The University of Wisconsin
Computer Sciences Department
1210 West Dayton Street
Madison, Wisconsin 53706

General Convergence Conditions in Nonlinear Programming and a Kuhn-Tucker Algorithm[1]

by

Hilbert K. Schultz

Computer Sciences Technical Report #140

November, 1971

(1) This research was supported by NSF Grant GJ-362 and an NSF Institutional Grant.

ABSTRACT

This paper presents a general definition of algorithmic convergence in mathematical programming and lists conditions which are sufficient for convergence in the sense of the definition. These conditions are also shown to be necessary for convergence in many applications. The definition and conditions are slight modifications of those given by Zangwill [1969 pp. 235 and 244]. Special cases of the assumptions required by Topkis and Veinott [1967] and Polak [1971] for their general algorithms are shown to imply that these general conditions are met, and hence their algorithms converge in the sense of our definition for these cases. The use of the theory is illustrated by proving convergence of the conditional gradient algorithm and a Kuhn-Tucker algorithm for which no prior convergence proof had been given.

1. Introduction

Most convergence proofs for algorithms in mathematical programming involve proving that each limit point of the sequence generated by the algorithm satisfies some sort of optimality criterion. Points satisfying this criterion are frequently called stationary points and for certain classes of problems such points solve the programming problem, although in many cases the criteria are only necessary. This paper defines convergence in the above terms and sets forth in a general context sufficient conditions for an algorithm to converge (in the sense of the definition) to a "stationary point". The conditions are also shown to be necessary for algorithmic convergence with certain additional assumptions which are satisfied by many algorithms.

In section 2 we present some basic definitions and conditions on an algorithm. These conditions are proven to be sufficient for convergence in section 3, and also necessary under certain additional assumptions. Section 4 presents a simple application of these conditions to proving convergence of a conditional gradient algorithm. A convergence proof for an algorithm based on attempting to satisfy the Kuhn Tucker conditions (the algorithm was suggested by Rosen [1965]) is given in section 5. In the last section we show how the convergence conditions of Polak [1971] and Topkis and Veinott [1967] are related to the conditions presented in section 2. Although we do not present a discussion in this report it is also possible to obtain convergence proofs for the gradient projection method of Rosen [1960] and Zoutendijk's [1960] feasible direction algorithm.

2. The Minimization Problem and Notation

The problem is to find $\bar{x}$ such that

(2.1) $$f(\bar{x}) = \min_{x \in X} f(x), \quad \bar{x} \in X \subset R^n$$

Let

(2.2) $$U = \{x \in X \mid f(x) = f(\bar{x})\},$$

the set of solution points. Assume further that there exists a set

(2.3) $$S \subset X, \quad S \neq \emptyset$$

called the set of stationary points. We make no other assumptions on X, U, or S at this time. We shall use the notation of Zangwill [1969] for subsequences. That is, if $x_1, x_2, \ldots,$ is a sequence, we shall denote subsequences by $\{x_j\}$, $j \in K$ where K is an <u>infinite</u> subset of the positive integers. We shall also adopt the convention of using primes to denote the transpose of matrices only, while vectors will be either a row or column vector depending on the context.

(2.4) Definition. An algorithm generating a sequence of points $x_1, x_2, \ldots$ is said to be convergent if the limit of every convergent subsequence is a stationary point, i.e. it is in S.

Remark: We adopt the convention that for a finite sequence the last element is assumed to be repeated so that an infinite sequence is generated (for theoretical purposes). Thus for an algorithm generating a finite sequence to be convergent the terminal point must be a stationary point.

Let $x_1, x_2, \cdots$ be a sequence generated by an algorithm and let Z be a continuous function from X into R. Suppose that the algorithm satisfies the following conditions:

(2.5) If the algorithm terminates at x_j, then $x_j \in S$.

If the algorithm generates an infinite sequence of points then

(2.6) $\forall k \; \exists L_k \ni \forall \ell \geq L_k \quad Z(x_\ell) \leq Z(x_k)$

and

(2.7) If $x_j \to \bar{x}$, $j \in K \subset \{1, 2, \cdots\}$ and $\bar{x} \notin S$ then $\exists x_k \ni Z(x_k) < Z(\bar{x})$.

Condition 2.6 says that for each k, from some point on, the terms are less than or equal to $Z(x_k)$. Thus an increase in Z is only allowed a finite number of times. The second condition, 2.7, is a relatively easy condition to verify for most algorithms as can be seen in sections 4 - 6 and it is the condition which requires using ε active constraints in gradient projection and feasible direction methods.

The definition we have taken for convergence is the one most commonly used (implicitly) in mathematical programming. The main difference between the approach taken here and that of Zangwill [1969] page 235 is that his definition of convergence and sufficient conditions for convergence place requirements on the algorithm for determining when there is no solution. However, that sort of approach detracts from the generality of the theory since most authors treat the existence of solutions separately from the algorithm.

In the next section we show that 2.5 - 2.7 are sufficient conditions for convergence without any assumptions on f or X. We also prove necessity but this requires some additional assumptions.

3. Convergence Theorems

(3.1) Theorem (Sufficiency)

If an algorithm satisfies 2.5 - 2.7 for a continuous Z then the algorithm is convergent, i.e. it satisfies 2.4.

Proof: If the algorithm terminates after a finite number of steps at x_j then $x_j \in S$ by 2.5. Therefore, suppose the algorithm generates an infinite number of points $x_1, x_2, \cdots$ and that $x_j \to \hat{x}$, $j \in K$. We must show $\hat{x} \in S$. Suppose $\hat{x} \notin S$, then by condition 2.7 $\exists\, x_k \ni$

$$Z(x_k) < Z(\hat{x})$$

But then by condition 2.6 $\exists\, L_k \ni$

$$Z(x_\ell) \leqq Z(x_k) < Z(\hat{x}) \qquad \forall\, \ell \geqq L_k$$

Hence

$$Z(x_j) \leqq Z(x_k) \qquad j \in K,\ j \geqq L_k$$

which implies $Z(\hat{x}) \leqq Z(x_k) < Z(\hat{x})$ since Z is continuous. This contradiction implies $\hat{x} \in S$.

Q.E.D.

(3.2) <u>Theorem</u> (Necessity)

If an algorithm is convergent, f is continuous, the x_j lie in a compact set for all j, S = U, and the algorithm terminates whenever $x_j \in U$, then the algorithm satisfies 2.5 - 2.7 with Z = f.

Proof: 2.5 is satisfied by the definition of convergence and the remark following it. Now suppose the algorithm generated an infinite sequence $x_1, x_2, \cdots$ and 2.6 is not satisfied for Z = f. Then $\exists\, k \ni \forall L_k \exists\, \ell > L_k \ni f(x_\ell) > f(x_k)$ so that we may take a subsequence $\{x_j\}$, $j \in K \ni$

(3.3) $f(x_j) > f(x_k)\ \forall\, j \in K$. Also since the x_j lie in a compact set we may take a further subsequence $\{x_j\}$, $j \in K' \subset K \ni x_j \to \hat{x}$ which implies that

(3.4) $$f(\hat{x}) \geq f(x_k).$$

But $f(x_k) > \min f(x),\ x \in X$ or the algorithm would have terminated at x_k by our hypothesis and thus $f(\hat{x}) > \min f(x),\ x \in X \Rightarrow \hat{x} \notin S$ which is a contradiction to the fact that the algorithm is convergent. Thus 2.6 is satisfied and 2.7 follows since every limit point is a stationary point by the definition of convergence.

Q.E.D.

We note that the hypothesis of Theorem 3.2 are common assumptions of many algorithms. We only require the continuity of f, the equivalence of stationary points and solutions to the minimization problem, the x_j remaining in a compact set, and that the algorithm recognize a solution when it finds one.

In section 4 we present a simple application of conditions 2.5 - 2.7 to proving convergence of the conditional gradient algorithm. In section 5 we present a convergence proof for an algorithm based on trying to satisfy the Kuhn Tucker conditions for a minimum point. It is a variation of an algorithm suggested by Rosen [1965]. In section 6 we show how the convergence conditions of Polak [1971] and Topkis and Veinott [1967] are related to the conditions 2.5 - 2.7 presented here.

4. Conditional Gradient Algorithm

We consider the problem of finding an $\bar{x}$ such that

(4.1) $f(\bar{x}) = \min f(x),\ x \in X \subset R^n$

and assume:

(4.2) $f \in C^1$, i.e. f is continuously differentiable and denote the gradient of f at x by $\nabla f(x)$.

(4.3) X is compact and convex

THE ALGORITHM

Given $x_0 \in X$

Step 1 Compute $\min\limits_{y \in X} \nabla f(x_j)y = \delta_j$. If $\delta_j < 0$ go to Step 2.

If $\delta_j \geqq 0$ terminate.

Step 2. $\min\limits_{\substack{x_j + \lambda y_j \in X \\ \lambda \geqq 0}} f(x_j + \lambda y_j) = f(x_{j+1})$.

Set $j + 1 \to j$ and return to Step 1.

Note $S = \{x \in X \mid \min\limits_{y \in X} \nabla f(x)y \geqq 0\}$ is the set of stationary points. We now verify that 2.5 - 2.7 are satisfied with $Z = f$. If $\delta_j \geqq 0$, x_j is optimal (see e.g. Levitin and Polak [1966]) so that 2.5 is satisfied. If $\delta_j < 0$ then there exists a $\bar{\lambda} > 0$ such that

$$f[x_j + \bar{\lambda} y_j] < f[x_j]$$

so that 2.6 is satisfied. Define $\theta(x) = \min_{y \in X} \nabla f(x)y$ and note that θ is continuous since X is compact.

For condition 2.7 suppose $x_j \to \bar{x}$, $j \in K$ and $\bar{x} \notin S$. Then $\exists$ $\hat{y} \in X \ni \theta(\bar{x}) = \nabla f(\bar{x})\hat{y} < 0$. Since X is compact, $\exists$ $\{x_j\}$, $j \in K' \subset K \ni$ $y_j \to \bar{y}$, $j \in K'$ and since θ is continuous $\nabla f(\bar{x})\hat{y} = \nabla f(\bar{x})\bar{y}$. Therefore, since $\nabla f(\bar{x})\bar{y} < 0$ $\exists$ $\bar{\lambda} > 0 \ni$

$$f[\bar{x} + \bar{\lambda}\bar{y}] = f(\bar{x}) - \bar{\varepsilon},\ \bar{\varepsilon} > 0 \Rightarrow$$

for sufficiently large j, $j \in K'$, that

$$f[x_j + \bar{\lambda}y_j] \leqq f(\bar{x}) - \frac{\bar{\varepsilon}}{2} \Rightarrow$$

$$f(x_{j+1}) \leqq f[x_j + \bar{\lambda}y_j] < f(\bar{x})$$

so condition 2.7 is met and the algorithm converges.

5. A Kuhn Tucker Algorithm

In examining algorithms for solving the problem of finding $\bar{x}$ such that

$$f(\bar{x}) = \min f(x),\ x \in X = \{x \mid g_i(x) \leqq 0,\ i = 1, \cdots, m\}$$

it becomes clear that two very broad and important classes of algorithms can be distinguished. The first class of algorithms is based on consideration of the entire feasible set X at each point. The versions of conditional gradient, gradient projection and Newton's method as discussed in Levitin and Polyak [1966] are examples of such techniques. Generally speaking the convergence theory of such methods is easier to establish (see e.g.

section 4) but the subproblem at each iteration is of a much higher degree of complexity than methods of class two. The second class of methods consists of those based only on the active constraints (those constraints for which $g_i(x) = 0$). Techniques such as those of Abadie [1970], Goldfarb [1969], McCormick [1970A] and [1970B], Ritter [1971], Rosen [1960], Zangwill [1967], and Zoutendijk [1960] and [1970] all fall into this category. In proving convergence of such algorithms some sort of anti-zigzagging procedure is usually required. One way of accomplishing this is to use ε tolerances (Zoutendijk [1960], Demjanov [1967], Zangwill [1969], Polak [1971]) and this technique will be used in proving convergence of the Kuhn Tucker algorithm given below. Thus 2.5 - 2.7 are shown to apply to algorithms in both classes. Mangasarian [1971] has devised a new approach to algorithms in the second class which has simplified the ε tolerance idea.

In the Kuhn Tucker algorithm we consider the linearly constrained nonlinear programming problem of finding $\bar{x}$ such that

(5.1) $$f(\bar{x}) = \min f(x),\ x \in X = \{x \mid n_i x - b_i \leqq 0,\ i = 1,\cdots,m\}$$

(5.2) $$n_i n_i = 1,\ i \in J = \{1,2,\cdots,m\}.$$

We use the notation:

(5.3) $$I(x_j,\varepsilon_j) = \{i \in J \mid -\varepsilon_j \leqq n_i x_j - b_i \leqq 0\}$$

(5.4) $$I(x_j) = I(x_j,0)$$

(5.5) N_{ε_j} = matrix of unit normals containing all n_i with $i \in I(x_j, \varepsilon_j)$

(5.6) N_I = matrix of unit normals containing all n_i, $i \in I$.

(5.7) U_I = vector in $R^{r(I)}$ where $r(I)$ is the number of constraints in I, and assume that:

(5.8) f is continuously differentiable and convex

(5.9) X is convex and compact.

The Kuhn Tucker conditions (Mangasarian [1969]) that $\bar{x}$ solve 3.1 are

(5.10) $\nabla f(\bar{x}) + N_J U_J = 0, \; U_J \in R^m$

(5.11) $N_J \bar{x} - b_j \leqq 0$

(5.12) $U_J [N_J \bar{x} - b_J] = 0$

(5.13) $U_J \geqq 0$

However, for an inactive constraint i, $n_i \bar{x} - b_i < 0$ so 5.12 $\Longrightarrow U_i = 0$ and thus conditions 5.10 - 5.13 can be written

(5.14) $\nabla f(\bar{x}) + N_{I(\bar{x})} U_{I(\bar{x})} = 0$

(5.15) $U_{I(\bar{x})} \geqq 0$ for a feasible point $\bar{x}$, where

$U_{I(\bar{x})} \in R^{r(I(\bar{x}))}$, $r(I(\bar{x}))$ = number of constraints in $I(\bar{x})$ and $U_i = 0$, $i \in J - I(\bar{x})$.

We now give a result due to Rosen [1965] in a somewhat more general form which shows that an attempt to verify the Kuhn Tucker conditions 5.14, 5.15 either leads to a feasible direction or shows that a point is optimal.

(5.16) Theorem [Rosen] Let $||v(\bar{u})||_2^2 \equiv \min_{\substack{u_I \geqq 0 \\ u_I \in R^{r(I)}}} ||\nabla f(\bar{x}) + N_I u_I||_2^2$

Then if $v(\bar{u}) \neq 0$,

(5.17) $N_I v(\bar{u}) \geqq 0$ and

(5.18) $\nabla f(\bar{x}) v(\bar{u}) = v(\bar{u}) v(\bar{u}) = ||v(\bar{u})||_2^2 > 0.$

Note that if $I = I(\bar{x})$ and $v(\bar{u}) = 0$ then $\bar{x}$ is optimal since the Kuhn Tucker conditions are satisfied.

Proof:

Let $p(u) = ||v(u)||_2^2$ so that $p(\bar{u}) = ||v(\bar{u})||_2^2 = v(\bar{u}) v(\bar{u})$ and set $\bar{g} = \nabla f(\bar{x})$

(5.19) Then from $v(\bar{u}) = \bar{g} + N_I \bar{u}$

$p(\bar{u}) = \bar{g}\bar{g} + 2\bar{g}N_I\bar{u} + \bar{u}N_I^{\prime}N_I\bar{u}$

and hence

$$(5.20) \qquad \frac{\partial p(\bar{u})}{\partial u_i} = 2\bar{g}n_i + 2\bar{u}N_I^{\prime}n_i = 2[\bar{g} + \bar{u}N_I^{\prime}]n_i = 2v(\bar{u})n_i \;,\; i \in I.$$

Thus we must have

$v(\bar{u})n_i \geq 0$ for $i \in I$ for suppose $v(\bar{u})n_j < 0$ for some j. Then we could increase u_j and decrease $p(\bar{u})$ since $\frac{\partial p}{\partial u_j} < 0$ which contradicts the fact that $\bar{u}$ is an optimal solution to 5.16. Thus 5.17 is proven.

Looking again at 5.20 we see that if $u_j > 0$ we must have $v(\bar{u})n_j = 0$ for if it were > 0 we could decrease u_j and decrease p which would again be a contradiction. Therefore, $v(\bar{u})n_i u_i = 0$, $i \in I$ since either $v(\bar{u})n_i = 0$ or $u_i = 0$ so that

$$(5.21) \qquad v(\bar{u})N_I\bar{u} = 0$$

Using 5.19 and 5.20 we have

$0 < v(\bar{u})v(\bar{u}) = v(\bar{u})[\nabla f(\bar{x}) + N_I\bar{u}] = v(\bar{u})\nabla f(\bar{x})$ i.e.

$$(5.22) \qquad \nabla f(x)v(u) > 0$$ so 5.18 is proven.

Q.E.D.

For convenience we now define a function $\emptyset$: $R^n \times R^1 \rightarrow R^1$ by

$$(5.23) \qquad \emptyset(x,\varepsilon) = \emptyset(x,I(x,\varepsilon)) = \min_{u\geq 0} ||\nabla f(x)+N_\varepsilon u||_2^2,\; u \in R^{r(I(x,\varepsilon))}.$$

and note that by the remark following theorem 5.16 if $\emptyset(x) = \emptyset(x,0) = 0$, x is optimal.

(5.24) Definition:

A point x is stationary if

$$x \in S = \{x \in X | \emptyset(x) = 0\}.$$

We can now state (see figure 1):

THE KUHN TUCKER ALGORITHM

Assume $x_1 \in X$ is given and $\varepsilon > 0$, $p > 0$ are fixed. Set $\varepsilon_1 = \varepsilon$, $p_1 = p$.

(5.25) Step 1. Compute $\emptyset(x_j)$. If $\emptyset(x_j) = 0$ terminate, otherwise go to step 2.

(5.26) Step 2. Compute $\emptyset(x_j,\varepsilon_j) = \delta_j$. If $\delta_j \geqq p_j$ set $s_j = v_{\varepsilon_j}$ where $\emptyset(x_j,\varepsilon_j) = ||v_{\varepsilon_j}||_2^2 \equiv ||\nabla f_j + N_{\varepsilon_j} u_{\varepsilon_j}||_2^2$. If $\delta_j < p_j$ go to step 3.

(5.27) Step 3. Set $\varepsilon_j = \frac{\varepsilon_j}{2}$, $p_j = \frac{p_j}{2}$ and return to step 2.

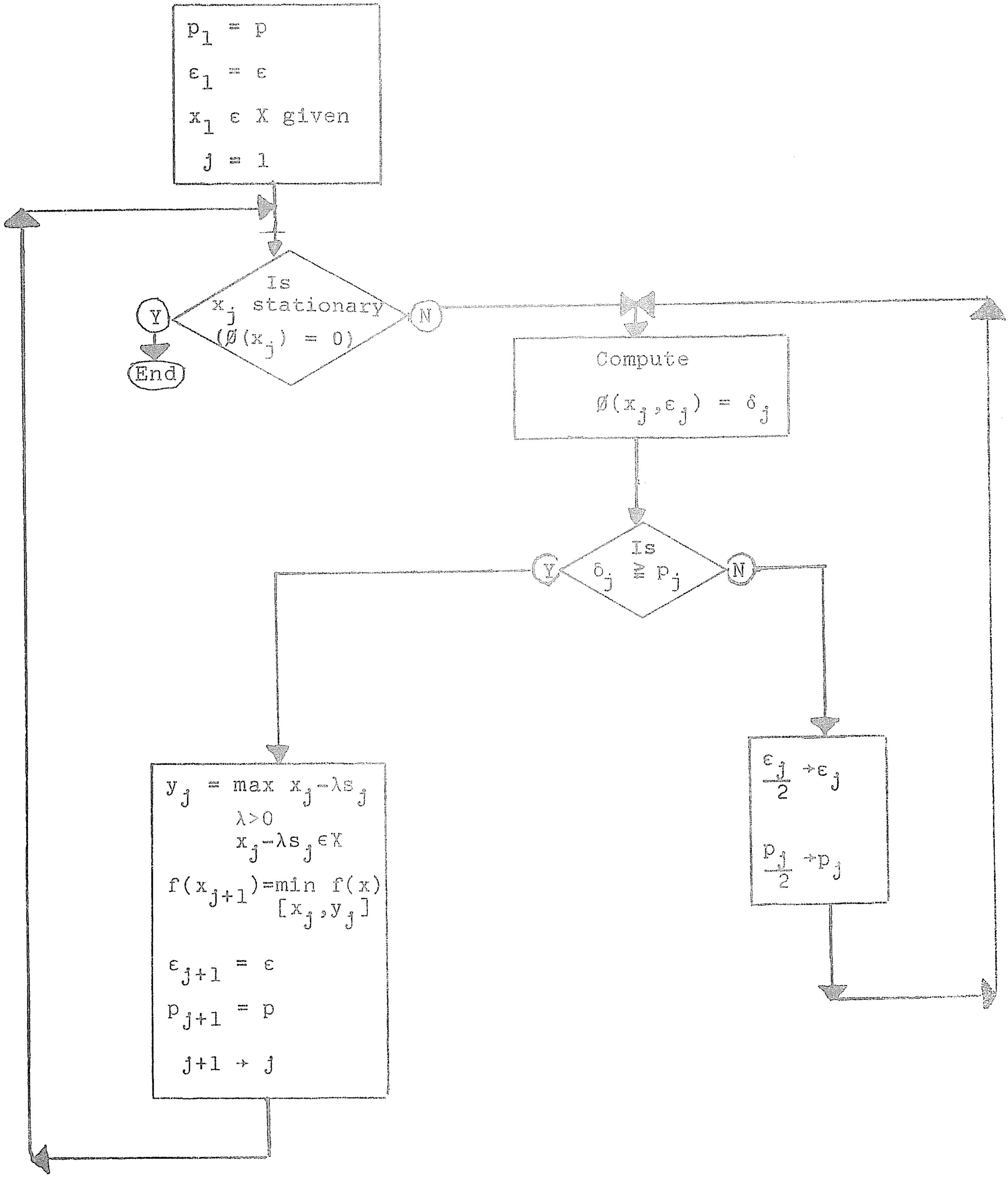
(5.28) Step 4. Determine y_j by $y_j = \max_{\substack{\lambda>0 \\ x_j-\lambda s_j \in X}} x_j - \lambda s_j = x_j - \lambda_j s_j$

and x_{j+1} by $f(x_{j+1}) = \min_{[x_j,y_j]} f(x)$. Set $\varepsilon_{j+1} = \varepsilon$, $p_{j+1} = p$, $j + 1 \rightarrow j$ and return to step 1.

Remarks: (1) The algorithm will only return to step 2 a finite number of times since for sufficiently small ε $\emptyset(x_j,\varepsilon_j) = \emptyset(x_j,I(x_j,\varepsilon_j)) = \emptyset(x_j,I(x_j,0)) = \emptyset(x_j) > 0$ by step 1. Thus once $I(x_j,\varepsilon_j) = I(x_j,0)$, $\emptyset$ remains constant with decreasing ε and we need only halve ε and p until $\emptyset(x_j) \geqq p_j$.

(2) Our theory also covers the case where $p = \varepsilon$ which is used by some authors (e.g. Polak [1971] for

Figure 1 Kuhn Tucker Algorithm Flowchart



gradient projection and feasible directions).

Before verifying conditions 2.5 - 2.7, we require the following theorem:

(5.29) Theorem. For a fixed $I \subseteq J$, $\phi(x,I) \equiv \min_{u_I \geq 0} ||\nabla f(x) + N_I u_I||_2^2$ is a continuous function of X.

Proof: Suppose $x_j \to \bar{x}$, x_j, $\bar{x} \in X$ and $\phi(x_j,I) \not\to \phi(\bar{x},I)$. Then $\exists\, \varepsilon > 0 \ni \forall N \exists\, j > N \ni$

(5.30) $|\phi(x_j,I) - \phi(\bar{x},I)| \geq \varepsilon$ i.e.

(5.31) $\phi(x_j,I) \geq \phi(\bar{x},I) + \varepsilon$ or

(5.32) $\phi(x_j,I) \leq \phi(\bar{x},I) - \varepsilon$.

Now one of 5.31 or 5.32 must occur infinitely often. Suppose it were 5.31. Then there is an infinite sequence $\{x_j\}$, $j \in L \subseteq \{1,2,\cdots\}$ such that

(5.33) $||\nabla f_j + N_I u_j||^2 \geq ||\nabla f(\bar{x}) + N_I \bar{u}||^2 + \varepsilon$, $j \in L$, where $\phi(x_j,I) = ||\nabla f_j + N_I u_j||^2$ and $\phi(\bar{x},I) = ||\nabla f(\bar{x}) + N_I \bar{u}||^2$. But since $\nabla f_j \to \nabla f(\bar{x})$, $||\nabla f_j + N_I \bar{u}||^2 \to ||\nabla f(x) + N_I \bar{u}||^2$ so that for sufficiently large j, $j \in L$,

(5.34) $||\nabla f(\bar{x}) + N_I \bar{u}||^2 - \frac{\varepsilon}{2} < ||\nabla f_j + N_I \bar{u}||^2 < ||\nabla f(\bar{x}) + N_I \bar{u}||^2 +$

Now combining the right inequality of 5.34 with 5.33 we have

$||\nabla f_j + N_I\bar{u}||^2 < ||\nabla f(\bar{x}) + N_I\bar{u}||^2 + \frac{\varepsilon}{2} < ||\nabla f(\bar{x}) + N_I\bar{u}||^2 + \varepsilon \leq ||\nabla f_j + N_I u_j||^2$ which is a contradiction since u_j minimizes $||\nabla f_j + N_I u_I||^2$ over $u_I \geq 0$. Thus 5.31 cannot occur infinitely often.

Suppose 5.32 occurs infinitely often, i.e.

(5.35) $||\nabla f_j + N_I u_j||^2 \leq ||\nabla f(\bar{x}) + N_I\bar{u}||^2 - \varepsilon \quad \forall j \in L$, $L \subseteq \{1, 2, \cdots\}$.

(5.36) This implies $\exists \bar{\varepsilon} > 0 \ni ||\nabla f_j + N_I u_j|| \leq ||\nabla f(\bar{x}) + N_I\bar{u}|| - \bar{\varepsilon} \quad \forall j \in L$ i.e. the inequality remains true without the square. Since $\nabla f_j \to \nabla f(\bar{x})$ we can take j so large that

(5.37) $||\nabla f_j - \nabla f(\bar{x})|| \leq \frac{\bar{\varepsilon}}{2}$ so that $||\nabla f(\bar{x}) + N_I u_j|| = ||\nabla f(\bar{x}) - \nabla f_j + \nabla f_j + N_I u_j|| \leq ||\nabla f_j + N_I u_j|| + ||\nabla f(\bar{x}) - \nabla f_j|| \leq ||\nabla f(\bar{x}) + N_I\bar{u}|| - \bar{\varepsilon} + \frac{\bar{\varepsilon}}{2}$ (by 5.36 & 5.37) $= ||\nabla f(\bar{x}) + N_I\bar{u}|| - \frac{\bar{\varepsilon}}{2} < ||\nabla f(\bar{x}) + N_I\bar{u}||$ which is a contradiction since $\bar{u}$ minimizes $||\nabla f(\bar{x}) + N_I u_I||$ over all $u_I \geq 0$. Thus 5.32 cannot occur infinitely often and hence we must have $\phi(x_j, I) \to \phi(\bar{x}, I)$.

Q.E.D.

We now verify that the algorithm is convergent via 2.5 - 2.7 with $Z = f$. Condition 2.5 is satisfied since the algorithm only terminates if $\phi(x_j) = 0$, i.e. $x_j \in S$. Condition 2.6 is satisfied by 5.18 since $-\nabla f_j v_j < 0 \Rightarrow f(x_{j+1}) < f(x_j)$. We next require:

(5.38) Lemma. If $x_j \to \bar{x}$, $j \in K$ and $\emptyset(\bar{x}) > 0$, then $\varepsilon_j \not\to 0$, $j \in K$.

Proof:

Suppose $\varepsilon_j \to 0 \Rightarrow p_j \to 0$, $j \in K$, since they are halved simultaneously. This also implies $\emptyset(x_j, 2\varepsilon_j) \leqq 2p_j$ for sufficiently large j, $j \in K$, or halving would not have been required. Since there are only finitely many constraints we may take a further subsequence $\{x_j\}$, $j \in K' \subset K$ such $\bar{I} \equiv I(x_j, 2\varepsilon_j)$ is a fixed set $\bar{I}$ for all $j \in K'$. Now $\bar{I} \subset I(\bar{x})$ since $2\varepsilon_j \to 0$, and $\emptyset(x_j, 2\varepsilon_j) \to 0$, $j \in K'$ since $2p_j \to 0$, but this implies $\emptyset(\bar{x}) = 0$ since $\emptyset(x_j, 2\varepsilon_j) = \emptyset(x_j, \bar{I}) \to \emptyset(\bar{x}, \bar{I})$ and $\bar{I} \subset I(\bar{x})$. This contradiction proves the assertion.

Q.E.D.

To verify condition 2.7 suppose that $x_j \to \bar{x}$, $j \in K$ and $\bar{x} \notin S$, i.e. $\emptyset(\bar{x}) > 0$.

By lemma 5.38 $\varepsilon_j \not\to 0$, $p_j \not\to 0$ so $\exists \varepsilon^* > 0$, $p^* > 0$ and a subsequence $\{x_j\}$, $j \in K' \subset K \ni \varepsilon_j \geqq \varepsilon^*$, $p_j \geqq p^*$, $j \in K'$. Thus $\emptyset(x_j, \varepsilon_j) \geqq p_j \geqq p^*$ and since there are only finitely many constraints we may take a further subsequence $\{x_j\}$, $j \in K'' \subset K' \ni I(x_j, \varepsilon_j) = I$, a fixed set $\forall j \in K''$. Then by the continuity of $\emptyset$ (theorem 5.29), $\emptyset(\bar{x}, I) \geqq p^*$. Also since $\varepsilon_j \not\to 0$ $I(\bar{x}) \subset I$. Now since X is compact y_j and λ_j lie in compact sets so that we may take a further subsequence $\{x_j\}$, $j \in K''' \subset K''$

$\ni$ $x_j \to \bar{x}$, $y_j \to \bar{y}$, $\lambda_j \to \bar{\lambda}$. We claim $\bar{\lambda} > 0$ for if $\bar{\lambda} = 0$ then for each $j \in K'''$ we must have $n_i y_j - b_i = 0$ and $n_i s_j < 0$ for some $i \in J$ or λ_j could be increased. Since there are finitely many constraints we may take a further subsequence $\{x_j\}$, $j \in K^{iv} \subset K''' \ni$ this occurs for the same i. But then since $\lambda_j \to 0$, $y_j \to \bar{x} \Rightarrow n_i\bar{x} - b_i = 0 \Rightarrow i \in I(\bar{x}) \subset I$ which is a contradiction to theorem 5.16 since $n_i s_j \geqq 0 \; \forall \, i \in I$. Thus $\lambda_j \to \bar{\lambda} > 0$, $j \in K^{iv}$. Now $\nabla f_j[y_j - x_j] = -\lambda_j \nabla f_j s_j = -\lambda_j \emptyset(x_j, \varepsilon_j)$ by theorem 5.16, and $-\lambda_j\emptyset(x_j,\varepsilon_j) \leqq -\lambda_j p^*$ $j \in K^{iv}$. Therefore since $x_j \to \bar{x}$, $y_j \to \bar{y}$ and ∇f_j is continuous we have

$$\nabla f(\bar{x})(\bar{y} - \bar{x}) \leqq -\bar{\lambda}p^* < 0.$$

Thus there exists $1 > \hat{\lambda} > 0$ and $\bar{\varepsilon} > 0 \ni$

$$f[\hat{\lambda}\bar{x} + (1 - \hat{\lambda})\bar{y}] = f(\bar{x}) - \bar{\varepsilon} \Rightarrow$$

for sufficiently large j by the continuity of f that

$$f[x_{j+1}] \leqq f[\hat{\lambda}x_j + (1-\hat{\lambda})y_j] \leqq f(\bar{x}) - \frac{\bar{\varepsilon}}{2} < f(\bar{x})$$

i.e. condition 2.7 is satisfied and thus the algorithm is convergent.

A version of this algorithm was implemented on the Univac 1108 at the University of Wisconsin by Teorey [1971]. The algorithm was used to solve test problems 1 and 7 of Colville's [1968] nonlinear programming study. Problem #1 was solved in a standard time of .0057 units and problem #7 was solved in a standard time of .0202 units indicating that the Kuhn Tucker algorithm is one of the faster methods.

6. Applications To Other Convergence Conditions

We now show that the general approaches to algorithmic convergence utilized by some other authors imply our general conditions. To that end we present section 2 of the article by Topkis and Veinott [1967] in an appropriate form for the minimization problem:

(6.1) $\min F(x),\ x \in X \subset R^n$

Some of the notation and definitions are:

(6.2) $d \in R^n$ is feasible direction at $x \in X$ if $\exists\, \delta > 0 \ni$ $x + sd \in s\ \forall s \ni 0 < s \leqq \delta$.

(6.3) A feasible direction $d \ni F(x) - s\delta > F(x+sd)$ $\forall s \ni 0 < s \leqq \delta$ is called a usable direction.

(6.4) If there is no usable direction at $x \in X$ then x is called a stationary point.

Let $x_0 \in X$ be given and define

(6.5) $X_0 = \{x \in X | F(x) \leqq F(x_0)\}$.

Topkis and Veinott make the following assumptions:

I. X is closed, F is continuous on X and X_0 is compact.

II. For every sequence $\{x_1, x_2, \ldots\}$ in X_0 there is a bounded direction function d which assigns to $\{x_0, x_1, \ldots, x_n\}$

a feasible direction $d(x_0,x_1,\dots,x_n) = d_n$ at x_n for $n = 0, 1, \cdots$. Moreover, there is a specified infinite set P of nonnegative integers such that any subsequence $\{(x_{n_k}, d_{n_k})\}$ of $\{(x_n,d_n) | n \in P\}$ converging to $(\bar{x}, \bar{d})$ has the property that

(a) for some $\delta > 0$, $x_{n_k} + sd_{n_k} \in X$ for $k = 1,2,\cdots$ and all s, $0 < s \leqq \delta$ and

(b) if $\bar{x} \in X_0$ and $\bar{d}$ is feasible but not usable at $\bar{x}$, then $\bar{x}$ is stationary.

III. There is a real valued upper semi-continuous step size function $f(x,y)$ defined on $X_0 x X$ for which $f(x,x+\lambda d)$ is continuous in λ for fixed $x \in X_0$, $d \in R^n$ and

(a) $f(x,y) \geqq F(y)$ and $f(x,x) = F(x)$ for all $x \in X_0$, $y \in X$ and

(b) if d is a usable direction for F at $x \in X_0$ then d is also usable for $f(x,\cdot)$ at x.

Under the above hypothesis the algorithm of Topkis and Veinott consists of computing x_{n+1} by

(1) $x_{n+1} = x_n + s_n d_n \qquad n = 0,1,\cdots$
where $d_n = d(x_0,x_1,\dots,x_n)$ and $s_n \geqq 0$ is chosen so that $x_n + s_n d_n \in X$ and such that $F(x_{n+1}) \leqq F(x_n)$ for $n \notin P$ and for $n \in P$ $f(x_n,x_n + s_n d_n) \leqq f(x_n,x_n + sd_n)$ for all $s \geqq 0$ with $x_n + sd_n \in S$.

Since Topkis and Veinott only discuss limit points of $\{x_j\}$, $j \in P$ our definition of convergence is not strictly applicable. However, there are several different ways in which our theory can be modified so that 2.5 - 2.7 are implied by the above. One is to require P to be the set of all positive integers and the other is redefine convergence to mean only convergence of all subsequences $\{x_j\}$, $j \in P$. We shall take the first alternative and let $P = \{1,2,\cdots\}$. Condition 2.5 is obviously satisfied. For an infinite sequence of points, we set Z = F in verifying 2.6 and 2.7. The second requirement 2.6 is then obvious since $F(x_{k+1}) \leqq F(x_k)$ for all k. For 2.7 suppose $x_j \to \bar{x}, d_j \to \bar{d}$, $j \in k$ and suppose $\bar{x}$ is not a stationary point. By IIa $\bar{d}$ is feasible so that by IIb $\bar{d}$ must be usable, and hence by IIIb $\bar{d}$ is usable for f. We let $\bar{s}$ be such that $f(\bar{x},\bar{x}+\bar{s}\bar{d}) \leqq f(\bar{x},\bar{x}+s\bar{d})$ $\forall\, s \geqq 0 \ni$ $\bar{x} + s\bar{d} \in X$. Then since $\bar{d}$ is usable for f

$$f(\bar{x},\bar{x}+\bar{s}\bar{d}) = f(\bar{x},\bar{x}) - \bar{\varepsilon}. \tag{6.6}$$

Now f is upper semi-continuous so that for $\frac{\bar{\varepsilon}}{2} > 0$ $\exists\, \delta > 0 \ni$ if $||x - \bar{x} - \bar{s}\bar{d}|| < \delta$ and $||y - \bar{x}|| < \delta$ then

$$f(y,x) < f(\bar{x},\bar{x}+\bar{s}\bar{d}) + \frac{\bar{\varepsilon}}{2}$$

Therefore taking $||x_j + \bar{s}d_j - \bar{x} - \bar{s}\bar{d}|| < \delta$ and $||x_j - \bar{x}|| < \delta$

$$\begin{aligned} F(x_{j+1}) &\leqq f(x_j,x_{j+1}) \leqq f(x_j,x_j+\bar{s}d_j) < f(\bar{x},\bar{x}+\bar{s}\bar{d}) + \frac{\bar{\varepsilon}}{2} \\ &= f(\bar{x},\bar{x}) - \bar{\varepsilon} + \frac{\bar{\varepsilon}}{2} = f(\bar{x},\bar{x}) - \frac{\bar{\varepsilon}}{2} \quad \text{(by 6.6)} \\ &< f(\bar{x},\bar{x}) = F(\bar{x}) \end{aligned}$$

i.e. 2.7 is satisfied and the algorithm converges.

We now show that the requirements of Polak [1971] page 15-16 imply the conditions 2.5, 2.6, 2.7 or modifications of them for the nonlinear programming problem 2.1. His algorithm model consists of a set valued search function A mapping X into the set of all nonempty subsets of X, written $A : X \to 2^X$. Polak considers points as being either desirable or nondesirable and assumes the existence of a stop function c which is either continuous at all nondesirable points $x \in X$ or else $c(x)$ is bounded from below for $x \in X$. We first consider the case where c is continuous at all nondesirable points.

(6.7) Polak's Algorithm Model: $A: X \to 2^X$, $c: X \to R^1$.

Step 0 Compute a $x_0 \in X$

Step 1 Set $i = 0$

Step 2 Compute $y \in A(x_i)$

Step 3 Set $x_{i+1} = y$

Step 4 If $c(x_{i+1}) \geqq c(x_i)$, stop; else set $i = i+1$ and go to step 2.

Polak also assumes that c satisfies:

(6.8) for every $x \in X$ which is not desirable there exists an $\varepsilon(x) > 0$ and a $\delta(x) < 0$ such that

$$c(x'') - c(x') \leqq \delta(x) \quad \forall \quad x' \in X \ni ||x'-x|| \leqq \varepsilon(x)$$

and for all $x'' \in A(x')$.

We shall use the term stationary points to describe the points he refers to as desirable points.

(6.9) Theorem. If c is continuous at all non-stationary points and satisfies (6.8) then the algorithm satisfies 2.5, 2.6, 2.7, i.e. it is convergent with $Z = c$.

Proof: The algorithm only terminates if $c(x_{i+1}) \geq c(x_i)$, i.e. x_i is stationary by (6.8) so 2.5 is satisfied. If the algorithm generates an infinite sequence of points $x_1, x_2, \cdots$ then 2.6 is satisfied with $Z \equiv c$, since $Z(x_{i+1}) < Z(x_i)$. Now suppose $x_i \to \bar{x}$, $i \in K$ and $\bar{x}$ is not a stationary point. Then by (6.8) $\exists \varepsilon(\bar{x}) > 0$, $\delta(\bar{x}) < 0 \ni ||x' - \bar{x}|| < \varepsilon(\bar{x}) \Rightarrow$

$$c(x'') - c(x') \leq \delta(\bar{x}) \quad \forall x'' \in A(x')$$

But since $x_i \to \bar{x}$, $i \in K$ for sufficiently large i, $i \in K$, $||x_i - \bar{x}|| < \varepsilon(\bar{x})$ and since $c(x_i) \to c(\bar{x})$ we may take $i \in K$ so large that $||x_i - \bar{x}|| < \varepsilon(\bar{x})$ and so that

(6.10) $$|c(x_i) - c(\bar{x})| < - \frac{\delta(\bar{x})}{2}$$

Then since $x_{i+1} \in A(x_i)$ we have

(6.11) $$c(x_{i+1}) - c(x_i) \leq \delta(\bar{x})$$ and from 6.10

(6.12) $$c(x_i) < c(\bar{x}) - \frac{\delta(\bar{x})}{2}$$ so that

combining 6.11 and 6.12 we have

$$c(x_{i+1}) \leqq c(x_i) + \delta(\bar{x}) < c(\bar{x}) - \frac{\delta(\bar{x})}{2} + \delta(\bar{x})$$

$$= c(\bar{x}) + \frac{\delta(\bar{x})}{2} < c(\bar{x})$$

i.e. 2.7 is satisfied and the algorithm is convergent.

Polak's alternate assumption on the stop function c, namely that it be bounded below on X does not imply 2.5 - 2.7 directly. However, it is possible to modify these requirements to make 2.6 a stronger condition in the interest of weakening the continuity requirement on Z. That is, we replace 2.6 by

(2.6') $Z(x_{k+1}) \leqq Z(x_k) \; \forall k \in \{1,2,\cdots\}$

and require that Z satisfy 6.8 at all non-stationary points as well as Z being bounded below on X. We omit the proofs here but these assumptions on Z together with 2.5, 2.6', 2.7 imply convergence in the sense of 2.4 and Polak's assumptions imply 2.5, 2.6' and 2.7.

BIBLIOGRAPHY

1. Abadie, J. [1970] Applications of the GRG Algorithm to Optimal Control Problems, in Integer and Nonlinear Programming, Abadie, J., Editor. American Elsevier Publishing Company.

2. Demjanov, V. F. [1967] On the Solution of Certain Minimax Problems, I, II. Kybernetica, 2, 58-66 (1966); 3, 62-66 (1967).

3. Goldfarb, D. [1969] Extension of Davidson's Variable Metric Method to Maximization under Linear Inequality and Equality Constraints. SIAM J. Appl. Math., 17, 739-764 (1969).

4. Levitin, E. S., Polyak, B. T. [1966] Constrained Minimization Methods, (Russian) Zh. vych. Mat. Mat. Fiz., 6, 787-823 (1966). Also translated in USSR Comput. Math. Math Phys., 6, 1-50 (1966).

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

6. Mangasarian, O. L. [1971] Dual, Feasible Direction Algorithms, Mathematics Research Center, University of Wisconsin, Report #1173, August 18, 1971.

7. McCormick, Garth P. [1970A] A Second Order Method for the Linearly Constrained Nonlinear Programming Problem, In Nonlinear Programming, Rosen, J. B., Mangasarian, O. L., Ritter, K., editors, Academic Press, New York.

8. McCormick, G. [1970B] The Variable Reduction Method for Nonlinear Programming, Management Science, Theory Series, 17, 146-160 (1970).

9. Polak, E. [1971] Computational Methods in Optimization, A Unified Approach, Academic Press, New York.

10. Ritter, K. [1971] A Superlinearly Convergent Method for Minimization Problems with Linear Inequality Constraints. Mathematics Research Center, University of Wisconsin, Report #1098, June 1971.

11. Rosen, J. B. [1960] The Gradient Projection Method for Nonlinear Programming, Part I, Linear Constraints, SIAM J. Appl. Math., 8, 181-217 (1960).

12. Rosen, J. B. [1965] Gradiant Projection as a Least Squares Solution of Kuhn Tucker Conditions, Computer Sciences Department, University of Wisconsin, Memorandum July 1965.

13. Theorey, T. J. [1971] Private Communication.

14. Topkis, D. M., and Veinott, A. F., Jr. [1967] On the Convergence of Some Feasible Direction Algorithms for Nonlinear Programming. SIAM J. Control, 5, 268-279 (1967).

15. Zangwill, W. I. [1967] An Algorithm for the Chebyshev Problem - With an Application to Concave Programming, Management Science, 14, 58-78 (1967).

16. Zangwill, W. I. [1969] Nonlinear Programming. Prentice Hall, Englewood Cliffs, New Jersey.

17. Zoutendijk, G. [1960] Methods of Feasible Directions, Elsevier Publishing Co., Amsterdam.

18. Zoutendijk, G. [1970] Some Algorithms Based on the Principle of Feasible Directions, In Nonlinear Programming, Rosen, J. B., Mangasarian, O. L., Ritter, K., editors, Academic Press, New York.