S(x_*, \varepsilon, \eta)\} + \varepsilon'. \end{aligned}$$

Therefore

$$\begin{aligned} \psi(x_k, \varepsilon, \eta) &= \mathrm{Min}\,\{\mathrm{Max}\,\{(d, g);\ d \in S(x_k, \varepsilon, \eta)\};\ ||g|| \leq 1\} \\ &\leq \mathrm{Min}\,\{\mathrm{Max}\,\{(d, g);\ d \in S(x_*, \varepsilon, \eta)\};\ ||g|| \leq 1\} + \varepsilon' \\ &\leq \psi(x_*, \varepsilon, \eta) + \varepsilon' \\ &\leq \psi(x_*, \varepsilon, \eta) + b/2 \\ &\leq -b/2\ . \end{aligned} \qquad ||$$

We also need the following three lemmas:

Lemma 5.5 (Cullum, et al.[4]) Let F be convex on R^n. Let the sequences $\{x_k\}$ and $\{g_k\}$ satisfy

$$F(x_{k+1}) \leq F(x_k + tg_k), \quad 0 \leq t \leq T,$$

$$g_k \longrightarrow g_* \quad \text{and} \quad x_k \longrightarrow x_*, \; k \to \infty.$$

Then $F'(x_*, g_*) \geq 0$.

Proof. By definition,

$$F'(x_*, g_*) = \lim_{t \to 0^+} [F(x_* + tg_*) - F(x_*)]/t.$$

But for any $t \in [0, T]$,

$$\begin{aligned} F(x_* + tg_*) &= \lim_{k \to \infty} F(x_k + tg_k) \\ &\geq \lim_{k \to \infty} F(x_{k+1}) \\ &\geq F(x_*) \qquad \text{(by continuity).} \end{aligned}$$

The desired result follows immediately. ||

Lemma 5.6 There exists a number N_3 such that for $k > N_3$ and $k \in K$, we have

$$R(x_*, 0) \subseteq R(x_k, \varepsilon).$$

Proof. To prove this statement, we use continuity. Suppose that $i \in R(x_*, 0)$, i.e., $F(x_*) = f_i(x_*)$. Since f_i is continuous, for k large enough, $k > N'_{3i}$ we have

$$|f_i(x_*) - f_i(x_k)| < \varepsilon/2.$$

Thus

$$F(x_*) - \varepsilon/2 < f_i(x_k) < F(x_*) + \varepsilon/2.$$

Similarly, since F is continuous, for $k > N''_3$

$$F(x_*) - \varepsilon/2 < F(x_k) < F(x_*) + \varepsilon/2.$$

Therefore, $k > N_3 = \mathrm{Max}\{\mathrm{Max}\{N'_{3i}; i \in R(x_*, 0)\}, N''_3\}$ and $k \in K$ imply

$$f_i(x_k) - F(x_k) > F(x_*) - \varepsilon/2 - F(x_*) - \varepsilon/2,$$

or $$f_i(x_k) > F(x_k) - \varepsilon, \text{ for all } i \in R(x_*, 0).$$ Thus,

$$i \in R(x_k, \varepsilon), \text{ for all } i \in R(x_*, 0).$$ ||

Lemma 5.7 For any $\varepsilon' > 0$, $\forall i \in R(x_*, 0)$ and $\forall s \in \partial f_i(x_*)$, there exists L_i such that for any $k > L_i$ in K, we can find $s' \in S(x_k, \varepsilon, \eta)$ satisfying

$$s = s' + t, \quad ||t|| < \varepsilon'.$$

Proof. By Lemma 5.6, we know that for $k \in K$, $k > N_3$, $R(x_*, 0) \subseteq R(x_k, \varepsilon)$ and then

$$k \in K,\ k > N_3 \Rightarrow i \in R(x_k, \varepsilon). \tag{5.16}$$

Since $f_i = \sum_{j=1}^{\ell} f_{ij}$, by Theorem 23.8 of [18], we can write s as

$$s = \sum_{j=1}^{\ell} s_j,\ s_j \in \partial f_{ij}(x_*). \tag{5.17}$$

For each $j = 1, 2, \ldots, \ell$, either f_{ij} is differentiable at x_* or it is not. We consider these as two cases.

A. f_{ij} is differentiable at x_*, i.e., $s_j = \nabla f_{ij}(x_*)$ and $\{\nabla f_{ij}(x_*)\} = \partial f_{ij}(x_*)$. Then, according to Corollary 24.5.1 of [18], for any $\varepsilon' > 0$, there exists L_{ij} such that $k \in K$ and $k > L_{ij}$ imply

$$\partial f_{ij}(x_k) \subset \partial f_{ij}(x_*) + \varepsilon' B/\ell. \tag{5.18}$$

Since $\partial f_{ij}(x_*) = \{\nabla f_{ij}(x_*)\} = \{s_j\}$, (5.18) means that we can find $s'_j \in \partial f_{ij}(x_k) \subseteq \partial f_{ij}(G_{ij}(x_k, \eta))$ such that

$$s_j = s'_j + t_j,\ ||t_j|| < \varepsilon'/\ell. \tag{5.19}$$

B. f_{ij} is not differentiable at x_*. Then for k large enough in K, $k > L_{ij}$, $x_* \in N(x_k, \eta)$ and $x_* \in G_{ij}(x_k, \eta)$. Hence for $k > L_{ij}$, if $s_j \in \partial f_{ij}(x_*)$, $s_j \in \partial f_{ij}(G_{ij}(x_k, \eta))$, i.e.,

$$s_j = s'_j + t_j,\ s_j = s'_j,\ t_j = 0. \tag{5.20}$$

Choosing now $k > L'_i \equiv \mathrm{Max}\,\{L_{ij};\ j = 1, \ldots, \ell\}$, we obtain, from (5.19), and (5.20),

$$s = \sum_{j=1}^{\ell} s_j = \sum_{j=1}^{\ell} (s'_j + t_j) = \sum_{j=1}^{\ell} s'_j + \sum_{j=1}^{\ell} t_j, \tag{5.21}$$

where

$$s'_j \in \partial f_{ij}(G_{ij}(x_k, \eta)),\ ||t_j|| < \varepsilon'/\ell,\ j = 1, 2, \ldots, \ell. \tag{5.22}$$

Defining

$$s' = \sum_{j=1}^{\ell} s_j', \; t = \sum_{j=1}^{\ell} t_j,$$

it is clear, from (5.21) and (5.22), that

$$s' \in \sum_{j=1}^{\ell} \partial f_{ij}(G_{ij}(x_k, \eta)) = S_i(x_k, \eta), \tag{5.23}$$

and

$$||t|| = ||\sum_{j=1}^{\ell} t_j|| < \varepsilon'.$$

Furthermore, for $k > L_i = \text{Max}\ \{L_i', N_3\}$, (5.16) implies $S_i(x_k, \eta) \subseteq S(x_k, \varepsilon, \eta)$ so that by (5.23) $s' \in S(x_k, \varepsilon, \eta)$. ||

We now possess the results necessary for the main theorem of this section.

Theorem 5.8 Any limit point x_* of a sequence generated by the algorithm is stationary, i.e., $\psi(x_*, \varepsilon, \eta) = 0$.

Proof. Consider the sequence $\{g_k\}$ defined by

$$\psi(x_k, \varepsilon, \eta) = \text{Max}\ \{(d, g_k);\ d \in S(x_k, \varepsilon, \eta)\}, \quad k \in K.$$

Since $\{g_k\} \subseteq B$, this sequence has a limit point $g_* \in B$. We then have

$$x_k \to x_*,\ g_k \to g_*, \text{ for } k \to \infty \text{ and } k \in K' \subseteq K.$$

We know that for some convex combination

$$\sum_i \lambda_i = 1,\ \lambda_i \geq 0,\ s_i \in \partial f_i(x_*), \quad i \in R(x_*, 0)$$

the directional derivative $F'(x_*, g_*)$ is given by

$$F'(x_*, g_*) = \text{Max}\ \{(d, g_*);\ d \in \partial F(x_*)\}$$

$$= (\sum_i \lambda_i s_i, g_*) \qquad \text{(by (2.4))}.$$

For each i of this convex combination, from Lemma 5.7, there exists L_i such that for $k > L_i$,

$$\forall \varepsilon' > 0,\ \exists\ s_i' \text{ such that } s_i = s_i' + t_i,\ s_i' \in S(x_k, \varepsilon, \eta),\ ||t_i|| \leq \varepsilon'.$$

Choosing $k > N_4 = \text{Max}_i\ \{L_i\}$, $k \in K$, we obtain

$$F'(x_*, g_*) = (\sum_i \lambda_i s_i', g_*) + (\sum_i \lambda_i t_i, g_*)$$

$$= (\sum_i \lambda_i s_i', g_k) + (\sum_i \lambda_i s_i', g_* - g_k) + (\sum_i \lambda_i t_i, g_*). \tag{5.24}$$

But,

$$(\sum_i \lambda_i s_i', g_k) \leq \text{Max } \{(d, g_k);\ d \in S(x_k, \varepsilon, \eta)\} \quad \text{(by convexity of S)}$$

$$\leq \psi(x_k, \varepsilon, \eta) \qquad \text{(by definition of } g_k \text{ and } \psi)$$

$$\leq -b/2 \quad \text{for } k > N_2 \qquad \text{(by Theorem 5.4).}$$

Thus, from (5.24) and for $k > \text{Max } \{N_2, N_4\}$,

$$F'(x_*, g_*) \leq -b/2 + (\sum_i \lambda_i s_i', g_* - g_k) + (\sum_i \lambda_i t_i, g_*). \tag{5.25}$$

Since $g_k \to g_*$ for $k \in K'$, there exists N_5 such that $k > N_5$ and $k \in K'$ imply

$$(\sum_i \lambda_i s_i', g_* - g_k) \leq ||\sum_i \lambda_i s_i'||\ ||g_* - g_k||$$

$$\leq M\ ||g_* - g_k|| < b/8.$$

Furthermore, choosing $0 < \varepsilon' < b/8$, since

$$(\sum_i \lambda_i t_i, g_*) \leq ||\sum_i \lambda_i t_i||\ ||g_*|| \leq \varepsilon',$$

(5.25) becomes

$$F'(x_*, g_*) \leq -b/4 < 0.$$

Considering now Lemma 5.5. with the sub-subsequences $\{x_k\}$ and $\{g_k\}$, $k \in K'$, $T = +\infty$, we have the desired contradiction. Thus, the assumption $\psi(x_*, \varepsilon, \eta) < 0$ is not valid and x_* is a stationary point. ||

6. The Implemented Algorithm

When actually executed on a computer, the result of the difference $F(x) - \varepsilon$, appearing in the definition of $R(x, \varepsilon)$, is not very different from $F(x)$ if $F(x)$ is a large number since in general ε is small. Consequently, $R(x, \varepsilon)$ might indeed be reduced to $R(x, 0)$ through roundoff error and this, in turn, could affect the convergence of the algorithm since, in effect, we no longer consider ε-binding functions.

To avoid this numerical problem, we redefine $R(x, \varepsilon)$, as is done in [6], and use instead

$$R'(x, \varepsilon) = \{i = 1, 2, \ldots, m;\ f_i(x) \geq F(x) - \varepsilon\, F(x)\}.$$

It is also necessary to suppose $F^* > 0$. Then, modifying in a straightforward manner Definitions 2.3 and 2.4, we have $S'(x, \varepsilon, \eta)$ and $\psi'(x, \varepsilon, \eta)$ replacing $S(x, \varepsilon, \eta)$ and $\psi(x, \varepsilon, \eta)$.

Making the obvious modifications, most of the previous results need no extra work, except for four of them that we consider now.

A. The inequalities in Theorem 3.3 should be modified as

$$(1 - \varepsilon)F(x) - 2\ell\eta M \leq F^* \leq F(x). \tag{6.1}$$

Hence, there is another appealing feature in the use of a "relative" ε. Rewriting (6.1) as

$$0 \leq (F(x) - F^*)/F(x) \leq \varepsilon + 2\ell\eta M/F(x), \tag{6.2}$$

suppose that the quantity $\varepsilon + 2\ell\eta M/F(x)$ is equal to 10^{-n}, where n is some positive integer. Then $F(x)$, whatever its magnitude, has at least n correct digits. This is best illustrated with an example:

$$F(x) = 1{,}000{,}010\ , \quad \varepsilon + 2\ell\eta M/F(x) = 10^{-5},$$

$$F(x) - F^* \leq 10^{-5}F(x)$$

$$\leq 10.00010 \quad \text{and}$$

$$1{,}000{,}000. \leq F^* \leq 1{,}000{,}010.$$

B. It is easy to check the correctness of the inclusion

$$R'(x, \varepsilon_2) \subseteq R'(x, \varepsilon_1) \quad \text{for} \quad 0 \leq \varepsilon_2 \leq \varepsilon_1.$$

Then, we still obtain

$$R(x, 0) = R'(x, 0) \subseteq R'(x, \varepsilon), \quad \forall \varepsilon > 0.$$

Hence, in the remark following Example 1

$$\partial F(x) \subseteq S'(x, \varepsilon, \eta)$$

remains correct.

C. It is necessary to rederive Lemma 5.1, using $R'(x, \varepsilon)$.

Lemma 6.1 There exists N_1 such that $k \in K$ and $k > N_1$ imply $R'(x_k, \varepsilon) \subseteq R'(x_*, \varepsilon)$.

Proof. Using the same argument as Dem'yanov [6], it is easy to establish that

$$\exists \varepsilon', \ 0 < \varepsilon < \varepsilon' \text{ such that } R'(x_*, \varepsilon) = R'(x_*, \varepsilon').$$

Since the f_i and F are continuous, we have

$\exists N_1'$ such that $k \in K$ and $k > N_1'$ imply

$|f_i(x_k) - f_i(x_*)| < (\varepsilon' - \varepsilon) F(x_*)/2, \quad i = 1, 2, \ldots, m$, and

$\exists N_1''$ such that $k \in K$ and $k > N_1''$ imply

$|F(x_k) - F(x_*)| < (\varepsilon' - \varepsilon)F(x_*)/2(1 - \varepsilon).$

Thus for $k \in K$ and $k > N_1 = \text{Max } \{N_1', N_1''\}$

$$F(x_*) - (\varepsilon' - \varepsilon)F(x_*)/2(1 - \varepsilon) < F(x_k), \tag{6.3}$$

$$f_i(x_k) < f_i(x_*) + (\varepsilon' - \varepsilon)F(x_*)/2. \tag{6.4}$$

Multiplying (6.3) by $1 - \varepsilon$ and considering also (6.4), we obtain

$$\begin{aligned}(1 - \varepsilon)F(x_*) - (\varepsilon' - \varepsilon)F(x_*)/2 &< (1 - \varepsilon)F(x_k) \\ &< f_i(x_k) \quad (\text{if } i \in R'(x_k, \varepsilon)) \\ &< f_i(x_*) + (\varepsilon' - \varepsilon)F(x_*)/2,\end{aligned}$$

that is,

$$(1 - \varepsilon')F(x_*) < f_i(x_*) \quad \forall i \in R'(x_k, \varepsilon). \tag{6.5}$$

But (6.5) means that $\forall i \in R'(x_k, \varepsilon)$, $i \in R'(x_*, \varepsilon')$. Since $R'(x_*, \varepsilon') = R'(x_*, \varepsilon)$, for $k > N_1$, we have $R'(x_k, \varepsilon) \subseteq R'(x_*, \varepsilon)$. ||

D. Lemma 5.6 needs also to be modified as follows:

Lemma 6.2 There exists N_3 such that for $k \in K$ and $k > N_3$, we have

$$R(x_*, 0) = R'(x_*, 0) \subseteq R'(x_k, \varepsilon).$$

Proof. Choose any $i \in R'(x_*, 0)$, so that $f_i(x_*) = F(x_*) = F_*$. Since f_i is continuous, there exists N'_{3i} such that for $k \in K$ and $k > N'_{3i}$,

$$|f_i(x_*) - f_i(x_k)| < \varepsilon F_*/2. \tag{6.6}$$

Similarly, with F, $\exists N''_3$ such that for $k > N''_3$,

$$|F(x_*) - F(x_k)| < \varepsilon F_*/2. \tag{6.7}$$

Hence, for $k > N_3 = \text{Max}\{\text{Max}\{N'_{3i};\ i \in R'(x_*, 0)\}, N''_3\}$, we get from (6.6) and (6.7),

$$\begin{aligned} f_i(x_k) &> f_i(x_*) - \varepsilon\, F_*/2 \\ &> F_* - \varepsilon\, F_*/2, \text{ and} \end{aligned} \tag{6.8}$$

$$\begin{aligned} -F(x_k) &> -F(x_*) - \varepsilon\, F_*/2 \\ &> -F_* - \varepsilon\, F_*/2. \end{aligned} \tag{6.9}$$

Adding (6.8) and (6.9) yields

$$\begin{aligned} f_i(x_k) - F(x_k) &> -\varepsilon\, F_* \\ &\geq -\varepsilon\, F(x_k) \qquad \text{(since } F(x_k) \geq F_*\text{)}, \end{aligned}$$

that is,

$$i \in R'(x_k, \varepsilon). \qquad ||$$

These are the four principal modifications to signal and, mutatis mutandis, the convergence proof remains valid.

7. Computation Results

We have implemented our algorithm with the modification given in Section 6.

To find the point of minimum norm in $S(x_k, \varepsilon, \eta)$, we used, depending on the problem considered, two different algorithms. For the minimax problems with $m > 1$ and $\ell = 1$, we used an algorithm of Wolfe [20]. Otherwise, for minisum problems, where $m = 1$ and $\ell \geq 1$, we employed an algorithm from Gilbert [10].

The line search was done with quadratic fits and worked well for most problems. We modified it when dealing with piecewise linear functions, since in this case a quadratic fit does not seem reasonable, and devised a different line search which has performed satisfactorily.

We solved three types of problems: minimax location, minisum location and approximation problems. Description of the results are given below.

Minimax location problems.

The problem to solve is

$$\underset{x \in R^n}{\text{Min}} \quad \text{Max} \{f_i(x);\ i = 1, 2, \ldots, m\}$$

where $f_i(x)$ may have one of the following expressions

a) $w_i \; ||x_k - a_i||_p$, for some p, $1 \leq p \leq \infty$,

b) $w_i \; ||x_k - x_j||_p$, for some p, $1 \leq p \leq \infty$,

and $x = (x_1, \ldots, x_k, \ldots, x_r) \in R^n$, $x_k \in R^s$, $k = 1, \ldots, r$ and $n = sr$.

The a_i are interpreted as known locations in R^s of existing facilities and the x_k are the locations of the new facilities, and where the w_i are positive weights.

Results of the first minimax location problems solved are given in Table 1. M indicates the number of existing facilities and N the number of new facilities to be located. We ran three Type 1 problems involving ℓ_2-norms, which were randomly generated with weights between 1 and 100 and existing facility locations in the square 1000 by 1000. Problem 2 was a selected minimax location problem posed by Love, Wesolowsky and Kraemer and considered in [7]. Problem 2 involves ℓ_2-norms. Problems 3 and 4 are ℓ_2-norm problems discussed in [7]. The data in Problem 5 is identical to the data in Problem 2, however ℓ_1-norms are used. Finally, Problem 6 is a randomly generated problem using ℓ_1-norms.

For comparative purposes, we have included the execution times on these problems using:

(a) The heuristic subgradient procedure of Hearn and Lowe [12] (all problems).

(b) A dual formulation solved by GRG [7], in the case of ℓ_2-norms (Problems 1 through 4).

(c) Another dual procedure studied by Dearing and Francis [5], in the case of ℓ_1-norms (Problems 5 and 6).

(d) The subgradient algorithm (all problems).

We have also listed the number of iterations using the subgradient algorithm. In those cases where the last iterate was not a stationary point, the line search could not generate a point to improve the objective function value.

	Problem	Execution Time				Iterations
		(a)	(b)	(c)	(d)	
1)	Random data, ℓ_2-norm	1.66s	2.03s	—	3.55s	25
	M = 50, N =1	1.15s	1.57s	—	1.52s	11*
		.95s	2.28s	—	2.52s	16
2)	Love et al, ℓ_2-norm M = 5, N = 2	.39s	2.06s	—	.66s	14*
3)	Triangle #1, ℓ_2-norm M = 3, N = 3	.64s	5.5s	—	1.11s	16*
4)	Triangle #2, ℓ_2-norm M = 3, N = 3	.80s	10.5s	—	1.40s	20*
5)	Love et al, ℓ_1-norm M = 5, N = 2	.24s	—	2.40s	.17s	6*
6)	Random data, ℓ_1-norm M = 20, N = 10	1.16s**	—	2.89s	6.94s	15

* The last iterate is a stationary point

** Used a fixed step length.

Table 1

For the problems of Table 1, the value chosen for ε was 5×10^{-6} and we set $\eta = 10^{-6}$ or 10^{-5}.

We also devised a problem of our own which involved mixed norms. The problem scenario reads as follows. Two ships have to be located in the Caribbean sea and must be ready to intervene, in case of trouble, at any one of nine given cities of the Caribbean Islands. Trouble may occur according to estimated probabilities which are used as weights. Furthermore, the two ships must be able to communicate and we must consider their mutual distance (Euclidean distance with weight 1). Thus, the problem is

$$\underset{\substack{x_1\in R^2\\ x_2\in R^2}}{\text{Min}} \quad \underset{i=1,\ldots,9}{\text{Max}} \{w_{i1}||x_1 - a_i||_{pi1}, w_{i2}||x_2 - a_i||_{pi2}, ||x_1 - x_2||\}.$$

The cities and their locations are displayed in Table 2. For some cities, the distance to a ship is well represented as Euclidean distance. But if one city is on the opposite side of the island with respect to the ship location, we can no longer use Euclidean distance but an ℓ_p-distance where p is chosen between 1 and 2 (see Table 2).

cities	locations		1st ship		2nd ship	
			w_{i1}	p_{i1}	w_{i2}	p_{i2}
Colon (Panama Canal)	11.4,	11.6	2.0	2.0	1.0	2.0
Caracas-LaGuaira (Venezuela)	35.3,	13.5	1.0	2.0	2.0	2.0
Havana (Cuba)	8.80,	37.2	1.5	1.1	1.0	1.4
Guantanamo (Cuba)	20.9,	30.6	1.5	1.5	1.0	1.9
Port-au-Prince (Haiti)	25.5,	28.0	1.5	1.4	1.5	1.2
Santo Domingo (Dom. Rep.)	29.7,	27.7	1.0	2.0	1.5	2.0
San Juan (Puerto Rico)	36.2,	27.8	0.5	1.8	1.0	1.7
Fort-de-France (Martinique)	45.5,	21.3	0.5	2.0	0.5	2.0
Montego Bay (Jamaica)	15.8,	28.2	0.5	1.1	0.5	1.8

Table 2

The optimal locations are $(13.817, 24.358)^T$ and $(25.818, 22.454)^T$, with $F^* = 26.0836$. This solution was found after 40 iterations and 3.15s of CPU time. As a final check, we used the locations for the ships given by the algorithm and measured the ℓ_p- distances between the ships and ports and the ℓ_p-distance between the ships. All ℓ_p-distances approximately agreed with the ℓ_p-distances used in the algorithm with one exception. The ℓ_p-distances from the ships to Port-au-Prince were not in agreement with the ℓ_p-distances used in the algorithm, but this was not important in our problem since at the optimal solution, the functions involving Port-au-Prince were not binding.

Minisum location problems.

The objective function to minimize has the form

$$F(x) = \sum_i f_i(x), \quad x \in R^n,$$

where the f_i are as in the minimax location problem.

Among many minisum problems that we solved, we give here the numerical results for two problems:

a) The 24 cities problem of Kuhn and Kuenne [13].

Starting from the center of gravity of the 24 existing facilities, $x_0 = (47.47945, 34.35616)$, the algorithm generates the result $x^* = (47.70779, 35.10391)$ in 6 iterations and .46s of CPU time, giving $F^* = .9528840$.

After 5 iterations, the procedure of Kuhn and Kuenne reaches the result (47.60, 35.32), which corresponds to an objective function value of .9541373.

b) A problem from Eyster et al. [8], with five existing and two new facilities.

The starting point is $x_0 = (0, 0, 0, 0)$ and after 23 iterations we got $x^* = (2.840055, 2.686639, 5.129290, 6.388482)$ with $F^* = 67.23856$. The CPU time for this problem is 1.07s. The HAP procedure used in [8], starting also at x_0, stopped after 45 iterations and gave $x^* = (2.840, 2.687, 5.126, 6.383)$. $F^* = 67.239$

In both problems a and b, we chose $\eta = 10^{-5}$.

Linear Approximation Problems

A linear approximation problem can be formulated as follows

$\underset{x \in R^n}{\text{Min}}$ $K(Ax - b)$, where A is an m x n matrix and b is a vector of R^m and where K is some norm in R^m measuring the discrepancy between a desired point, b, and an approximation, Ax, of this point.

Popular choices for K are the ℓ_1-norm and the ℓ_∞-norm, so that we solve either

$$\underset{x \in R^n}{\text{Min}} \sum_{i=1}^{m} |(Ax)_i - b_i|,$$

or

$$\underset{x \in R^n}{\text{Min}} \ \text{Max} \ \{|(Ax)_i - b_i|; \ i = 1, 2, \ldots, m\}.$$

With the ℓ_∞-norm, and thus a minimax problem, we solved the examples given by Barrodale and Young [1, p. 115]. For these problems, we chose $\varepsilon = 10^{-4}$ and $\eta = 10^{-16}$. The results are presented in Table 3. The solutions we found are in total agreement with those of [1]. Barrodale and Young did not report any computation times or numbers of iterations.

Problem	Execution Time	Iterations	Optimal Value
1) Example 1 m = 33, n = 3	.85s	12*	$.6733 \times 10^{-4}$
2) Example 2 m = 41, n = 6	9.90s	58*	$.1514 \times 10^{-5}$
3) Example 3, Case 1 m = 51, n = 5	2.54s	41*	$.1039 \times 10^{-5}$
4) Example 3, Case 2 m = 51, n = 5	2.86s	39*	$.1757 \times 10^{-5}$

* The last iterate is a stationary point.

Table 3

References

[1] I. Barrodale and A. Young, "Computational experience in solving linear operator equations using the Chebyshev norm", in Numerical Approximation to Functions and Data, Hayes S. G. (ed.) (The Athlone Press, London, 1970) 115-142.

[2] D. P. Bertsekas and S. Mitter, "A descent numerical method for optimization problems with nondifferentiable cost functionals", SIAM Journal on Control 11 (1973) 637-652.

[3] F. H. Clarke, "Generalized gradients and applications", Trans. Amer. Math. Soc. 205 (1975) 247-262.

[4] J. Cullum, W. E. Donath and P. Wolfe, "The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices", Mathematical Programming Study 3 (1975) 35-55.

[5] P. M. Dearing and R. L. Francis, "A network flow solution to a multifacility minimax location problem involving rectilinear distances", Transportation Science 8 (1974) 126-141.

[6] V. F. Dem'yanov and V. N. Malozemov, Introduction to minimax, (John Wiley, 1974).

[7] J. Elzinga, D. W. Hearn and W. D. Randolph, "Minimax multifacility location with Euclidean distances", (1976) to appear in Transportation Science.

[8] J. W. Eyster, J. A. White and W. W. Wierwille, "On Solving Multifacility Location Problems Using a Hyperboloid Approximation Procedure", A.I.I.E. Trans. 5 (1973) 1-6.

[9] A. Feuer, "An implementable mathematical programming algorithm for admissible fundamental function", Ph.D. Dissertation, Department of Mathematics, Columbia University, New York (1974).

[10] E. G. Gilbert, "An iterative procedure for computing the minimum of a quadratic form on a convex set", SIAM Journal on Control 4 (1966) 61-80.

[11] J. Greenstadt, "Variations on variable metric methods", Maths of Comp., 24 (1970) 1-22.

[12] D. W. Hearn and T. J. Lowe, "A subgradient procedure for the solution of minimax location problems", Department of Industrial and Systems Engineering Research Report 76-6, University of Florida (1976).

[13] H. W. Kuhn and R. E. Kuenne, "An efficient algorithm for the numerical solution of the generalized Weber problem in spatial economics", Journal of Regional Science 4 (1962) 21-33.

[14] C. Lemaréchal, "An extension of Davidon methods to nondifferentiable functions", Mathematical Programming Study 3 (1975) 95-109

[15] D. G. Luenberger, Introduction to linear and nonlinear programming, (Addison-Wesley, 1973).

[16] R. Mifflin, "An algorithm for nonsmooth optimization", School of Organization and Management, Yale University (1975).

[17] Pshenichnyi, Necessary conditions for an extremum, (Marcel Dekker, New York, 1971).

[18] R. T. Rockafellar, Convex analysis (Princeton University Press, N. J., 1970).

[19] P. Wolfe, "A method of conjugate subgradients for minimizing non-differentiable functions", Mathematical Programming Study 3, (1975) 145-173.

[20] P. Wolfe, "Algorithm for a least distance programming problem", Mathematical Programming Study 1 (1974) 190-205.