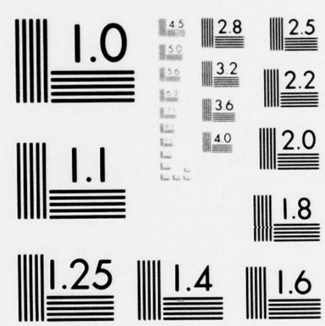
AD-A062 089 FLORIDA UNIV GAINESVILLE DEPT OF INDUSTRIAL AND SYS--ETC F/G 12/1
A FEASIBLE DIRECTION SUBGRADIENT ALGORITHM FOR A CLASS OF NONDI--ETC(U)
OCT 78 J CHATELON, D HEARN, T J LOWE DAHC04-75-G-0150
UNCLASSIFIED RR-78-13 ARO-12640.35-M NL

| OF |

AD
A062089

END
DATE
FILMED
3-79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0

ARO 12640.35-M

LEVEL

(12)

RESEARCH REPORT

Industrial & Systems
Engineering Department
University of Florida
Gainesville, FL. 32611

78 12 04.007

A FEASIBLE DIRECTION SUBGRADIENT ALGORITHM FOR A CLASS OF NONDIFFERENTIABLE OPTIMIZATION PROBLEMS

Research Report No. 78-13

by

Jacques Chatelon*
Donald Hearn**
Timothy J. Lowe†

October 1978

**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-76-17810 and ENG-76-24294.

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.

*ITT, Paris, France
†Krannert Graduate School of Management, Purdue University, West Lafayette, IN 47907.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: ~~78-13~~	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): A Feasible Direction Subgradient Algorithm for a Class Nondifferentiable Optimization Problem	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER: 78-13
7. AUTHOR(s): Jacques Chatelon, Donald Hearn, Timothy J. Lowe	8. CONTRACT OR GRANT NUMBER(s): DAHC04-75-G-0150, ENG-76-17810, ENG-76-24294
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: 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: October, 1978
	13. NUMBER OF PAGES: 24
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

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

We present an implementable feasible direction subgradient algorithm for minimizing the maximum of a finite collection of functions subject to constraints. It is assumed that each function involved in defining the objective function is the sum of a finite collection of basic convex functions and that the number of different subgradient sets associated with nondifferentiable points of each basic function is finite on any bounded set. It is demonstrated that under certain conditions, including continuous differentiability of the constraints and a

20. (cont'd)

regularity condition of the feasible region, that the algorithm generates a feasible sequence which converges to an ε-optimal solution.

The results of some computational experiments are included.

TABLE OF CONTENTS

Abstract

We present an implementable feasible direction subgradient algorithm for minimizing the maximum of a finite collection of functions subject to constraints. It is assumed that each function involved in defining the objective function is the sum of a finite collection of basic convex functions and that the number of different subgradient sets associated with nondifferentiable points of each basic function is finite on any bounded set. It is demonstrated that under certain conditions, including continuous differentiability of the constraints and a regularity condition of the feasible region, that the algorithm generates a feasible sequence which converges to an ε-optimal solution.

The results of some computational experiments are included.

1. Introduction

In this paper, we develop an implementable feasible direction algorithm for solving a class of nondifferentiable nonlinear programming problems of the form:

$$(P)\quad \min_{x \in C} F(x),$$

where

$$F(x) \equiv \max \{f_i(x) = \Sigma_{j=1}^{\ell} f_{ij}(x) \; ; \; i = 1, \ldots, m\},$$

f_{ij} finite, convex, not necessarily differentiable,

$$C \equiv \{x \in R^n \; ; \; H(x) \leqq 0\},$$

$$H(x) \equiv \max \{h_i(x) \; ; \; i = 1, \ldots, K\},$$

$h_i(x)$ convex on R^n.

Our present work is related to an earlier paper [1] where we presented an algorithm for solving an unconstrained version of problem (P).

Our approach follows that of Dem'yanov and Malozemov [5] who employ subgradients to devise a descent algorithm for minimizing the maximum of continuously differentiable convex functions, i.e., problem (P) with all f_i differentiable and $C = R^n$. For related literature on nondifferentiable optimization methods, the surveys by Mifflin [13, 14] are recommended. He traces, e.g., the development of heuristic methods by Held and Karp [8] and Held, Wolfe and Crowder [9]; the convergent methods of Polyak [15] and Bertsekas and Mitter [4]; and the conjugate-type methods of Lemarechal [10] and Wolfe [17]. Mifflin's algorithm [13] for problems with "weakly upper semismooth" functions [14] builds on the notion of generalized gradients introduced by Clarke [2] for Lipschitz functions. While (P) is a special case of that problem type, the algorithm developed here differs in that it is a feasible direction method.

Much of the notation in this paper is similar to the notation in [1]. Denote n dimensional Euclidean space as R^n and $||x||_p$ as the ℓ_p-norm of $x \varepsilon R^n$, where $||x||$ is the ℓ_2-norm. Given a point $x \varepsilon R^n$, the Euclidean ball about x of radius η is $N(x, \eta)$; when $x = 0$ and $\eta = 1$, $B = N(0, 1)$ is the Euclidean unit ball. For a function f defined on R^n let $\partial f(x)$ be the subgradient set of f at x and let $f'(x, d)$ be the directional derivative of f at x in the direction d. Given a set $S \subset R^n$, Conv (S) is the convex hull of S and Nr(S) is the element of minimum Euclidean norm in S. Also denote $\partial f(S) \equiv \cup \{\partial f(x) \; ; \; x \varepsilon S\}$.

For convenience, the functions f_i in problem (P) are assumed to be the sum of exactly ℓ functions f_{ij}, where perhaps for some i, some of the f_{ij} functions are identically zero on R^n. Clearly, both F and H are continuous, finite, convex functions on R^n.

Given $\varepsilon \geqq 0$ and $\mu \geqq 0$, at any $x \varepsilon R^n$ we define

$$R(x, \varepsilon) \equiv \{i \varepsilon\{1, 2, \ldots, m\} \; ; \; f_i(x) \geqq F(x) - \varepsilon\},$$

and

$$Q(x, \mu) \equiv \{i \varepsilon\{1, 2, \ldots, K\} \; ; \; 0 \geqq h_i(x) \geqq -\mu\}.$$

With these definitions, we can interpret f_i, $i \varepsilon R(x, \varepsilon)$, as an "$\varepsilon$-binding" objective function at x; and h_i, $i \varepsilon Q(x, \mu)$, as a "μ-binding" constraint at x. When $H(x) \leqq -\mu$, we say that x is a <u>μ-feasible</u> point and the set of all x in R^n where $H(x) \leqq -\mu$ is the <u>μ-feasible set</u>.

Given $x \varepsilon R^n$ and $\eta \geqq 0$, we define

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

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

In addition, let

$$S^1(x, \varepsilon, \eta) \equiv \cup \{S_i(x, \eta) \; ; \; i \varepsilon R(x, \varepsilon)\}.$$

To handle constraints, we use a procedure somewhat similar to that presented in [13]. At any $x \varepsilon R^n$, we consider subgradients of μ - binding constraints by defining $S^2(x, \mu) \equiv \cup\{\partial h_i(x) ; i \varepsilon Q(x, \mu)\}$ and letting $S(x, \varepsilon, \mu, \eta) \equiv \text{Conv}\ \{S^1(x, \varepsilon, \eta) \cup S^2(x, \mu)\}$.

We note that for an unconstrained problem, $S(x, \varepsilon, \eta, \mu) = S^1(x, \varepsilon, \eta)$ is precisely the enlargement of the subgradient set considered in [1].

We assume the functions f_{ij} are LFS (locally finitely subdifferentiable), which means that in any closed bounded Euclidean ball, the number of different subgradient sets of f_{ij} corresponding to the points of nondifferentiability, is finite. In [1], we cite several examples of LFS functions, including examples from location theory and linear approximation problems.

Associated with $S(x, \varepsilon, \mu, \eta)$, the function Ψ measures the proximity of $S(\cdot)$ to zero:

$$\Psi(x, \varepsilon, \mu, \eta) \equiv \min\ \{\max\ \{(g, d) ; d \varepsilon S(x, \varepsilon, \mu, \eta)\} ; g \varepsilon B\}.$$

It is easily established that

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

We note that Ψ is well defined since S is a nonempty, compact, convex subset of R^n. Further Ψ is always nonpositive. When $\Psi(x, \varepsilon, \mu, \eta) = 0$, we call x a stationary point, and any x where $\Psi(x, \varepsilon, \mu, \eta) < 0$ is a nonstationary point. In the next section we consider properties of stationary and nonstationary points.

2. Stationary and Nonstationary Points

In the unconstrained problem, information given by the value of Ψ is relatively easy to exploit [1] since the set $S(x, \varepsilon, \eta)$ is derived solely from the f_i functions which are ε-binding. In the constrained problem this is no longer the case, since, to construct $S(x, \varepsilon, \mu, \eta)$, we also consider the subgradient sets of the μ-binding constraints. Consequently, stationarity in the constrained problem will not always imply a lower bound on the minimum value of F on C, F*, as is the case with the unconstrained problem. In the presence of constraints, we show in the next theorem that, while it may be possible to obtain a lower bound on F*, it may well be the case that the only implication of stationarity is the emptiness of the interior of the μ-feasible set.

Before proving this theorem, we assume that the minimum, F*, of F(x) over C exists. In addition, we assume that there exists an upper bound, δ, on the norm of any subgradient of any function f_{ij} or h_i at any point in C.

Theorem 2.1 Let $x \in C$ be a stationary point.

a) If $H(x) < -\mu$, then

$$F(x) \geqq F^* \geqq F(x) - \varepsilon - 2\ell\eta\delta. \tag{2.1}$$

b) If $-\mu \leqq H(x) \leqq 0$, at least one of the following is true:

b1) (2.1) holds,

b2) The interior of the μ-feasible set is empty,

b3) $F(z) \geqq F(x) - \varepsilon - 2\ell\eta\delta$, for all μ-feasible z.

Proof: Stationarity at x is equivalent to $0 \in S(x, \varepsilon, \mu, \eta)$, which can occur if and only if there exists $g_1, \ldots, g_q$, where $g_i \in \{S^1(x, \varepsilon, \eta) \cup S^2(x, \mu)\}$ and $0 \in \mathrm{Conv}(g_1, \ldots, g_q)$. Index the g_i so that $g_i \in S^1(x, \varepsilon, \eta)$ for $1 \leqq i \leqq q_1$ and $g_i \in S^2(x, \mu)$ for $q_1 + 1 \leqq i \leqq q$ Thus for i, $1 \leqq i \leqq q_1$, $g_i \in \partial f_{(i)}(y_i)$, where $f_{(i)}(x) \geqq F(x) - \varepsilon$ and $||x - y_i|| \leqq \eta$; and for i, $q_i + 1 \leqq i \leqq q$, $g_i \in \partial h_{(i)}(x)$, where $0 \geqq h_{(i)}(x) \geqq -\mu$.

We note in the case where $q_1 = q$ (which certainly holds in case a)), that $g_i \in S^1(x, \varepsilon, \eta)$, $1 \leqq i \leqq q$, and thus by Theorem 3.2 of [1], (2.1) holds. Thus suppose $q_1 < q$, in which case $0 = q_1$ or $0 < q_1$.

If $0 = q_1$, then $g_i \in S^2(x, \mu)$ for all i. By the subgradient inequality, $h_{(i)}(z) \geqq h_{(i)}(x) + (g_i, z-x) \;\; \forall z \in R^n$, and further $h_{(i)}(x) \geqq -\mu$ and $H(z) \geqq h_{(i)}(z)$. Thus

$$H(z) \geqq h_{(i)}(z) \geqq -\mu + (g_i, z-x) \quad \forall z \in R^n.$$

Taking the convex combination over all $i = 1, \ldots, q$, yields $H(z) \geqq -\mu$, $\forall z \in R^n$, and hence the μ-feasible set has an empty interior, which establishes case b2).

Consider the remaining case, $1 \leqq q_1 < q$. As in [1], since each function f_i is Lipschitz,

$$F(z) \geqq F(x) - \varepsilon + (g_i, z-x) - 2\ell\eta\delta \;, \forall z \in R^n, \; 1 \leqq i \leqq q_1; \text{ and}$$

thus for all $z \in R^n$,

$$F(z) \geqq F(x) - \varepsilon - 2\ell\eta\delta + \max\{(g_i, z-x) : i = 1, \ldots, q_1\}. \quad (2.2)$$

For any μ-feasible z, and $i = q_1 + 1, \ldots, q$,

$$-\mu \geqq h_{(i)}(z) \geqq h_{(i)}(x) + (g_i, z-x) \geqq -\mu + (g_i, z-x),$$

which implies $(g_i, z-x) \leqq 0$. Writing zero as a convex combination of all g_i leads to $\max\{(g_i, z-x) : i = 1, \ldots, q_1\} \geqq 0$ for all μ-feasible z. Thus the final term of (2.2) may be deleted and case b3) is established.

<u>Remark 2.1</u> The definition of q_1 in the previous proof implies that if a point g_i is in both sets $S^1(x, \varepsilon, \eta)$ and $S^2(x, \mu)$, it should be considered as a point of $S^1(x, \varepsilon, \eta)$. This forces q_1 to be equal to q if possible and thus to obtain the more useful cases (a) or (b1) in which we have bounds on the constrained minimum of F.

<u>Remark 2.2</u> Cases (b1), (b2), and (b3) are not mutually exclusive. When several different convex combinations of elements in $S^1(x, \varepsilon, \eta)$ and $S^2(x, \mu)$ are 0, (b1), (b2), and (b3) can occur simultaneously.

To illustrate the different situations described in Theorem 2.1 we consider the following example.

<u>Example 2.1</u> Min $F(x)$, $F(x) \equiv ||x - (.5, -.5)||$

subject to

$h_1(x) \equiv ||x||_1 - 1.6 \leq 0,$

$h_2(x) \equiv ||x - (1, -1)||_1 - 1.6 \leq 0,$

Let $\varepsilon = 0$, $\eta = \sqrt{2} \times 10^{-1}$. It is clear that $\delta = \sqrt{2}$ and $x^* = (.5, -.5)$ is the optimal solution to this problem. For different values of μ, stationarity may have distinct implications.

(a) Let $\mu = 0$. The point $y_0 = (.6, -.4)$ is feasible, with $H(y_0) = \text{Max}\{-.6, -.6\} = -.6 < -\mu$. y_0 is also stationary because $(.5, -.5)$ is in $N(y_0, \eta)$. We observe that $F(y_0) = \sqrt{2} \times 10^{-1}$, $F(y_0) - 2\ell\eta\delta = -.2588$ and we obtain case (a) of Theorem 2.1, that is,

$$\sqrt{2} \times 10^{-1} \geq 0 = F^* = \text{Min}\{F(z) : z \in C\} \geq -.2588.$$

(b) Let $\mu = 0.6$. y_0 remains stationary with $H(y_0) = -.6 = -\mu$. In this case, in addition to the lower bound on F^*, $\{z \in R^2 : H(z) < -.6\}$ is empty, and both (b1) and (b2) are obtained. On the other hand, observe that $y_1 = (1, 0)$ is stationary since $H(y_1) = h_1(y_1) = h_2(y_1) = -.6 = -\mu$ so that both points $(1, -1) \in \partial h_1(y_1)$ and $(-1, 1) \in \partial h_2(y_1)$ are in $S(y_1, \varepsilon, \mu, \eta)$. However, the only conclusion is that the set $\{z \in R^2 : H(z) < -.6\}$ is empty.

Now let $\mu = 0.6$, and add a third constraint, $h_3(x) = -x_1 - x_2 \leq 0$, to Example 2.1. Consider $y_2 = (.8, -.2)$. Since $H(y_2) = \text{Max}\{-.6, -.6, -.6\} = -.6 = -\mu$, y_2 is feasible. Moreover y_2 is stationary. It is easy to verify that $S(y_2, \varepsilon, \mu, \eta)$ is the square with vertices $(-1, 1)$, $(1, -1)$, $(-1, -1)$, $(1, 1)$. Here the lower bound obtained is only valid on the μ-feasible set, which is the line segment with end points y_2 and $(1, 0)$. With $F(y_2) = 3\sqrt{2} \times 10^{-1}$, this lower bound is positive and equal to .0242. Hence, at y_2, both (b2) and (b3) hold.

Remark 2.3 It is clear that case (b2) of Theorem 2.1 is relatively undesirable since nothing can be said about the objective function itself. To discard this case, a constraint qualification, similar to Slater's constraint qualification [12] is necessary. Thus, if the set $\{z \in R^n : H(z) < -\mu\}$ is nonempty, case (b2) cannot occur.

Having dealt with stationary points in Theorem 2.1, we now consider nonstationary points. Given a nonstationary point x, the subgradient sets in $S(x, \varepsilon, \mu, \eta)$ relative to the functions f_i ensure that we can find a descent direction, and the subgradient sets ∂h_i, if any, ensure that this descent direction is feasible.

Theorem 2.2 If $\Psi(x, \varepsilon, \mu, \eta) < 0$, there exists a feasible descent direction for F at X.

Proof: Let $g_o \neq 0$ be the element of minimum norm in $S(x, \varepsilon, \mu, \eta)$, i.e.,

$$0 > \Psi(x, \varepsilon, \mu, \eta) = -||g_o|| = -||Nr(S(x, \varepsilon, \mu, \eta))||.$$

Define $d_o = -g_o/||g_o||$. If $H(x) = 0$, then for any i such that $h_i(x) = H(x) = 0$, $\partial h_i(x) \subset S(x, \varepsilon, \mu, \eta)$ so that

$$\begin{aligned} h_i'(x, d_o) &= \max\{(g, d_o) ; g \in \partial h_i(x)\} \\ &\leqq \max\{(g, d_o), g \in S(x, \varepsilon, \mu, \eta)\} \\ &= -\min\{(g, -d_o), g \in S(x, \varepsilon, \mu, \eta)\} \\ &= -1/||g_o|| \min\{(g, g_o), g \in S(x, \varepsilon, \mu, \eta)\} \\ &= -||g_o|| < 0. \end{aligned}$$

Hence, d_o is a feasible direction at x. On the other hand, if $H(x) < 0$, the direction d_o is feasible since the functions h_i are continuous.

In either case above, since $\partial F(x) \subset S(x, \varepsilon, \mu, \eta)$ it follows from Theorem 3.3 of [1] that $F'(x, d_o) < 0$. Therefore d_o is a descent direction for F at x.

3. The Algorithm

The algorithm exploits the results of the previous section. We choose positive values for the three parameters ε, μ, and η. Let x_o be a feasible starting point, set $k = 0$ and go to Step 1.

Step 1

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

Step 2

If $\Psi(x_k, \varepsilon, \mu, \eta) = 0$, stop; x_k is a stationary point.
If $\Psi(x_k, \varepsilon, \mu, \eta) = 0 < 0$, define g_k as the element of minimum norm in $S(x, \varepsilon, \mu, \eta)$ and let $d_k = -g_k/||g_k||$. Perform a restricted line search along d_k, finding t_k such that

$$F(x_k + t_k d_k) = \text{Min}\{F(x_k + td_k) \ ; \ t \geq 0, H(x_k + td_k) \leq 0\}.$$

Update $x_k \rightarrow x_{k+1} = x_k + t_k d_k$, $k \rightarrow k + 1$ and return to Step 1.

In the next section, under supplementary assumptions, we prove that limit points of the algorithm are stationary points for problem (P).

4. Proof of Convergence

In this section we assume that the algorithm does not stop, but generates an infinite sequence $\{x_k\}$ converging to some limit x_*. The proof that x_* is stationary depends on the approximation of $S(x_*, \varepsilon, \mu, \eta)$ by $S(x_k, \varepsilon, \mu, \eta)$. Although the $S(x_k, \varepsilon, \mu, \eta)$, do not necessarily converge to $S(x_*, \varepsilon, \mu, \eta)$, for sufficiently large k, $S(x_k, \varepsilon, \mu, \eta)$ is contained in $S(x_*, \varepsilon, \mu, \eta)$ plus an epsilon ball. In addition to the assumption that the functions f_i are LFS, two additional assumptions are required in the proofs:

Assumption 4.1 There exists some $x_o \in C$, a starting point for the algorithm, such that the intersection, X, of C with the level set

$\{x \varepsilon R^n : F(x) \leqq F(x_o)\}$ is nonempty and bounded. By continuity of F and H, X is also closed.

Assumption 4.2 The constraint functions, h_i, i = 1, 2, ... K are continuously differentiable as well as convex.

The first assumption guarantees a solution point x^* for (P) and insures that a limit point x_* exists. The assumption that the h_i are continuously differentiable is required for the stationarity proof, especially Lemma 4.5. In the next section we show by counterexample that in the absence of this assumption convergence to a nonstationary point is possible.

Lemma 4.1 For k sufficiently large,

$$Q(x_k, \mu) \subset Q(x_*, \mu).$$

Proof: Follows from Assumption 4.2.

We now show that $S(x_k, \varepsilon, \mu, \eta)$ approximates the set $S(x_*, \varepsilon, \mu, \eta)$. In the proof of this result, we use Theorem 5.2 of [1].

Lemma 4.2 For any $\gamma > 0$, there exists N_1 such that

$$S(x_k, \varepsilon, \mu, \eta) \subset S(x_*, \varepsilon, \mu, \eta) + \gamma B, \; k > N_1.$$

Proof: By definition,

$$S(x_k, \varepsilon, \mu, \eta) = \text{Conv}\,(S^1(x_k, \varepsilon, \eta) \cup S^2(x_k, \mu)). \qquad (4.1)$$

Since $S^1(x_k, \varepsilon, \eta)$ is identical to $S(x_k, \varepsilon, \eta)$ in [1], by Theorem 5.2 of [1], there exists N_1' such that

$$S^1(x_k, \varepsilon, \eta) \subset S^1(x_*, \varepsilon, \eta) + \gamma B \;, \; k > N_1'. \qquad (4.2)$$

Now consider the sets $S^2(\cdot)$ in (4.1). From Corollary 24.5.1 of [16], since the functions h_i are assumed continuously differentiable, for each $i \varepsilon Q(x_*, \mu)$, there exists L_i such that

$$\partial h_i(x_k) \subset \partial h_i(x_*) + \gamma B \;, \; k > L_i. \qquad (4.3)$$

It follows from the definition of $S^2(\cdot)$, (4.3) and Lemma 4.1 that there exists N_1'' such that

$$S^2(x_k, \mu) \subset S^2(x_*, \mu) + \gamma B \ , \ k > N_1''. \qquad (4.4)$$

Letting $N_1 = \max\{N_1', N_1''\}$ and using (4.2), (4.3) and (4.4), the result follows.

<u>Corollary 4.3</u> If $\Psi(x_*, \varepsilon, \mu, \eta) < -2\gamma < 0$, then $\Psi(x_k, \varepsilon, \mu, \eta) < -\gamma$, $k > N_1$.

<u>Proof:</u> See Corollary 5.3 of [1].

The following lemma is a partial converse of Lemma 4.2. It shows that any subgradient of any binding function at x_* can be approximated by an element of $S(x_k, \varepsilon, \mu, \eta)$ for large k.

<u>Lemma 4.4</u> Choose any $\alpha > 0$ and let $s \in \partial f_i(x_*)$; $i \in R(x_*, 0)$. Then for k larger than some N_2 there exists $s' \in S(x_k, \varepsilon, \mu, \eta)$ and t such that

$$s = s' + t \ , \ ||t|| < \alpha.$$

<u>Proof:</u> The proof follows from the proof of Lemma 5.4 of [1] upon noting that $s' \in S(x_k, \varepsilon, \eta) \subset S(x_k, \varepsilon, \mu, \eta)$.

<u>Lemma 4.5</u>. *For any k greater than the* N_1 *of* Corollary 4.3, there exists some $T > 0$, independent of k, where with d_k as chosen in Step 2 of the algorithm,

$$H(x_k + t\, d_k) \leqq 0, \text{ and}$$

$$F(x_{k+1}) \leqq F(x_k + t\, d_k), \text{ for all } t \in [0, T].$$

<u>Proof:</u> By Assumption 4.2, the h_i are continuously differentiable on R^n and thus from Remark 2 of [5, p. 270], for any i = 1, ..., k, there exists $T_{oi} > 0$ such that for all $t \in [0, T_{oi}]$,

$$h_i(x + td) = h_i(x) + t(\nabla h_i(x), d) + \sigma_i(x, d\ ;\ t), \qquad (4.5)$$

where σ_i is a function with the property that $\sigma_i(x, d\ ;\ t)/t \to 0$ uniformly in $x \in X$ and d, $||d|| = 1$, as $t \to 0$. Because of the uniform convergence of $\sigma_i(x, d\ ;\ t)/t$, we can choose T'_{oi}, $0 < T'_{oi} \leqq T_{oi}$, such that

$$t \in [0, T'_{oi}] \text{ implies } |\sigma_i(x, d\ ;\ t)| < \gamma\, t/2, \qquad (4.6)$$

for all $x \in X$ and all d where $||d|| = 1$.

Letting $T'_o \equiv \min\{T'_{oi} \; ; \; i = 1, \ldots, K\}$, it follows from (4.5) and (4.6) that for any $i = 1, \ldots, K$, any $x \in X$ and any d, $||d|| = 1$, if $t \in [0, T'_o]$, then

$$h_i(x + td) \leqq h_i(x) + t(\nabla h_i(x), d) + \gamma\, t/2. \tag{4.7}$$

At a point x_k, for any $i = 1, \ldots, K$, either $i \in Q(x_k, \mu)$ or not. If $i \in Q(x_k, \mu)$, then with $h_i(x_k) \leqq 0$ and choosing $d = d_k$ in (4.7),

$$h_i(x + td_k) \leqq t(\nabla h_i(x_k), d_k) + \gamma\, t/2. \tag{4.8}$$

Noting that $\nabla h_i(x_k) \in S(x_k, \varepsilon, \mu, \eta)$, with $k > N_1$, and from Corollary 4.3 that $(\nabla h_i(x_k), g_k) \leqq \Psi(x_k, \varepsilon, \mu, \eta) \leqq -\gamma$, gives, from (4.8),

$$h_i(x_k + td_k) \leqq -\gamma t + \gamma t/2 < 0. \tag{4.9}$$

Consider any $i \notin Q(x_k, \mu)$. By uniform continuity of the functions h_i on the compact set C, there exists some $T''_o > 0$ such that for all x, y in C, $||x - y|| \leqq T''_o$ implies

$$|h_i(x) - h_i(y)| < \mu \text{ for all } i = 1, \ldots, K. \tag{4.10}$$

For any $i \notin Q(x_k, \mu)$, $h_i(x_k) < -\mu$ and so from (4.10), with $t \in [0, T''_o]$,

$$h_i(x_k + td_k) \leqq h_i(x_k) + \mu < -\mu + \mu = 0. \tag{4.11}$$

Letting $T = \min\{T'_o, T''_o\} > 0$, from (4.9) and (4.11) and the definition of H,

$$H(x_k + td_k) \leqq 0 \;,\; t \in [0, T].$$

The second conclusion of the Lemma follows since Step 2 of the algorithm determines x_{k+1} where

$$F(x_{k+1}) = \min\{F(x_k + td_k) \; ; \; t \geqq 0,\; H(x_k + td_k) \leqq 0\}$$
$$\leqq \min\{F(x_k + td_k) \; ; \; t \in [0, T)\}.$$

In the convergence proof, we make use of the following result from Cullum, Donath and Wolfe [3].

Lemma 4.6 Let F be convex on R^n and let the sequences $\{x_k\}$ and $\{d_k\}$ satisfy $x_k \to x_*$, $d_k \to d_*$, and $F(x_{k+1}) \leqq F(x_k + td_k)$, $0 \leqq t \leqq T$. Then $F'(x_*, d_*) \geqq 0$.

We now prove the main result of this section that the limit point of any convergent sequence generated by the algorithm is a stationary point.

Theorem 4.7 $\Psi(x_*, \varepsilon, \mu, \eta) = 0$.

Proof: Suppose, to the contrary, that $\Psi(x_*, \varepsilon, \mu, \eta) < -2\gamma < 0$. Since $d_k \to d_* \varepsilon B$ there exists N_3 such that

$$\ell\delta ||d_k - d_*|| < \frac{\gamma}{4}, \ k > N_3. \qquad (4.12)$$

The directional derivative of F at x_* is

$$F'(x_*, d_*) = \text{Max}\{(g, d_*) \ ; \ g \ \varepsilon \ \partial F(x_*)\} = (\Sigma\lambda_j s_j, d_*),$$

where $\Sigma\lambda_j s_j$ is a convex combination of elements s_j and each $s_j \ \varepsilon \ \partial f_i(x_*)$ for some $i \ \varepsilon \ R(x_*, 0)$. Choose a single N_2 so large that Lemma 4.4 holds for all such s_j with $\alpha = \frac{\gamma}{4}$. Let $k > \text{Max}\{N_1, N_2, N_3\}$, where N_1 is from Lemma 4.2. Employing Lemma 4.4,

$$\begin{aligned} F'(x_*, d_*) &= (\Sigma\lambda_j s_j', d_*) + (\Sigma\lambda_j t_j, d_*) \\ &= (\Sigma\lambda_j s_j', d_k) + (\Sigma\lambda_j s_j', d_* - d_k) + (\Sigma\lambda_j t_j, d_*) \\ &\leqq (\Sigma\lambda_j s_j', d_k) + (\Sigma\lambda_j s_j', d_* - d_k) + ||\Sigma\lambda_j t_j|| \ ||d_*|| \\ &\leqq (\Sigma\lambda_j s_j', d_k) + (\Sigma\lambda_j s_j', d_* - d_k) + \frac{\gamma}{4}. \end{aligned}$$

By definition of d_k and from Corollary 4.3,

$$\begin{aligned} (\Sigma\lambda_j s_j', d_k) &\leqq \text{Max}\{(g, d_k) \ ; \ g \ \varepsilon \ S(x_k, \varepsilon, \mu, \eta)\} \\ &= \Psi(x_k, \varepsilon, \mu, \eta) \leqq -\gamma \ . \end{aligned}$$

Further, $||s_j'|| \leqq \ell\delta$ and from (4.12),

$$(\Sigma\lambda_j s_j', d_* - d_k) \leqq ||\Sigma\lambda_j s_j'|| \ ||d_* - d_k|| \leqq \frac{\gamma}{4}.$$

Combining these results gives

$$F'(x_*, d_*) \leqq \frac{-\gamma}{2} < 0 \ ,$$

which contradicts Lemma 4.6 (with T as in Lemma 4.5). Therefore, $\Psi(x_*, \varepsilon, \mu, \eta) = 0$.

When 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 be reduced to $R(x, 0)$ through roundoff error and this, in turn, could affect the convergence of the algorithm.

To avoid this numerical problem, redefine $R(x, \varepsilon)$, as is done in [1], 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 assume $F^* > 0$ and to modify the definition of $S^1(x, \varepsilon, \eta)$ given in Section 1. Then the results of Sections 2 and 4 hold with the exception that the inequalities (2.1) are modified to be $(1 - \varepsilon)F(x) - 2\ell\eta\delta \leq F^* \leq F(x)$.

5. Computational Results and a Non Convergent Example

In this section, numerical results are presented for several constrained minisum location problems and a constrained minimax location problem. To find the point of minimum norm in $S(x, \varepsilon, \mu, \eta)$, Wolfe's algorithm [18] was used for minimax problems, and Gilbert's algorithm [6] was used for minisum problems. The line search was done with quadratic fits, but was modified so as to yield a feasible point at each iteration. All the programs were written in FORTRAN and run on an IBM 370/165.

The constrained minisum problems are from [11]. Three new facilities are to be located in the plane relative to five existing facilities. There is a single linear constraint on the location of the third new facility, (x_{31}, x_{32}), namely, $x_{31} + x_{32} - 3 \leqq 0$.

The problem to solve is

$$\text{Min } F(X_1, X_2, X_3) = \Sigma_{r=1}^{3} \Sigma_{s=1}^{5} w_{rs} \ ||X_r - A_s||_p$$
$$+ \Sigma_{1 \leqq r < t \leqq 3} v_{rt} \ ||X_r - X_t||_p$$

subject to:

$$x_{31} + x_{32} - 3 \leqq 0,$$

where $X_r = (x_{r1}, x_{r2})$ is the location of the r^{th} new facility; $A_s = (a_{s1}, a_{s2})$ is the fixed location of the s^{th} existing facility, and w_{rs} and v_{rt} are known positive weights. Note that the function F is the sum of ($\ell = 18$) LFS functions and that the single constraint is continuously differentiable.

From [11], $A_1 = (2, 3)$, $A_2 = (4, 2)$, $A_3 = (5, 4)$, $A_4 = (3, 5)$ and $A_5 = (6, 7)$, and the w_{rs} and v_{rt} data are as in Table 1. The problem was solved with three different values for p (p = 1, 1.78, and 2). In all cases the starting point was $X_1 = X_2 = X_3 = (0, 0)$ and μ was 10^{-5}. For the problem with p = 1, $\eta = 10^{-5}$, and for the problems with p = 1.78 and p = 2, both $\eta = 10^{-3}$ and $\eta = 10^{-5}$ were used. A summary of the computational results is given in Table 2, where the upper and lower bounds on F* are as given in [11].

We remark that for the problems with $\eta = 10^{-5}$ and p = 1.78 and p = 2, termination occured when the maximum number of iterations allowed (150) was reached, but progress was still possible. For the other problems, the algorithm reached stationarity. The comparison with the results of [11] is rather difficult to make, since no numbers of iterations or computing times are reported in that reference.

As a constrained minimax location problem, the Caribbean Islands problem formulated in [1], was modified to include the four constraints $||X_1 - A_1||^2 \leqq 144$, $||X_1 - A_4||^2 \leqq 121$, $||X_2 - A_1||^2 \leqq 225$, and

$||X_2 - A_2||^2 \leqq 144$. The points A_i, i = 1, 2, 3 and 4 represent the locations of four of the Caribbean cities in the problem (A_1 = (11.4, 11.6), A_2 = (35.3, 13.5), A_3 = (8.80, 37.2), and A_4 = (20.9, 30.6)).

Parameters were given values $\varepsilon = 5 \times 10^{-6}$, $\eta = 10^{-5}$ and $\mu = 10^{-5}$. Starting from $(X_1, X_2) = (x_{11}, x_{12}, x_{21}, x_{22})$ = (15, 22, 26, 11), the algorithm terminated after 15 iterations and 3.03 seconds CPU time at the stationary point (13.902, 23.201, 24.546, 18.824) with a function value of 28.024. The value of H at the stationary point was -1.07×10^{-14}.

With a fifth constraint added, $||X_1 - X_2||^2 \leqq 118$, and starting from (19, 20, 24, 16), the answer (14.585, 23.170, 24.488, 18.705) with F = 28.158 and H = 0 was obtained in 24 iterations and 2.36 seconds of CPU time.

We now give an example which shows that if the constraints of problem (P) *are not continuously* differentiable, the sequence generated by the algorithm may converge to a nonstationary point. Consider the problem:

$$\min_{X \in R^3} F(X), \quad F(X) \equiv -2x_2 + x_3$$

subject to

$$h_1(X) \equiv \text{Max}\{3x_1 + x_2 - 2x_3, -3x_1 + x_2 - 2x_3\} \leqq 0$$

$$h_2(X) \equiv x_3 - 1 \leqq 0 ,$$

which has the optimal solution $X^* = (0, 2, 1)$ with $F(X^*) = -3$. The functions $F(x)$ and $h_2(X)$ are continuously differentiable on R^3, but $h_1(X)$ is not. In fact $h_1(X)$ is LFS with possible subgradients sets of {(3, 1, -2)}, {(-3, 1, -2)} or Conv({(3, 1, -2), (-3, 1, -2)}). To simplify notation, define $\mathcal{S}_1$ = Conv({(0, -2, 1), (3, 1, -2)}), and $\mathcal{S}_2$ = Conv({(0, -2, 1), (-3, 1, -2)}). It is easy to verify that $Nr(\mathcal{S}_1)$ = (1, -1, 0) and $Nr(\mathcal{S}_2)$ = (-1, -1, 0). Since F is differentiable, set $\eta = 0$.

Also it is clear that $\varepsilon = 0$ is a legitimate choice. Further, let $\mu = .5$ and let $X_0 = (1, -3, 0)$ be the starting point for the algorithm, where $F(X_0) = 6$.

Since $h_1(X_0) = \text{Max}\{0, -6\} = 0$ and $h_2(X_0) = -1$, $Q(X_0, \mu) = \{1\}$, and $S(X_0, \varepsilon, \mu, \eta) = \mathcal{S}_1$. Thus the feasible descent direction given by the algorithm is $-\text{Nr}(\mathcal{S}_1) = (-1, 1, 0)$. To find X_1, minimize $F(X_0 + t(-1, 1, 0)) = -2t + 6$, for $t \geqq 0$ and $(X_0 + t(-1, 1, 0))$ feasible. The minimum in the feasible set occurs for $t = 3/2$ which gives $X_1 = (-1/2, -3/2, 0)$, $F(X_1) = 3$.

At X_1, $h_1(X_1) = \text{Max}\{-3, 0\}$, $h_2(X_1) = -1$ and $S(X_1, \varepsilon, \mu, \eta) = \mathcal{S}_2$. With $-\text{Nr}(\mathcal{S}_2) = (1, 1, 0)$, minimize $F(X_1 + t(1, 1, 0))$ over $t \geqq 0$ and $(X_1 + t(1, 1, 0))$ feasible, obtaining $t = 3/4$ and $X_2 = (1/4, -3/4, 0)$ with $F(X_2) = 3/2$.

X_2 yields $S(X_2, \varepsilon, \mu, \eta) = \mathcal{S}_1$ and the minimization in the direction $-\text{Nr}(\mathcal{S}_1)$ gives $X_3 = (-1/8, -3/8, 0)$, $F(X_3) = 3/4$. Continuing, it can be shown, using inductive arguments, that for any k,

$$X_k = ((-1)^k/2^k\ ,\ - 3/2^k\ ,\ 0) \text{ and } F(X_k) = 3/2^{k-1}.$$

Consequently, the sequence $\{X_k\}$ converges to $X_* = (0, 0, 0)$, while $F(X_k) \to F(X_*) = 0$. However, X_* is <u>not</u> a stationary point. This can be seen from the fact that $h_1(X_*) = 0$, $h_2(X_*) = -1$, yielding $Q(X_*, \mu) = \{1\}$, and $S(X_*, \varepsilon, \mu, \eta) = \text{Conv}(\{(0, -2, 1), (3, 1, 2), (-3, 1, -2)\})$, where $0 \notin S(X_*, \varepsilon, \mu, \eta)$. Further, $-\text{Nr}(S(X_*, \varepsilon, \mu, \eta)) = (0, 1/2, 1/2)$ which is a feasible descent direction for F at X_*.

r \ s	1	2	3	4	5
1	1	1	6	1	6
2	4	1	1	1	1
3	1	1	1	1	1

w_{rs}

r \ t	1	2	3
1	-	1	1
2	-	-	1
3	-	-	-

v_{rt}, $r < t$

Table 1: Weights for the Minisum Problem

p	1	1.78		2	
η	10^{-5}	10^{-3}	10^{-5}	10^{-3}	10^{-5}
Number of iterations	63	90	150	134	150
CPU time	6.62s	14.39s	14.72s	16.24s	13.77s
Lower bound on F*	90	70.25804		68.18406	
Value obtained	90.00008	70.27462	70.28511	68.23939	68.28341
Upper bound on F*	90	70.4299		68.64550	
Constraint value	-8.17×10^{-6}	0	-1.23×10^{-5}	0	0

Table 2. Numerical results for the constrained minisum location problem.

References

[1] Chatelon, J. A., Hearn, D. W., and Lowe, T. J., "A Subgradient Algorithm for Certain Minimax and Minisum problems," Math. Prog., to appear.

[2] Clarke, F. H., "Generalized Gradients and Applications," Trans. Amer. Math. Soc.; 205, 247-262, 1975.

[3] Cullum, J., Donath, W. E., and Wolfe, P., "The Minimization of Certain Nondifferentiable Sums of Eigenvalues of Symmetric Matrices," Math. Prog. Study 3, 35-55, 1975.

[4] Bertsekas, D. P., and Mitter, S. K., "A Descent Numerical Method for Optimization Problems with Nondifferentiable Cost Functionals," SIAM J. on Control, 11, 637-652, 1973.

[5] Dem'yanov, V. F., and Malozemov, V. N., Introduction to Minimax, John Wiley, 1974.

[6] Gilbert, E. G. "An Iterative Procedure for Computing the Minimum of a Quadratic Form on a Convex Set," SIAM J. on Control 4, 61-80, 1966.

[7] Goldstein, A. A., "Optimization of Lipschitz Continuous Functions," Math. Prog. 13, 14-22, 1977.

[8] Held, M. and Karp, R. M., "The Traveling Salesman Problem and Minimum Spanning Trees: Part II," Math. Prog. 1, 6-25, 1971.

[9] Held, M. Wolfe, P., and Crowder, H. P., "Validation of Subgradient Optimization," Math. Prog. 6, 62-88, 1974.

[10] Lemarechal, C., "An Extension of Davidon Methods to Non-differentiable Functions," Math. Prog. Study 3, 95-109, 1975.

[11] Love, R. F., "The Dual of a Hyperbolic Approximation to the Generalized Constrained Multifacility Location Problem with ℓ_p Distances," Manag. Sci. 21, 22-33, 1974.

[12] Mangasarian, O. L., Nonlinear Programming, McGraw-Hill, New York, 1969.

[13] Mifflin, R., "An Algorithm for Constrained Optimization with Semismooth Functions," Math. of O.R., 2, 191-207, May 1977.

[14] Mifflin, R., "Semismooth and Semiconvex Functions in Constrained Optimization," SIAM J. on Control and Optimization, 1977.

[15] Polyak, B. T., A General Method of Solving Extremum Problems," Soviet Math. Doklady 7, 72-75, 1966.

[16] Rockafellar, R. T., Convex Analysis, Princeton University Press, Princeton, N. J., 1970.

[17] Wolfe, P. "A Method of Conjugate Subgradients for Minimizing Non-differentiable Functions," Math. Prog. Study 3, 145-173, 1975.

[18] Wolfe, P., "Finding the Nearest Point in a Polytope," Math. Prog. 11, 128-144, 1976.