Discussion Paper No. 308

"The Fixed Point Approach to Nonlinear Programming"*

by

Romesh Saigal

December, 1977

*This research was partially supported by Grant No. MCS 77-03472 from the National Science Foundation.

The Fixed Point Approach to Nonlinear Programming

R. Saigal

Northwestern University

Evanston, Illinois

ABSTRACT

In this paper we consider the application of the recent algorithms that compute fixed points in unbounded regions to the nonlinear programming problem. It is shown that these algorithms solve the inequality constrained problem with functions that are not necessarily differentiable. The application to convex and piecewise linear problems is also discussed.

THE FIXED POINT APPROACH TO

NONLINEAR PROGRAMMING

R. Saigal

1. Introduction

In this paper we consider the problem

$$\min g_0(x) \tag{1.1}$$

$$g_i(x) \leqq 0 \qquad i = 1, \ldots, m \tag{1.2}$$

where g_0 and g_i are arbitrary functions on a set $X \subset R^n$, the n-dimensional Euclidean space. The set X may be discrete, but for simplification, we assume that the convex hull of X is R^n, and that there exists a subdivision of R^n with vertices in X. One such example of X is the grid of integers, for which efficient triangulation procedures exist. See Todd [16].

Our approach is to consider piecewise linear approximations g_i^{ℓ}, $i = 0, \ldots, m$ instead, and then to solve the continuous problem by the fixed point algorithms. We thus obtain an approximate solution to (1.1-2). Such an approach has been successfully used in [8]. We note that g_i^{ℓ} are not differentiable. Since they are piecewise linear, a notion of a generalized subdifferential can be readily defined. This is the same as the generalized gradient of Clarke [1], and for convex g_i^{ℓ}, the same as the subdifferential of convex functions, Rockafellar [9]. Since the fixed point algorithms of Eaves and Saigal [3] and Merrill [6] can successfully find fixed points of certain point-to-set mappings, we formulate this problem as such a point-to-set mapping problem which can then be solved by these algorithms.

Hansen [4], Hansen and Scarf [5], and Eaves [2] had recognized the potential of applying these methods to nonlinear programming, but the full potential was explored by Merrill [6]. In section 3, we present extensions of several of his results for the convex case. Traditional descent type methods for solving this problem are summarized in Mifflin [7]. Also, in [7], a steepest descent type algorithm, using the mapping of [6], is presented.

In section 2, we present a brief overview of the fixed point algorithms; in section 3 these algorithms are applied to the constrained and unconstrained convex problems; in section 4 we introduce piecewise linear mappings and establish the necessary and sufficient conditions for local minimization of the constrained and the unconstrained problems; in section 5 the application of the algorithms is discussed; and in section 6 we discuss the computational aspects. Finally, in the appendix, we present the computational experience of solving some fairly large nondifferentiable problems.

2. The Algorithms

We now give a brief description of the algorithm of Eaves and Saigal [3] implemented on the subdivision J3 of $R^n x\ (0,D]$. We will assume that the nonlinear programming problem is being solved by this algorithm.

The triangulation J3 of $R^n x(0,D]$ has vertices in $R^n x\{D.2^{-k}\}$ for $k = 0, 1, 2, \ldots$. Also, $v = (v_1, \ldots, V_{n+1})$ is a vertex if $v_{n+1} = D.2^{-k}$ for some integer k and v_i/v_{n+1} is an integer for each i. In case, for a vertex v, v_i/v_{n+1} is an odd integer, it is called a central vertex. Any simplex in J3 then has a unique representation by a triplet (v,π,s) where v is a central vertex, π is a permutation of $\{1, \ldots, n+1\}$ and s is an n-vector with $s_i \in \{-1, +1\}$. A complete description of J3 can be found in Saigal [11], and Todd [16].

Given a point to set mapping ℓ from R^n into nonempty subsets of R^n, and a 1-1 linear mapping r from R^n into R^n, we say a n-simplex $\sigma = (v^1, \ldots, v^{n+1})$ is

(a) r - complete if $0 \in \text{hull}\{ r(\sigma) \}$,

(b) $\ell \cup r$ - complete if $0 \in \text{hull}\ \{r(\sigma) \cup \ell(\sigma)\}$,

(c) ℓ - complete if $0 \in \text{hull}\ \{\ell(\sigma)\}$.

These algorithms, starting with a unique r-complete simplex σ_0 containing the unique zero of r, generate a sequence of $\ell \cup r$- complete simplexes $\sigma_o, \sigma_1, \sigma_2, \ldots, \sigma_k, \ldots$. In case these simplexes lie in a bounded region, it can be readily shown that if they are from J3, there is a subsequence $\sigma_{i_1}, \sigma_{i_2}, \sigma_{i_3}, \ldots, \sigma_{i_k}, \ldots$, of ℓ- complete simplexes such that $\sigma_{i_k} \subset R^n x\ \{D.2^{-k}\}$ and, so the diameter of σ_{i_k} approaches zero as k approaches ∞.

Thus, the failure of the algorithm implies that it has generated $\ell \cup r$- complete simplexes sufficiently far from x_0, the unique zero of r.

In applications to nonlinear programming, we will freqently choose

$$r(x) = x - x_0$$

and define a point-to-set mapping ℓ such that if, for some x, $0 \varepsilon \ell(x)$, then x is a solution to our problem.

We now prove the following result:

Theorem 2.1: Let $\sigma_1, \sigma_2, \ldots, \sigma_k, \ldots$ be a sequence of ℓ- complete simplexes which lie in a bounded set and the diamater ε_k of σ_k approaches 0. In addition, let ℓ be a upper semi-continuous point-to-set mapping with $\ell(x)$ nonempty, compact and convex subsets of R^n. Then, if x is a cluster point of $\{x^k\}^\infty_{k=0}$ with $x_k \varepsilon \sigma_k$, then $0 \varepsilon \ell(x)$.

Proof:

Since $\{\sigma_k\}^\infty_{k=0}$ lie in a bounded region, say B, under the hypothesis on ℓ, $\ell(\overline{B})$ is compact, hence $\ell(B)$ is bounded ($\overline{B}$ is the closure of B). Now, as $\sigma_k = (v^1{}_k, v^2{}_k, \ldots, v^{n+1}{}_k)$ is ℓ- complete, there exist $y_{i,k} \varepsilon \ell(v^i{}_k)$ and $\lambda_{i,k} \geqq 0$, $\sum_i \lambda_{i,k} = 1$ such that $\sum_i \lambda_{i,k} y^*_{i,k} = 0$. Since $0 \leqq \lambda_i, k \leqq 1$, $y^*_{i,k} \varepsilon \ell(B)$ on some common subsequence $\lambda_{i,k} \to \lambda_i$ for all i and $y^*_{i,k} \to y^*_i$ for all i. Thus

$$\sum_i \lambda_i y^*_i = 0$$

$$\Sigma \lambda_i = 1$$

and $\lambda_i \geqq 0$.

Also, as $\mathrm{dia}(\sigma^k)$ approaches 0, on some subsequence $v^i{}_k \to x$ for all i. Since $y^*_{i,k} \varepsilon \ell(v^i{}_k)$, using the upper semi-continuity of ℓ we have $y^*_i \varepsilon \ell(x)$, and since $\ell(x)$ is convex, we have our result.

3. Convex Case

In this section, we will consider the applications of fixed point algorithms for solving (1.1-2) when the underlying functions are convex, not necessarily differentiable. We will make the simplifying assumption that the functions are defined over all of R^n, and that they are finite.

3.1 Unconstrained Case

We now consider the problem of minimizing g_0 when the set $\{x : g_0(x) \leqq g_0(x_0)\}$ is bounded for some x_0, and the function g_0 is convex.

The subdifferential set $\partial g_0(x)$ of a convex function g_0 at x is the set of all vectors x* in R^n such that

$$g_0(y) \geqq g_0(x) + \langle x^*, y-x\rangle \text{ for all } y \text{ in } R^n \qquad (3.1)$$

and under our assumption this set is nonempty, closed and bounded, [9, Theorem 23.4].

A trivial consequence of (3.1) is the following theorem.

Theorem 3.1: Under the above conditions on g_0, $\bar{x}$ solves (1.1) if and only if $0 \in \partial g_0(\bar{x})$.

We now show that the algorithms of Section 2 implemented with

$$r(x) = x - x_0$$

$$\ell(x) = \partial g_0(x)$$

for an arbitrary starting point x_0 and initial grid size ε_0 will converge to a solution of (1.1). Let $B(x,\varepsilon) = \{y : ||y-x||<\varepsilon\}$, and $M(x,\varepsilon) = \sup \{g_0(y) : y\varepsilon B(x,\varepsilon)\}$.

Theorem 3.2: Starting with the unique r- complete simplex containing x_0, the fixed point algorithms will succeed in generating ℓ- complete simplexes of diameters $\varepsilon \leqq \varepsilon_0$.

Proof:

Assume that the algorithm does not compute a ℓ- complete simplex of diameter $\varepsilon < \varepsilon_0$. Now, define

$$D = \{x : g_0(x) \leqq M(x_0, \varepsilon_0)\}$$

Since D is bounded, we can find a $\ell \cup r$- complete simplex $\sigma = (v^1, \ldots, v^{n+1})$ sufficiently far from x_0 such that for some $x \in \sigma$, $\langle v^i - x_0, x - x_0 \rangle > 0$, and such that $v^i \notin D$ for each $i = 1, \ldots, n+1$. Now, as g_0 is convex, by (3.1) for every $y^* \varepsilon g_0(v^i)$ and every i

$$g_0(v^i - x + x_0) \geqq g_0(v^i) + \langle x_0 - x, y^* \rangle$$

Since $v^i - x + x_0 \in D$, we have $\langle x_0 - x, y^* \rangle > 0$ for all $y^* \varepsilon g_0(v^i)$ and every i. Hence, using Farkas lemma we claim that σ cannot be $\ell \cup r$- complete, and we have a contradiction.

Now, let $\sigma = (v^1, \ldots, v^{n+1})$ be a ℓ- complete simplex of diameter $\varepsilon > 0$. Then there exist $y^*_i \varepsilon \partial g_0(v^i)$, $\lambda_i \geqq 0$, $i = 1, \ldots, n+1$, $\Sigma \lambda_i = 1$ such that $\Sigma \lambda_i y^*_i = 0$. We can then prove that:

Theorem 3.3: Let $\bar{x}$ be a solution to (3.1). Then there is an $x \in \sigma$ such that

$$g_0(\bar{x}) \geqq g_0(x) - \sum_1^{n+1} \lambda_i \langle v^i, y^*_i \rangle \tag{3.2}$$

and $M(\bar{x}, \varepsilon) \geqq g_0(x)$. (3.3)

Proof:

Let $x = \Sigma \lambda_i v^i$. From (3.1)

$$g_0(\bar{x}) \geqq g_0(v^i) + \langle \bar{x} - v^i, y^*_i \rangle$$

Hence

$$g_0(\bar{x}) \geqq \sum_{i=1}^{n+1} \lambda_i g_0(v^i) - \Sigma\lambda_i <v^i, y^*_i>$$

$$\geqq g_0(x) - \Sigma\lambda_i <v^i, y^*_i>$$

and the first part follows. Also,

$$g_0(v^i - x + \bar{x}) \geqq g_0(v^i) + <\bar{x} - x, y^*_i>$$

and so

$$\Sigma\lambda_i g_0(v^i - x + \bar{x}) \geqq \Sigma\lambda_i g_0(v^i) \quad \geqq g_0(x).$$

But, as $v^i - x + \bar{x} \in B(\bar{x}, \varepsilon)$, we have our result.

Note that in Theorem 3.3, (3.2) gives a computable lower bound on the minimum value of g(x) and can thus be used as a stopping rule. Also, (3.3) shows that the algorithm is converging to a minimum.

3.2 Constrained Case

We now consider the problem (1.1-2) when the functions g_i are convex functions. Define,

$$s(x) = \max \{g_i(x) : 1 \leqq i \leqq m\}$$

and we note the s is also a convex function. We now assume that the set $\{x : s(x) \leqq s(x_0)\}$ is bounded for some x_0. Now, define the mapping

$$\ell(x) = \begin{cases} \partial g_0(x) & \text{if } s(x) < 0 \\ \partial g_0(x) + \partial s(x) & \text{if } s(x) = 0 \\ \partial s(x) & \text{if } s(x) > 0 \end{cases} \tag{3.4}$$

Theorem 3.4: Let $\bar{x}$ be such that $0 \in \ell(\bar{x})$. Then $\bar{x}$ solves (1.1-2) or indicates that (1.2) has no solution.

Proof:

There are three cases.

Case (i). $s(\bar{x}) > 0$. In this case $0 \in \partial s(\bar{x})$, and thus $\bar{x}$ is a global minimizer of s, and hence the constraint set (1.2) is empty.

Case (ii). $s(\bar{x}) < 0$. In this case $0 \in \partial g_0(\bar{x})$ and hence $\bar{x}$ is a global minimizer of g_0. Since it also satisfies (1.2), $\bar{x}$ solves (1.1-2).

Case (iii). $s(\bar{x}) = 0$. In this case, there is a $z^* \in \partial g_0(\bar{x})$ and $y^* \in \partial s(\bar{x})$ such that $z^* + y^* = 0$. Let $I(\bar{x}) = \{i: g_i(\bar{x}) = 0\}$. Then, $\partial s(\bar{x}) = \text{hull}\{\bigcup_{i\in I(\bar{x})} \partial g_i(\bar{x})\}$ and so there are numbers $\lambda_i \geqq 0$, $i \in I(\bar{x})$, $\Sigma\lambda_i = 1$, and $y^*_i \in \partial g_i(\bar{x})$ such that $y^* = \Sigma\lambda_i y^*_i$. Hence, $z^* + \Sigma\lambda_i y^*_i = 0$. Now, let y satisfy (1.2). Then, from (3.1),

$$g_0(y) \geqq g_0(\bar{x}) + \langle z^*, y-\bar{x}\rangle$$

$$g_i(y) \geqq g_i(\bar{x}) + \langle y^*_i, y-\bar{x}\rangle \qquad i \in I(\bar{x})$$

Hence

$$\begin{aligned} g_0(y) &\geqq g_0(y) + \Sigma\lambda_i g_i(y) \\ &= g_0(\bar{x}) + \langle z^* + \Sigma\lambda_i y^*_i, y-\bar{x}\rangle \\ &= g_0(\bar{x}) \end{aligned}$$

and hence $\bar{x}$ solves (1.1-2).

We now show that the algorithm initiated with

$$r(x) = x - x_0$$

for arbitrary x_0 and $\ell(x)$ as in (3.4) will find a ℓ- complete simplex.

Theorem 3.5: Let $\varepsilon > 0$ be arbitrary, and let the algorithm implement the mapping r above. Then, for each $\varepsilon > 0$, the algorithm will find a ℓ-complete simplex of diameter ε.

Proof:

Let

$$M(x_0,\varepsilon) = \max\{\sup\{ s(x) : x \in B(x_0,\varepsilon)\}, 0\}$$

and $D = \{x : s(x) \leqq M(x_0,\varepsilon)\}$. By assumption, D is bounded. Now, assume that the algorithm fails. Hence, it generates a simplex $\sigma = (v^1, \ldots, v^{n+1})$ of diameter $\varepsilon > 0$ such that σ is $\ell \cup r$- complete, and sufficiently far from D, i.e., $\sigma \not\subset D$ and for every $x \in \sigma$, $\langle v^i - x_0, x-x_0 \rangle > > 0$. Also $s(v^i) > 0$ for all i. Now, consider the point $v^i - x + x_0 \varepsilon D$. Then

$$s(v^i - x + x_0) \geqq s(v^i) + \langle x_0 - x, y^*_i \rangle \quad \text{for all } y^*_i \ \varepsilon \partial s(v^i).$$

Since $s(v^i) \notin D$, we have $\langle x-x_0, y^*_i \rangle > > 0$ for all $y^*_i \varepsilon \partial s(v^i)$ and all i; and, from Farkas' lemma, σ cannot be $\ell \cup r$- complete, a contradiction.

We now assume that there is no solution to (1.2). Hence $s(x) > 0$ for all x, and thus (3.4) reduces to $\ell(x) = \partial s(x)$. A consequence of Theorem 3.3 is the following.

Theorem 3.6: Let $s(x) > 0$ for all x, and that $\{x : s(x) \leqq s(x_0)\}$ is bounded for some x_0. Then, the algorithm will detect the infeasibility of (1.2) in a finite number of iterations.

Proof:

For each $\varepsilon > 0$, the algorithms compute a $\ell \cup r$- complete simplex in a finite number of iterations. Also, since $s(x) > 0$, it will attempt to minimize $s(x)$. Now, let $\sigma = (v^1, \ldots, v^{n+1})$ be an ℓ- complete simplex of size $\varepsilon > 0$ found by the algorithm. Then, there are $y^*_i \ \varepsilon \ \partial \ s(v^i)$ such that $\Sigma\lambda_i y^*_i = 0$, $\Sigma\lambda_i = 1$, $\lambda_i = 0$ has a solution. Also, from Theorem 3.3, if $\bar{x}$ minimizes s,

$$\begin{aligned} s(\bar{x}) &\geqq s(x) + \Sigma\lambda_i \langle v^i, y^*_i \rangle \\ &= s(x) + \Sigma\lambda_i \langle v^i - x, y^*_i \rangle \end{aligned}$$

Now, define $D = \{x : s(x) \leq M(\bar{x}, \varepsilon)\}$

when $M(\bar{x}, \varepsilon) = \sup \{s(x) : x \in B(\bar{x}, \varepsilon)\}$. From Theorem 3.3, $x \in D$ and $\sigma \subset B(D, \varepsilon)$. Define $N \geq ||y^*||$ for $y^* \in \partial s(x)$, $x \in B(D, \varepsilon)$. Then

$$||\Sigma\lambda_i \langle v^i - x, y^*_i\rangle|| \leq N\varepsilon$$

as $\varepsilon \to 0$, and $s(x) > 0$, for some sufficiently small $\varepsilon > 0$, $(s(x) + \Sigma\lambda_i \langle v^i, y^*_i\rangle > 0)$, and hence we are done.

In addition, we can obtain a lower bound on the optimal value of the objective function in this case as well. Let $\sigma = (v^1, \ldots, v^{n+1})$ be δ- complete and let $v^1, \ldots, v^r$ be labeled by $y^*_i \in \partial g_0(v^i)$ and $v^{r+1}, \ldots, v^{n+1}$ be labeled by $y^*_i \in \partial s(v^i)$. Hence $s(v^i) < 0$, $i = 1, \ldots, r$ and $s(v^i) \geq 0$ for $i = r+1, \ldots, n+1$. Also, let $\Sigma\lambda_i y^*_i = 0$, $\Sigma\lambda_i = 1$, $\lambda_i \geq 0$, $\theta = \sum_1^r \lambda_i$ and $\hat{x} = \frac{1}{\theta}\sum_1^r \lambda_i v^i$. Then we can prove:

<u>Theorem 3.7</u>: Let $\bar{x}$ solve (1.1-2), then

$$g_0(\bar{x}) \geq g_0(\hat{x}) - \frac{1}{\theta}\Sigma\lambda_i\langle v^i, y^*_i\rangle.$$

<u>Proof</u>:

Using (3.1) we get, for $i = 1, \ldots, r$

$$g_0(\bar{x}) \geq g_0(v^i) + \langle \bar{x} - v^i, y^*_i\rangle$$

and for $i = r+1, \ldots, n+1$, we get

$$s(\bar{x}) \geq s(v^i) + \langle \bar{x} - v^i, y^*_i\rangle$$

Hence, as $s(\bar{x}) \leq 0$

$$g_0(\bar{x}) \geq \frac{1}{\theta}\sum_1^r \lambda_i g_0(\bar{x}) + \frac{1}{\theta}\sum_{r+1}^{n+1} \lambda_i s(\bar{x})$$

$$\geq \frac{1}{\theta}\sum_1^r \lambda_i g_0(v^i) + \frac{1}{\theta}\sum_{r+1}^{n+1} \lambda_i s(v^i) + \frac{1}{\theta}\sum_1^{n+1} \lambda_i \langle \bar{x} - v^i, y^*_i\rangle$$

$$\geq g_0(\hat{x}) - \frac{1}{\theta}\sum_1^{n+1} \lambda_i \langle v^i, y^*_i\rangle$$

and we have our result.

4. Piecewise Linear Functions and Nonlinear Programming

In this section we establish the notation and prove some basic results for nonlinear programs with piecewise linear functions.

Cells and Manifolds

A cell is the convex hull of a finite number of points and half lines (half lines are sets of the type $\{x : x = a + tb,\ t \geqq 0\}$ where a and b are fixed vectors in R^n).

The dimension of a cell is the maximum number of linearly independent points in the cell. We will call an n dimensional cell an n-cell.

Let τ be a subset of an n-cell σ. If $x, y \in \sigma$ $0 < \lambda < 1$, $(1-\lambda)x + \lambda y \in \tau$ implies that x,y in τ then τ is called a face of a cell σ. A simple fact is that faces are cells. Also faces that are (n-1)- cells are called facets of the cell, and that are 0-cells are called vertices of the cell.

$\phi \neq \mathcal{m}$ be a collection of n-cells in R^n. Let $M = \bigcup_{\sigma \in \mathcal{m}} \sigma$.

We call $(M,\mathcal{m})$ a subdivided n- manifold if

(4.1) Any two n-cells of $\mathcal{m}$ that meet, do so on a common face.

(4.2) Each (n-1)- face of a cell lies in at most two n-cells.

(4.3) Each x in M lies in a finite number of n-cells in $\mathcal{m}$.

If $(M,\mathcal{m})$ is a subdivided n- manifold for some $\mathcal{m}$, we call M a n- manifold.

Piecewise Linear Functions

Let M be a n- manifold, then the function

$$g\colon M \to R$$

is called piecewise linear on a subdivision $\mathcal{m}$ of M if

(4.4) g is continuous

(4.5) Given a cell σ in m, there exists an affine function g_σ : $R^n \to R$ such that $g|\sigma(x) = g_\sigma(x)$ (i.e., g restricted to σ is g_σ).

Generalized Subdifferentials

Let M be a n-manifold, and m be its subdivision. Let $g : M \to R$ be a piecewise linear function. Then, for each $x \varepsilon M$, we define a generalized subdifferential set $\partial g(x)$ as follows:

From (4.3), x lies in a finite number of n-cells, $\sigma_1, \sigma_2, \ldots, \sigma_r$ in m say. Let

$$\nabla g_{\sigma_i} = a_i$$

(where ∇f is the gradient vector of f). Then, we define

$$\partial g(x) = \text{hull } \{a_1, \ldots, a_r\}$$

and we note that if g, in addition, is convex, then $\partial g(x)$ is the subdifferential of g at x, Rockafellar [9]; and, as g is locally Lipschitz continuous, $\partial g(x)$ is the generalized gradient of Clarke [1]. In that case, the theorem below is known, but we will use the piecewise linearity of g to establish it.

Theorem 4.1: If x is a local minimum of g_0, then $0 \varepsilon \partial g_0(x)$.

Proof:

Assume $\bar{x}$ is a local minimum but $0 \notin \partial g_0(\bar{x})$. Now, let $\bar{x} \varepsilon \sigma_1 \cap \sigma_2 \cap \ldots \cap \sigma_r$. Then $\partial g_0(\bar{x}) = \text{hull } \{a_1, \ldots, a_r\}$. Hence, from Farkas' lemma, there is a $z \neq 0$ such that $\langle z, a_i \rangle < 0$ for $i = 1, \ldots, r$. Let $\varepsilon > 0$ be sufficiently small so that $B(\bar{x}, \varepsilon) \subset \bigcup \sigma_i$. Then $\bar{x} + \theta z \varepsilon B(\bar{x}, \varepsilon)$ for sufficiently small $\theta > 0$. Assume $\bar{x} + \theta z \varepsilon \sigma_j$ for some j_j. Hence

$$\begin{aligned} g_0(\bar{x} + \theta z) &= \langle a_j, \bar{x} \rangle + \theta \langle a_j, z \rangle - \gamma_i \\ &= g_0(\bar{x}) + \theta \langle a_j, z \rangle < g_0(\bar{x}) \end{aligned}$$

and we have a contradiction to the fact that $\bar{x}$ is a local minimum.

Given a point-to-set mapping Γ from R^n to nonempty subsets of R^n, we say Γ is weakly monotone at $\bar{x}$ with respect to $\bar{y}^* \varepsilon \Gamma(\bar{x})$ on F if there is an $\varepsilon > 0$ such that for all $x \; \varepsilon \; B(\bar{x}, \varepsilon) \cap F$.

$$<x-\bar{x},\ y^* - \bar{y}^*> \;\geqq 0 \text{ for all } y^* \text{ in } \Gamma(x).$$

We can then prove:

<u>Theorem 4.2</u>: $\bar{x}$ is a local minimum of g_0 if and only if $0 \; \varepsilon \; \partial g_0(\bar{x})$ and ∂g_0 is weakly monotone at $\bar{x}$ with respect to 0 on R^n.

<u>Proof</u>:

Let $\bar{x}$ lie in the cells $\sigma_1, \sigma_2, \ldots, \sigma_n$. Then $\partial g_0(\bar{x}) =$ hull $\{a_1, \ldots a_r\}$. For some sufficiently small $\varepsilon > 0$, let $B(\bar{x},\varepsilon) \subset \bigcup \sigma_i$ To see the if part, let $0 \; \varepsilon \; \partial g_0(\bar{x})$ and let $\partial g_0(\bar{x})$ be weakly monotone with respect to zero at $\bar{x}$. Hence, for some $\varepsilon > 0$, for all $x \; \varepsilon \; B(\bar{x},\varepsilon)$ we have

$$<x-\bar{x},\ a_i> \;\geqq 0 \qquad \text{where } a_i \varepsilon \;\; \partial g_0(x) \subset \partial g_0(\bar{x})$$

Hence,

$g_0(x) - g_0(\bar{x}) = \;<a_i,x> - \gamma_i - <a_i,\bar{x}> + \gamma_i \;\geqq 0$, and so $\bar{x}$ is a local minimum of g_0. To see the only if part, let $0 \; \varepsilon \;\; \partial g_0(\bar{x})$ and $\partial g_0(\bar{x})$ not weakly monotone with respect to 0. Then, for a sufficiently small $\varepsilon > 0$ such that $B(\bar{x},\varepsilon) \subset \bigcup \sigma_i$, there is an $x \; \varepsilon \; B(\bar{x},\varepsilon)$ and a $a_i \; \varepsilon \;\; \partial g_0(x)$ such that $<x-\bar{x},\ a_i> \; < 0$. Since $\partial g_0(x) \subset \partial g_0(\bar{x})$, we have

$$\begin{aligned} g_0(x) - g_0(\bar{x}) &= \;<a_i,x> - \gamma_i - <a_i,\bar{x}> + \gamma_i \\ &= \;<x - \bar{x},\ a_i> \\ &< \;0 \end{aligned}$$

which is a contradiction.

The Constrainted Problem

Let $g_i : R^n \to R$ be piecewise linear functions on subdivided manifolds (R^n, m^i), respectively, for each $i = 0, \ldots, m$. We now consider the constrained minimization problem (1.1-2).

For a generic point x in R^n we define $\sigma_1{}^i, \sigma_2{}^i, \ldots, \sigma_{r_i}{}^i$ as the n-cells of m^i in which x lies, and $g_i|\sigma_j{}^i(y) = \langle a_j{}^i, y\rangle - \gamma_j{}^i$, for each $i = 0, \ldots, m$. Also note that, by definition, r_i is finite. Also, there exists $\varepsilon > 0$ such that $B(x,\varepsilon) \subset \bigcup_j \sigma_j{}^i$ for each i. We are now ready to establish the necessary conditions for $\bar{x}$ to be a local minimum of (1.1-2).

Theorem 4.3: Let $\bar{x}$ be a local minimum of g_0 over all x satisfying (1.2). Then

(i) There exists $\lambda_i \geqq 0$ such that

$$\lambda_i g_i(\bar{x}) = 0, \quad i = 1, \ldots, m.$$

(ii) There exists $y^* \in \partial g_0(\bar{x})$, $z^*_i \in \partial g_i(\bar{x})$ such that

$$0 = y^* + \sum_{i=1}^{m} \lambda_i z^*_i$$

Proof:

Let $\bar{x}$ be a local minimum, and $0 \notin \partial g_0(\bar{x}) + \text{cone}(C)$ where $C = \bigcup_{i \in I(\bar{x})} \partial g_i(\bar{x})$, $I(\bar{x}) = \{i : g_i(\bar{x}) = 0\}$ and cone $(C) = \{y : y = \sum_{i=1}^{r} \lambda_i x_i,\ x_i \in C,\ \lambda_i \geqq 0\}$. (It can be readily confirmed that (i) and (ii) hold if and only if $0 \in \partial g_0(\bar{x}) + \text{cone}(C)$). Then, from Farkas lemma, since both $\partial g_0(\bar{x})$ and C are convex combinations of a finite number of vectors, there exists a z such that

$$\langle z, y^*\rangle < 0 \quad \text{for all } y^* \in \partial g_0(\bar{x})$$

$$\langle z, y^*\rangle \leqq 0 \quad \text{for all } y^* \in C.$$

Now, consider $x = \bar{x} + \theta z$ for sufficiently small $\theta > 0$ such that for $i \notin I(\bar{x})$, $g_i(x) < 0$, and $x \in B(\bar{x},\varepsilon)$. Hence, for some $a_i^0 \in \partial g_0(\bar{x})$, $a_i^0 \in \partial g_0(x)$. Hence $g_0(x) - g_0(\bar{x}) = \langle a_i^0, x\rangle - \gamma_i^0 - \langle a_i^0, \bar{x}\rangle + \gamma_i^0 = \theta\langle a_i^0, z\rangle < 0$. Also, for $i \in I(\bar{x})$, there is a $a_j^i \in \partial g_i(\bar{x})$ such that $a_j^i \in \partial g_i(x)$. Hence $g_i(x) - g_i(\bar{x}) = \langle a_j^i, x\rangle - \gamma_j^i - \langle a_j^i, \bar{x}\rangle + \gamma_j^i = \langle z, a_j^i\rangle \leqq 0$. Since $g_i(\bar{x}) = 0$, we get a contradiction that $\bar{x}$ is not a local minimum.

We now prove a sufficiency condition.

<u>Theorem 4.4</u>: Let $\bar{x}$ be a point such that

(i) There exist $\lambda_i \geqq 0$ for which

$$\lambda_i g_i(\bar{x}) = 0 \quad i = 1, \ldots, m.$$

(ii) Define the map $\Gamma(x) = \partial g_0(x) + \sum_{i=1}^{m} \lambda_i \, \partial \, g_i(x)$.

Then $0 \in \Gamma(\bar{x})$.

(iii) $\Gamma(x)$ is weakly monotone at $\bar{x}$ with respect to 0 on the set

$$F = \{x : g_i(x) \leqq 0, \ i = 1, \ldots, m\}.$$

Then $\bar{x}$ is a local minimum of g_0 on F.

<u>Proof</u>:

Let $x \in B(\bar{x},\varepsilon) \cap F$, and ε sufficiently small so that $B(\bar{x},\varepsilon) \subset \bigcup_j \sigma_i^j$ for each $i = 0, \ldots, m$. Then

$$g_0(x) - g_0(\bar{x}) \geqq g_0(x) + \sum_{i=1}^{m} \lambda_i g_i(x) - g_0(\bar{x}) - \sum_{i=1}^{m} \lambda_i g_i(\bar{x})$$

$$\geqq \langle x-\bar{x}, a_{j_0}^0\rangle + \sum_{i \in I(\bar{x})} \lambda_i \langle x-\bar{x}, a_{j_i}^i\rangle$$

$$\geqq 0$$

since $a_j^0 + \Sigma\lambda_i a_{j_i}^i \in \Gamma(x)$, and $\Gamma(x)$ is weakly monotone with respect to 0 at x, and so $\bar{x}$ is a local minimum.

5. The Fixed Point Approach to PL Nonlinear Programming

We will consider the application of the fixed point algorithm of [3] to the case where g_0 is piecewise linear on some subdivision of R^n, and g_i are convex functions. In this case, the mapping ℓ,(3.4), is applicable. As is evident from Theorem 3.5, in this case we will prove that the algorithms will find a "stationary point" $\bar{x}$ such that $0 \in \ell(\bar{x})$. In certain special cases, the progression of the algorithm will indicate if $\bar{x}$ is a local minimum, Saari and Saigal [10], for the general case considered here, $\bar{x}$ may be a local maximum or a saddle point of the function g_0.

That the algorithm will compute a stationary point can be established in a manner similar to the proof of Theorem 3.5. Since s is convex, the part of Theorem 3.4 pertaining to the nonexistence of a solution to (1.2) also carries through. The convergence of the algorithms can also be proved under the relaxed hypothesis that s be convex outside some bounded region; i.e., if D is a bounded set containing x_0, then for each $x \notin D$, and $y^* \in \partial s(x)$,

$$s(z) = s(x) + \langle y^*, z-x \rangle \quad \text{for all } z \notin D.$$

Then, starting with $r(x) = x-x_0$ and ℓ as defined in (3.4) we can prove:

<u>Theorem 5.1</u>: For any $\varepsilon > 0$, starting with the unique r-complete simplex containing x_0, the fixed point algorithm will generate ℓ- complete simplexes of size $\leqq \varepsilon(> 0)$, and thus will compute a stationary point of (1.1-2).

Proof:

Let $N = \max\ \{0, \sup\ \{s(x) : x \in B(x_1,\varepsilon) \cup D\}$ for an arbitrary x_1 such that $B(x_1,\varepsilon) \cap D = \phi$ and let $\overline{D} = D \cup \{x : s(x) \leqq N\}$. By assumption $\overline{D}$ is bounded, and s is convex outside $\overline{D}$. Now, assume that for some $\varepsilon > 0$ the algorithm fails to compute a ℓ- complete simplex. Then, there is a $\ell\upsilon r$- complete simplex of diameter $< \varepsilon$ sufficiently far from $\overline{D}$; i.e., $\sigma \notin \overline{D}$ and for every $x \in \sigma$, $\langle v^i - x_0, x - x_1 \rangle > 0$. Also, $s(v^i) > 0$ for all i. Now consider the point $v^i - x + x_1 \in \overline{D}$. Also, by assumption, $v^i - x + x_1 \notin D$. Hence $s(v^i - x + x_1) \geqq s(v^i) + \langle x_i - x, y^*_i \rangle$ for all $y^*_i \in \partial s(v^i)$ and as $v^i \notin \overline{D}$ and $v^i - x + x_1 \varepsilon \overline{D}$ we have $\langle x - x_1, y^*_i \rangle > 0$ for all $y^*_i \varepsilon \partial s(v^i)$; for all i. Thus σ is not $\ell\upsilon r$-complete, a contradiction.

6. Computational Considerations

As is evident from the sections 3 and 5, the convergence of the fixed point algorithms can be established under some general conditions on the problem, and differentiability is not necessary. Computational experience indicates that the computation burden increases when the underlying mappings are not smooth. For smooth mappings, under the usual conditions, the fixed point algorithms can be made to converge quadratically, Saigal [14]. This can be observed by comparing the solution of three nondifferentiable nonlinear programming problems implementing the mapping (3.4) presented in the appendix, Tables A.1-3, with the solution of a smooth problem of eighty variables in Table A.4.

On such a problem, reported in Netravali and Saigal [8], the growth of the number of function evaluations with the number of variables was tested. The results were as anticipated by the works of Saigal [11] and Todd [16]. It was predicted in these works that the function evaluations grow as $0(n^2)$, where n is the number of variables. (See [8, 4.1].)

REFERENCES

[1] Clarke, F.H., "Generalized Gradients and Applications," Transactions of the American Mathematical Society, 205 (1975), 247-262.

[2] Eaves, B.C., "Nonlinear Programming Via Kakutani Fixed Points," Working Paper No. 294, Center for Research in Management Science, University of California, Berkeley, 1970.

[3] Eaves, B.C. and R. Saigal, "Homotopies for Computation of Fixed Points on Unbounded Regions," Mathematical Programming, 8 (1972), 134-145.

[4] Hansen, T., "A Fixed Point Algorithm for Approximating the Optimal Solution of a Concave Programming Problem," Cowles Foundation Discussion Paper No. 277, Yale University, New Haven, Connecticut, 1969.

[5] Hansen, T. and H. Scarf, "On the Applications of a Recent Combinational Algorithm," Cowles Foundation Discussion Paper No. 272, Yale University, New Haven, Connecticut, 1969.

[6] Merrill, O.H., "Applications and Extensions of an Algorithm that Computes Fixed Points of Certain Upper-Semi Continuous Point to Set Mappings," Ph.D. Dissertation, University of Michigan, Ann Arbor, 1972.

[7] Mifflin, R., "An Algorithm for Constrained Optimization with Semismooth Functions," Mathematics of Operations Research, 2 (1977), 191-207.

[8] Netravali, A. and R. Saigal, "Optimum Quantizer Design Using a Fixed Point Algorithm," The Bell System Technical Journal, 55 (1976), 1423-1435.

[9] Rockafellar, R.T., Convex Analysis, Princeton University Press, 1970.

[10] Saari, D.G. and R. Saigal, in preparation.

[11] Saigal, R., "Investigations into the Efficiency of Fixed Point Algorithms," Fixed Points - Algorithms and Applications, S. Karamardian (ed.), Academic Press, New York (1977).

[12] Saigal, R., "Fixed Point Computing Methods," Encyclopedia of Computer Science and Technology, Vol. 8, Marcel Dekker, Inc., 1977.

[13] Saigal, R., "On Paths Generated by Fixed Point Algorithms," Mathematics of Operations Research, 1 (1976), 359-380.

[14] Saigal, R., "On the convergence rate of algorithms for solving equations that are based on methods of complementary pivoting," Mathematics of Operations Research, 2 (1977), 108-124.

[15] Scarf, H., "The Approximation of Fixed Points of a Continuous Mapping, SIAM Journal of Applied Mathematics, 16 (1967), 1328-1343.

[16] Todd, M.J., "On Triangulations for Computing Fixed Points," Mathematical Programming 10 (1976), 322-346.

[17] Todd, M.J., The Computation of Fixed Points and Applications, Springer-Verlage, Berlin-Heidelburg, 1976.

[18] Todd, M.J., "New Fixed Point Algorithms for Economic Equilibria and Constrained Optimization," Tech. Report No. 362, School of Operations Research, Cornel University, Ithaca, New York, 1977.

[19] Wolfe, P., "A method of conjugate subgredients for minimizing nondifferentiable functions," in Nondifferentiable Optimization, M.L. Balinski and P. Wolfe, eds., Mathematical Programming Study, 3 (1975), North-Holland, Amsterdam, 145-173.

APPENDIX

We now give some computational experience with solving non-differentiable optimization problems of fairly large number of variables. For comparison purposes, we also give the results of solving an eighty-variable smooth problem (where the convergence has been accelerated).

Problem 1

This is a 7 variable problem. It is a version of the problem considered by Natravali and Saigal [8]. The value of entropy on the entropy constraint is 2.7, and this is the 19th run in the series of runs done on this problem.

Problem 2

This is a 43 variable problem considered in Elsner, W.B., "A Descent Algorithm for the Multihour Sizing of Traffic Networks," Bell System Technical Journal, 56 (1977), 1405-1430.

This run was made on a piecewise linear version, while the problem formulated by Elsner was piecewise smooth. The function is convex.

Problem 3

This is the following 15 variable convex piecewise smooth problem:

$$\min_{x} \max_{1 \leq j \leq 5} f_i(x)$$

where

$$f_j(x) = 2 \sum_{i=1}^{5} c_{ij}x_{10+i} + 3d_j x^2_{10+j} + e_j - \sum_{i=1}^{10} a_{ij}x_i$$

$$j = 1, \ldots, 5$$

and the data a_{ij}, c_{ij}, d_{ij}, e_j are the same as that for problem 10 in the appendix of Himmelblau, D.M., Applied Nonlinear Programming, McGraw-Book Company, 1972.

Problem 4

Is the 80 variable problem considered by Kellogg, Li and Yorke, "A Constructive Proof of Brouwers Fixed Point Theorem and Computational Results," SIAM Journal of Numerical Mathematics 13 (1976): 473-483.

The results of the above four problems are summarized in Tables A.1-4, respectively.

TABLE A.1

Grid of Search	Number of Function Evaluations	Number of Simplexes Searched
8.0	34	121
4.0	111	160
2.0	28	54
1.0	286	690
0.5	539	1,034
0.25	487	918

Constrained minimization problem with piecewise linear objective function and one piecewise linear constraint in seven variables.

TABLE A.2

Grid of Search	Number of Function Evaluations	Number of Simplexes Searched
6.55	951	1,431
3.27	469	469
1.63	400	400
0.81	526	526
0.41	623	623
0.20	1,081	1,081
0.10	885	885

Unconstrained minimization of a piecewise linear convex function of 43 variables.

TABLE A.3

Grid of Search	Number of Function Evaluations	Number of Simplexes Searched
3.87	2,259	3,099*
1.94	381	381
0.97	244	244
0.48	667	667
0.24	347	347
0.12	253	253
0.06	364	364

Unconstrained minimization of a piecewise smooth convex function of 15 variables.

TABLE A.4

Grid of Search	Number of Function Evaluations	Number of Simplexes Searched
8.74	1,573	1,750
4.47	149	149
2.24	106	106
1.18	97	97
0.56	88	88
0.24	84	84
0.14	81	81
0.03	84	84
0.002	81	81
0.000009	82	82

Zero finding problem for a smooth function of 80 variables. The accelerated algorithm has been used to solve this problem.