24

THE VALIDITY OF A FAMILY OF OPTIMIZATION METHODS

by

Robert Meyer

Technical Report #28

July 1968

THE VALIDITY OF A FAMILY OF OPTIMIZATION METHODS

Abstract

A family of iterative optimization methods, which includes most of the well-known algorithms of mathematical programming, is described and analyzed with respect to the properties of its accumulation points. It is shown that these accumulation points have desirable properties under appropriate assumptions on a relevant point-to-set mapping. The conditions under which these assumptions hold are then discussed for a number of algorithms, including steepest descent, the Frank-Wolfe method, feasible direction methods, and some second-order methods. Five algorithms for a special class of nonconvex problems are also analyzed in the same manner. Finally, it is shown that the results can be extended to the case in which the subproblems constructed are only approximately solved and to algorithms which are composites of two or more algorithms.

1. Semi-Continuity and Mathematical Programming

The concepts of upper and lower semi-continuity for point-to-set mappings have been studied by a number of prominent mathematicians, including Hausdorff, [1] Berge, [2] and McShane [3]. Several similar definitions of the two concepts have been formulated, and some comparisons may be found in a recent paper by Jacobs [4]. The following definitions, which are essentially the same as those given in Debreu [5], will be used in this paper: a point-to-set mapping Ω with domain G and range consisting of subsets of a set R is said to be (1) upper semi-continuous (u.s.c.) at a point y belonging to G if $y_i \to y$, $\{y_i\} \subset G$, and $z_i \to z$ with $z_i \in \Omega(y_i)$ for each i, imply $z \in \Omega(y)$; (2) lower semi-continuous (l.s.c.) at a point y belonging to G if $z \in \Omega(y)$, $y_i \to y$, $\{y_i\} \subset G$, imply the existence of an integer m and a sequence $\{z_m, z_{m+1}, \ldots\}$ with the properties that (a) $z_i \in \Omega(y_i)$ for $i \geq m$ and (b) $z_i \to z$; and (3) continuous at a point y if it is both upper and lower semi-continuous at y. Note that these definitions are meaningful whenever the notion of convergence is defined in both G and R. In particular, they are valid if G and R are subsets of topological or metric spaces. (If $\Omega(y)$ is a single point for every $y \in G$, i.e., a function, then it is easily seen that l.s.c. at a point implies u.s.c. and hence continuity at that point. Similarly, if Ω is single-valued and R is a sequentially compact subset of a topological space, then it is true that u.s.c. at a point implies l.s.c. and hence continuity at that point.

However, it is easy to construct set-valued mappings that are only u.s.c. or only l.s.c. even when R is a compact subset of E^n. Examples displaying this behavior appear below. These notations should not be confused with numerical upper and lower semi-continuity for real-valued functions, which have quite different definitions.)

An important class of point-to-set mappings consists of those mappings that involve the linearization of some or all of the constraints defining a set about a point. Let M denote the set $S \cap \{z \mid u(z) \geq 0\} \cap \{z \mid v(z) = 0\}$, where S is a closed subset of a Banach space B and u and v are continuously Frechet differentiable vector-valued functions. For a point $y \in B$, we will say that the "linearization" of M about y is the set

$$\Gamma m(y) := S \cap \{z \mid u(y) + u'(y)(z-y) \geq 0, \quad v(y) + v'(y)(z-y) = 0\} .$$

Note that if $y \in M$, then $\Gamma m(y)$ is non-empty, since $y \in \Gamma m(y)$. We will now show that the point-to-set mapping Γm is u.s.c. at every point of B. For, let $y \in B$, and let $y_i \rightarrow y$. If $z_i \in \Gamma m(y_i)$ for each i and $z_i \rightarrow z$, then it follows from the closure of S that $z \in S$, and it follows from the continuity of u' and v' at y that $u(y) + u'(y)(z-y) = \lim (u(y_i) + u'(y_i)(z_i - y_i)) \geq 0$ and $v(y) + v'(y)(z-y) = \lim (v(y_i) + v'(y_i)(z_i - y_i)) = 0$, so that $z \in \Gamma m(y)$, proving u.s.c. . Without additional hypotheses, however, it is not true that Γm is l.s.c. This fact was demonstrated by Rosen [6], and it can also be deduced from the following very simple example where B is taken to be the real line.

EXAMPLE: Take $S = E^1$, $v \equiv 0$, and $u(z) = z^3$. Let $y_i = 1/i$, so that $y_i \to 0$ as $i \to \infty$ and $\Gamma m(y_i) = \{z \mid y_i^3 + 3y_i^2(z-y_i) \geq 0\} = \{z \mid z \geq 2/3i\}$. However, $\Gamma m(0) = E^1$, and it is clear that Γm is u.s.c. but not l.s.c. at the point 0.

In the case that $B = E^n$ and S is a convex set consisting of the points satisfying $f(z) \geq 0$, where f is continuous and vector-valued, the next theorem gives sufficient conditions for l.s.c. of Γm in the neighborhood of a point. We adopt the convention of calling the inequality constraint $f_i(z) \geq 0$ $(u_i(y) + u_i'(y)(z-y) \geq 0)$ active at the point $\bar{z} \in \Gamma m(y)$ if $f_i(\bar{z}) = 0$ $(u_i(y) + u_i'(y)(\bar{z}-y) = 0)$. The (possibly vector-valued) function consisting of active constraint functions at $\bar{z} \in \Gamma m(y)$ is understood to consist of those functions $f_i(z)$ and $u_i(y) + u_i'(y)(z-y)$ which correspond to active inequality constraints at $\bar{z}$ as well as the function $v(y) + v'(y)(z-y)$.

THEOREM 1.1: Under the preceding assumptions on B and S, the point-to-set mapping Γm is l.s.c. in a neighborhood of a point y^* if the set $\Gamma m(y^*)$ contains a point z^* at which the gradients to the active constraint functions at z^* are linearly independent.

PROOF: See Appendix.

We will now obtain three basic results relating semi-continuity of point-to-set mappings to mathematical programming. Similar results may be found in Berge [2] and Debreu [5]. It will be assumed that f is a real-valued function defined and continuous on $R \times G$ and that the optimal value function $\mu(y) := \underset{z \in \Omega(y)}{\text{minimum}} \; f(z, y)$ is well-defined for every $y \in G$.

<u>LEMMA 1.2</u>: If Ω is u.s.c. at a point $y^* \in G$ and R is sequentially compact, then μ is (numerically) lower semi-continuous at y^*.

PROOF: Let $y_i \rightarrow y^*$, $\{y_i\} \subset G$. Then there exist sequences $\{y_{n_i}\}$ and $\{z_{n_i}\}$ such that $\mu(y_{n_i}) = f(z_{n_i}, y_{n_i})$, $z_{n_i} \rightarrow z^*$, and $\mu(y_{n_i}) \rightarrow \underline{\lim}\ \mu(y_i)$ as $i \rightarrow \infty$. It follows from u.s.c. that $z^* \in \Omega(y^*)$, and thus $\underline{\lim}\ \mu(y_i) = \lim \mu(y_{n_i}) = f(z^*, y^*) \geq \mu(y^*)$.

If the compactness hypothesis is deleted, the conclusion is no longer valid. An example illustrating this is given in the Appendix. However, compactness is not required in the following complementary result.

<u>LEMMA 1.3</u>: If Ω is l.s.c. at a point $y^* \in G$, then μ is (numerically) upper semi-continuous at y^*.

PROOF: Let $z^* \in \Omega(y^*)$ be such that $\mu(y^*) = f(z^*, y^*)$, and let $\{y_i\}$ be an arbitrary sequence in G converging to y. Choose $\{y_{n_i}\}$ and $\{z_{n_i}\}$ such that $\mu(y_{n_i}) \rightarrow \overline{\lim}\ \mu(y_i)$ and $z_{n_i} \rightarrow z^*$, with $z_{n_i} \in \Omega(y_{n_i})$. We then have $\mu(y^*) = f(z^*, y^*) = \lim f(z_{n_i}, y_{n_i}) \geq \lim \mu(y_{n_i}) = \overline{\lim}\ \mu(y_i)$.

Combining the previous two lemmas, we obtain:

<u>THEOREM 1.4</u>: If Ω is continuous at a point $y^* \in G$ and R is sequentially compact, then μ is (numerically) continuous at y^*.

The next theorem reflects a slightly different viewpoint. It shows that continuity of Ω is a sufficient condition for the limit of a set of solutions to solve the limiting problem. Note that compactness does not enter directly into the statement of the result.

<u>THEOREM 1.5:</u> Let $M(y)$ denote the subset of $\Omega(y)$ consisting of all points z such that $f(z, y) = \mu(y)$. If Ω is continuous at $y^* \in G$, then the point-to-set mapping M is u.s.c. at y^*.

PROOF: Let $\{y_i\} \subset G$ converge to y^*, and let $z_i \in M(y_i)$ for each i, with $z_i \to z^*$. Since Ω is u.s.c. at y^*, it follows that $z^* \in \Omega(y^*)$, and thus $\mu(y^*) \leq f(z^*, y^*)$. On the other hand, by lemma 1.3, $f(z^*, y^*) = \lim \mu(y_i) \leq \mu(y^*)$.

Let us now suppose that $R \subset G$ and that we have a continuous function φ defined on G with the property that $y' \in M(y)$ implies $\varphi(y') < \varphi(y)$ unless $y \in M(y)$. (This will sometimes be referred to as the <u>strict monotonicity property</u>.) Consider the algorithm defined as follows:

(a) Choose an arbitrary $y_0 \in G$.

(b) Let $y_{i+1} = y_i$ if $y_i \in M(y_i)$; otherwise let $y_{i+1} \in M(y_i)$.

<u>THEOREM 1.6:</u> If $\{y_i\}$ is contained in a sequentially compact set and y^* is an accumulation point of $\{y_i\}$ at which Ω is continuous, then $y^* \in M(y^*)$.

PROOF: If the conclusion were false, we would have $\varphi(\bar{y}) < \varphi(y^*)$ for all $\bar{y} \in M(y^*)$ by the assumption on φ. We will show that this leads to a contradiction. Let subsequences $\{y_{n_i}\}$ and $\{y_{n_i+1}\}$ be chosen so that $y_{n_i} \rightarrow y^*$ and $y_{n_i+1} \rightarrow y'$. It follows from the previous theorem that $y' \in M(y^*)$, so that $\varphi(y') < \varphi(y^*)$. However, since $\{\varphi(y_i)\}$ is a monotone decreasing sequence we have $\varphi(y') = \lim \varphi(y_{n_i+1}) = \lim \varphi(y_i) = \lim \varphi(y_{n_i}) = \varphi(y^*)$, a contradiction.

In a large number of well-known algorithms, it is the case that $f(z, y) = \varphi(z)$, and the preceding result may be sharpened.

<u>THEOREM 1.7</u>: If $\varphi(z) = f(z, y)$ and y^* is an accumulation point of $\{y_i\}$ at which Ω is l.s.c., then $y^* \in M(y^*)$.

PROOF: Again suppose that the conclusion is false, and let $\bar{y} \in M(y^*)$, so that $\varphi(\bar{y}) < \varphi(y^*)$. Since Ω is assumed l.s.c. at y^*, there exists a sequence of points $\{z_{n_i}\}$ with $z_{n_i} \in \Omega(y_{n_i})$ for each n_i and such that $z_{n_i} \rightarrow \bar{y}$. Thus we conclude that $\varphi(y^*) > \varphi(\bar{y}) = \lim \varphi(z_{n_i}) \geq \lim \varphi(y_{n_i+1}) = \varphi(y^*)$, which cannot hold.

Theorem 1.6 is similar to a result established previously by Zangwill [7]. Zangwill, however, directly assumes u.s.c. of the point-to-set mapping $M(z)$, and considers how this property may be demonstrated for various algorithms, rather than stating a theorem in terms of the properties of the underlying point-to-set mapping $\Omega(z)$. One of the advantages of considering only $\Omega(z)$ is the elimination of the compactness hypothesis in the case $f(z, y) = \varphi(z)$. (This hypothesis cannot be eliminated in theorem 1.6. A counterexample is given in the Appendix.)

2. An Application: Reverse-Convex Programming

Consider the following problem,

I: minimize $\varphi(z)$

subject to $z \in F := S \cap \{z \mid u(z) \geq 0\}$,

where S is a closed, convex subset of E^n; u is a vector-valued, convex and continuously differentiable function; and φ is continuous and real-valued on F. We shall further assume that F is bounded, which is easily seen to imply that F is compact. The curious feature of problems of the form I is the non-convexity of the feasible region F. The convexity of u implies that the region $\{z \mid u(z) \leq 0\}$ given by the <u>reverse</u> inequalities is convex (see Figure 1). For this reason, the sets $U := \{z \mid u(z) \geq 0\}$ and F will be called <u>reverse-convex</u> and the problem I a <u>reverse-convex minimization problem</u>.

Such problems arise, for example, when we wish to determine the minimum of a function in a region from which an open sphere about a point has been removed.

It is easy to show that even with a linear objective function such a problem may have a local minimum that is not a global minimum. The numerical methods proposed below, like all numerical methods for problems with infinitely many feasible points, can take into account only the local behavior of the function, and therefore these methods can at best be expected to determine a local minimum for a problem of the form I . The reasons why, in general, it is impossible to "solve" the problem I (i.e., obtain the global minimum) is thus not that the methods are faulty, but that the problem is in some sense not well-posed. From a physical standpoint, however, such problems are not unusual.

In order to establish iterative procedures for problems of type I, we first consider a technique for generating convex subsets of F. This is conveniently done by "linearization." That is, we define $W(y) := \{z \mid u(y) + u'(y)(z-y) \geq 0\}$, the "linearization" of the set U about the point y (here $u'(y)$ is the Jacobian of u evaluated at y); and we let $\Gamma(y) := S \cap W(y)$. For every $y \in F$ the set $\Gamma(y)$ has the three important properties: (1) $\Gamma(y)$ is convex and compact, (2) $\Gamma(y) \subset F$, and (3) $y \in \Gamma(y)$. (This was first pointed out by Rosen [6].) Property (2) is an immediate consequence of the convexity of u, and properties (1) and (3) are obvious. Note that by the results of section 1, the point-to-set mapping Γ is everywhere u.s.c.

Consider now the following sub-problem derived from the problem I:

$$R(y): \quad \text{minimize} \quad \varphi(z) \quad \text{subject to } z \in \Gamma(y) .$$

If $y \in F$, by property (1) above, $R(y)$ has a solution; by property (2), every point at which the minimum value is attained must be in F; and by property (3), if z^* solves $R(y)$, $\varphi(z^*) \leq \varphi(y)$. Of course, it is not likely that a solution of the sub-problem $R(y)$ can be obtained by numerical means unless the objective function has some convexity property. (For example, if φ is strictly quasi-convex [8], a local minimum for $R(y)$ will be a global minimum.) In many problems of interest the objective function will be linear, so that if S is a polytope (the intersection of a finite number of half-spaces), the problem $R(y)$ can be solved by linear programming (LP). In any event the following iterative scheme proposed by Rosen [6] is mathematically well defined:

Method A : (a) Choose an arbitrary $y_0 \in F$.

(b) Given y_i, let y_{i+1} be a solution of $R(y_i)$.

By the above discussion, method A yields a sequence of feasible points satisfying $\varphi(y_{i+1}) \leq \varphi(y_i)$, with strict inequality holding if y_i does not solve $R(y_i)$. Because F is compact, $\{y_i\}$ must have at least one accumulation point, and every accumulation point must lie in F. Note that in method A, unlike the iterative procedure in section 1, we do not require that $y_{i+1} = y_i$ if y_i solves $R(y_i)$. This restriction was included in section 1 merely to assure that $\{\varphi(y_i)\}$ satisfied $\varphi(y_{i+1}) \leq \varphi(y_i)$. By an immediate application of theorem 1.7 then, the following theorem holds:

<u>THEOREM 2.1:</u> If Γ is l.s.c. at an accumulation point y^* of a sequence $\{y_i\}$ generated by method A, then y^* solves $R(y^*)$.

There are several aspects of the previous theorem that warrant further discussion. The first point to be noted is that the compactness of F is only used to guarantee that the subproblems R(y) have solutions and that the sequence $\{y_i\}$ has at least one accumulation point. It follows that compactness can be replaced by those two hypotheses. In this case we might as well consider φ to be a real functional defined on a reverse-convex subset F of a Banach space, since the proof of the previous theorem was based solely on theorem 1.7. It should be pointed out that while theorem 1.1 gives a sufficient condition for l.s.c of Γ at y^* when $F \subset E^n$, sharper results have been obtained in [9]. In particular, if S is determined by constraints of the form $g(z) \geq 0$, where g is vector-valued and differentiable, it is sufficient that there be a point $z^* \in \Gamma(y^*)$ such that the gradients to the active constraints at z^* form a positively linearly independent set (i.e., no non-trivial non-negative combination of the vectors of the set vanishes). With regard to the conclusion of the theorem only one observation will be made here. (This topic is examined in some detail in [9].) If $\Gamma(y^*)$ satisfies some form of constraint qualification at y^* (which is the case, for example, if S is a polytope), then the fact that y^* solves $R(y^*)$ implies that the Kuhn-Tucker (K-T) first-order necessary conditions for a solution of I are satisfied at y^*. This is obvious, for at y^* the K-T conditions for the two problems $R(y^*)$ and I are identical.

If the function φ is differentiable on some open set containing F, we can construct the following sub-problem for each point $y \in F$:

$$L(y): \qquad \text{minimize} \quad \varphi'(y)z$$
$$\text{subject to} \quad z \in \Gamma(y) .$$

Of course, if φ is linear affine, the solutions to $L(y)$ coincide with the solutions of $R(y)$. However, even for the class of quasi-concave functions (which includes all linear affine functions), we have the crucial property that if y does not solve $L(y)$, then every solution $\bar{y}$ of $L(y)$ satisfies $\varphi(\bar{y}) < \varphi(y)$. This follows from the (differential) definition of quasi-concavity [8], which requires that $\varphi'(y)(\bar{y}-y) < 0$ imply $\varphi(\bar{y}) < \varphi(y)$. Consider now the following iterative method:

Method B:
(a) Choose an arbitrary $y_0 \in F$.
(b) Let $y_{i+1} = y_i$ if y_i solves $L(y_i)$; otherwise, let y_{i+1} be any solution of $L(y_i)$.

The following is an immediate consequence of theorem 1.6:

<u>THEOREM 2.2:</u> Let φ be quasi-concave and continuously differentiable on some open set containing F. If y^* is an accumulation point of a sequence $\{y_i\}$ generated by method B and Γ is continuous at y^*, then y^* solves $L(y^*)$.

A comparison of theorems 2.1 and 2.2 is in order. The former is valid for all continuous objective functions (although from a numerical standpoint we can apply method A only to objective functions with certain convexity properties), whereas the latter holds only for continuously differentiable quasi-concave functions (although method B is numerically feasible whenever φ is differentiable). Although the last theorem specifies that Γ be continuous at y^*, it follows from a result of section 1 that Γ is u.s.c. everywhere, so that only l.s.c. at y^* need be assumed or verified. In order to apply theorem 1.6 to prove the previous theorem, it is necessary that F be compact. As noted above, the compactness of F does not play so crucial a role in the proof of theorem 2.1. Finally, if we assume again that $\Gamma(y^*)$ satisfies some type of constraint qualification at y^*, we conclude that if y^* solves $L(y^*)$, then y^* satisfies the K-T conditions for problem I .

In the event that φ is continuously differentiable but not quasi-concave, it is still possible to obtain algorithms in which the objective is linearized and which have the required properties if we assume that φ is twice continuously differentiable. These are based upon the observation that the linear part of the objective function dominates in a region sufficiently close to the point linearized about. Let the constant $R > 0$ be chosen so that $\frac{1}{2}(z_2-z_1)^T \varphi''(z_1)(z_2-z_1) \leq R\,\|z_2-z_1\|^2$ for all $z_1, z_2 \in F$ (the norm is arbitrary but fixed for the remainder of the section). Given a point $y \in F$, let $M(y)$ be the set of solutions of $L(y)$. Let α be a fixed element of $(0, 1)$ and for $z \neq y$ define the real-valued function

$$K(y, z) := \text{minimum} \quad \{\alpha \cdot \varphi'(y)(y-z) \cdot \|z-y\|^{-1} \cdot R^{-1},\ \|z-y\|\} .$$

For the three algorithms below, it is assumed that the initial point y_0 is chosen arbitrarily from the feasible set F, and that y_{i+1} is taken to be y_i if y_i solves $L(y_i)$. Hence, we will specify the method for choosing y_{i+1} with the understanding that y_i does not solve $L(y_i)$. In the following algorithms, z_i^* is an arbitrary element of $M(y_i)$.

ALGORITHM C: Choose y_{i+1} to minimize $\varphi'(y_i)z$ over the set $\Gamma(y_i) \cap \{z \mid \|z-y_i\| \leq K(y_i, z_i^*)\}$.

ALGORITHM D: Let $y_{i+1} = y_i + K(y_i, z_i^*)\cdot(z_i^* - y_i)\cdot\|z_i^* - y_i\|^{-1}$.

ALGORITHM E: Choose an element θ from the fixed interval $[\beta, \gamma]$, where $0 < \beta \leq \gamma < 1$, and let $y_{i+1} = y_i + \theta^j \cdot (z_i^* - y_i)$, where j is the smallest non-negative integral exponent for which the inequality $\varphi(y_{i+1}) \leq \varphi(y_i) + (1-\alpha)\cdot\theta^j\cdot\varphi'(y_i)(z_i^* - y_i)$ is satisfied. (It will be shown that the previous inequality is satisfied for all sufficiently large j, so that the algorithm is well-defined.)

THEOREM 2.3: Let φ be twice continuously differentiable on some open set containing F, and let the sequence $\{y_i\}$ be generated by one of the three procedures above. If y^* is an accumulation point of $\{y_i\}$ at which Γ is continuous, then y^* solves $L(y^*)$.

PROOF: See Appendix.

In general, the three previous algorithms will yield three different points if applied to a given point. A typical situation is shown in Figure 2, where the points C, D, and E correspond to the application of algorithms C, D, and E respectively. The dotted lines represent level lines of the linearized objective function $\varphi'(y_i)z$, and the square in the interior of $\Gamma(y_i)$ represents the set, $\{z \mid \|z-y_i\| \leq K(y_i, z_i^*)\}$. The figure illustrates a case in which we have chosen to work with a norm whose level surfaces are the surfaces of similar polyhedra. When these types of norms are used, and, in addition, S is a polytope, it follows that in order to obtain y_{i+1} via algorithm C from y_i and z_i^*, we need only solve an LP problem. Hence, in this case we would solve two LP problems in order to move from y_i to y_{i+1} using algorithm C. On the other hand, regardless of the norm used, when S is a polytope, only one LP problem must be solved when algorithm D is used to obtain a successor to y_i. However, for both algorithms C and D an estimate on the upper bound of the norm of the Hessian matrix $\varphi''(z)$ is needed, and this may not be easily obtained. For algorithm E this estimate is unnecessary, and y_{i+1} is obtained by y_i by solving one LP problem (assuming again that S is a polytope) and performing a finite number of evaluations of φ.

In the case that S is a polytope, method B and the method corresponding to algorithm C are special cases of the MAP method of Griffith and Stewart [10]. For the classes of minimization problems for which they are intended, the former two methods resolve the previously unsolved problem of step-size limits for the MAP method. Algorithms D and E can be contrasted with the well-known Frank-Wolfe algorithm in the special case when F is a polyhedron. (This will occur if S is a polytope and u is linear, and will mean that $\Gamma(y) = F$ for all y .) The Frank-Wolfe algorithm consists of choosing y_{i+1} to be a point on the line segment connecting y_i and z_i^* which satisfies $\varphi(y_{i+1}) \leq \varphi(y_i) + (1-\alpha)[\varphi(y_i^*) - \varphi(y_i)]$, where y_i^* is a point which minimizes φ on that line segment and $\alpha \in [0, 1)$. Algorithms D and E require no knowledge of the minimum of the function φ on line segments, and hence enjoy something of a theoretical advantage over the Frank-Wolfe scheme.

3. Applications to Other Mathematical Programming Algorithms

In this section we shall indicate how the results of section 1 may be applied to a number of well-known algorithms of mathematical programming.

3.1 Unconstrained Minimization Methods.

In the notation of section 1 , let $f(z, y) = \varphi(z)$ and $\Omega(y) = \{y + \lambda \cdot D(\varphi(y)) \mid \lambda \geq 0\}$, where the composite function $D(\varphi)$ (which can be thought of as a direction-assigning function) is continuous for all continuously differentiable φ and has the property that $\varphi'(y) D(\varphi(y)) \leq 0$ with equality if and only if $\varphi'(y) = 0$. (When $D(\varphi(y))$ is chosen to be $(-\varphi'(y))^T$, the corresponding algorithm (see section 1) is the method of steepest descent. If φ is twice continuously differentiable and has a positive definite Hessian matrix at each point, then we may choose $D(\varphi(y)) = -[\varphi'(y)\varphi''^{-1}(y)]^T$. The corresponding algorithm is then a modification of the Newton-Raphson second order method.) It is clear that Ω is everywhere l.s.c. and that the iterates have the required monotonicity property. We thus conclude that an accumulation point y^* of such a method must solve the problem: minimize $\varphi(z)$ subject to $z \in \Omega(y^*)$. This implies that $\varphi'(y^*) = 0$, and if φ is convex, y^* must be the global minimum of the unconstrained minimization problem.

3.2 Feasible Direction Methods.

Topkis and Veinott [12] recently studied the properties of a general feasible direction algorithm which contains as special cases the feasible direction methods of Zoutendijk [13], the Frank-Wolfe method [11], and second-order feasible direction methods. We will show below how the same general algorithm can be studied by the techniques of section 1. Again we consider a general minimization problem of the form I, but we will assume here that the set $U = \{z \mid u(z) \geq 0\}$ is convex (rather than reverse-convex as assumed in section 2). All other assumptions on the feasible set F, including compactness, are assumed to hold. We define the set $\Omega(y)$ to be those pairs (v, z) satisfying

$$v \geq \varphi'(y)(z-y) + \frac{1}{2}(z-y)^T H(y)(z-y) ,$$

$$v \geq -[u_i(y) + u_i'(y)(z-y)] \quad \text{for all } i ,$$

and $z \in S \cap (B + y)$, where H is a continuous mapping from E^n into the set of all positive semi-definite $n \times n$ matrices and B is a compact convex neighborhood of the origin. Letting $\mu(y) :=$ minimum $\{v \mid (v, z) \in \Omega(y)\}$, the iterative procedure proposed by Topkis and Veinott is as follows:

(a) Choose an arbitrary $y_0 \in F$

(b) Given y_i, let y_i^* be chosen so that $(v_i^*, y_i^*) \in \Omega(y_i)$ and $v_i^* = \mu(y_i)$; if $\mu(y_i) = 0$, let $y_{i+1} = y_i$, and if not, let y_{i+1} be a point in the intersection of F with the line segment connecting y_i and y_i^* such that $\varphi(y_{i+1}) \leq \varphi(z)$ for all z in the intersection.

It is shown in the Appendix that the mapping Ω as defined above is continuous on S and that $\varphi(y_{i+1}) < \varphi(y_i)$ if $\mu(y_i) < 0$. By a slight modification of the proof of theorem 1.6, it follows that a limit point y^* of the iterative procedure just described has the property that $\mu(y^*) = 0$. If some form of constraint qualification holds at y^* (for a particular case, see [12]), the relation $\mu(y^*) = 0$ implies that the Kuhn-Tucker necessary conditions for a solution of problem I must be satisfied at y^*. The Kuhn-Tucker conditions are also sufficient for optimality when φ is pseudo-convex and the constraint functions are quasi-concave (see Mangasarian [14]).

4. Generalizations

Because of such factors as finite arithmetic and rounding errors, there is little hope of obtaining exact analytic solutions to optimization problems on digital computers. One can expect at best very good approximations to the true solutions. In the theory developed in the preceding sections, however, the availability of exact solutions at each iteration was assumed. We will now show how theorems 1.6 and 1.7, upon which most of the results of this paper are based, can be strengthened to provide for a certain type of approximate solution. (This type of approximation was considered by Dem'yanov and Rubinov [15] in a paper dealing with a convex programming method in Banach space. Other approximations, such as the class considered by Topkis and Veinott [12], can be handled in a similar manner.)

Let α be a fixed element of the open interval (0, 1), and, using the notation and assumptions introduced for the statement of theorem 1.6, let the sequence $\{\bar{y}_i\}$ be constructed in the following manner:

(a) Choose an arbitrary $y_0 \in G$.

(b) Let $\bar{y}_{i+1} = \bar{y}_i$ if $\bar{y}_i \in M(\bar{y}_i)$; otherwise let $\bar{y}_{i+1}$ be an element of $\Omega(\bar{y}_i)$ satisfying $\varphi(\bar{y}_i) - \varphi(\bar{y}_{i+1}) \geq \alpha \cdot (\varphi(\bar{y}_i) - \varphi(y_i^*))$, where $y_i^* \in M(\bar{y}_i)$.

Roughly this means that at each iteration at least a fixed fraction of the theoretically possible decrease in φ is attained.

<u>THEOREM 4.1</u>:

If $\{\bar{y}_i\}$ and $\{y_i^*\}$ are contained in sequentially compact sets and $\bar{y}$ is an accumulation point of $\{\bar{y}_i\}$ at which Ω is continuous, then $\bar{y} \in M(\bar{y})$.

Proof: As in the proof of theorem 1.6, we assume that the conclusion is false, and show a contradiction. It follows from the assumptions preceding 1.6 that $\varphi(y') < \varphi(\bar{y})$ for all $y' \in M(\bar{y})$. Now let subsequences $\{\bar{y}_{n_i}\}$, $\{\bar{y}_{n_i+1}\}$, and $\{y^*_{n_i}\}$ be chosen so that $\bar{y}_{n_i} \to \bar{y}$, $\bar{y}_{n_i+1} \to \tilde{y}$, and $y^*_{n_i} \to y^*$. It follows that $y^* \in M(\bar{y})$ and that

$$0 < \varphi(\bar{y}) - \varphi(y^*) = \lim (\varphi(\bar{y}_{n_i}) - \varphi(y^*_{n_i}))$$

$$\leq \alpha^{-1} \cdot \lim (\varphi(\bar{y}_{n_i}) - \varphi(\bar{y}_{n_i+1})) = 0 \text{, which cannot hold.}$$

By an analogous modification of the proof of theorem 1.7, we obtain:

<u>THEOREM 4.2</u>:

If $\varphi(z) = f(z, y)$ and $\bar{y}$ is an accumulation point of $\{y_i\}$ at which Ω is l.s.c., then $\bar{y} \in M(\bar{y})$.

Another computational aspect of algorithms that can be easily dealt with by the techniques of this paper is that of accelerating convergence by periodically taking a step in a direction other than that prescribed by the basic algorithm being used or taking slightly larger or slightly smaller steps than those prescribed. (The validity of procedures so modified has also been discussed by Topkis and Veinott [12].) It should be observed that the proofs of theorems 1.6 and 1.7 depended only on the monotonicity of the sequence $\{\varphi(y_i)\}$ and the fact that y^* was the limit of a subsequence $\{y_{n_i}\}$ whose successor points were constructed by an algorithm with certain specified properties. Thus, if, with the goal of accelerating convergence, an algorithm without those properties is used periodically, we can conclude nevertheless that convergence of the iterates to a point y^* at which Ω is continuous (or l.s.c. in the case of theorem 1.7) implies that $y^* \in M(y^*)$.

A further extension of theorems 1.6 and 1.7 can be made if we note that the proofs still go through if we assume only that $\{\varphi(y_i)\}$ converges (i.e., it need not be monotonic) and that the strict monotonicity property holds at y^* (instead of everywhere). This extension is useful when Kelley's cutting-plane algorithm [16] is analyzed by the techniques of section 1 .

APPENDIX

The following property of sequences in normed spaces will be needed in the proof of theorem 1.1.

LEMMA: If $z_i \to z$ as $i \to \infty$ and $z_{ij} \to z_i$ and $j \to \infty$ $(i = 1, 2, \ldots)$, then there exist n_j $(j = 1, 2, \ldots)$ such that $z_{n_j j} \to z$ as $j \to \infty$.

PROOF: Let $N(1)$ be chosen such that $\| z_i - z \| < 1$ for $i \geq N(1)$, and let $N'(1)$ be chosen such that $\| z_{N(1)_j} - z_{N(1)} \| < 1$ for $j \geq N'(1)$. Suppose now we have chosen $N(1), N(2), \ldots, N(k)$ and $N'(1), N'(2), \ldots, N'(k)$. Choose $N(k+1)$ and $N'(k+1)$ so that $\| z_i - z \| < 1/k+1$ for $i \geq N(k+1)$, $\| z_{N(k+1)_j} - z_{N(k+1)} \| < 1/k+1$ for $j \geq N'(k+1)$, and $N'(k+1) > N'(k)$. Let $N(0) = 1$ and define $n_j = N(\ell)$ when $N'(\ell) \leq j < N'(\ell+1)$. It is easily verified that the sequence so defined satisfies $z_{n_j j} \to z$ as $j \to \infty$.

<u>PROOF OF THEOREM 1.1</u>: We will first show that the linear independence hypothesis is equivalent to assuming that there exists a point z' such that $f(z') > 0$, $u(y^*) + u'(y^*)(z' - y^*) > 0$, $v(y^*) + v'(y^*)(z' - y^*) = 0$, and that the Jacobian matrix $v'(y^*)$ has full row rank. For, we may choose a vector d such that $v'(y^*)d = 0$ and such that the inner product of d with each gradient to an active inequality constraint function at z^* is positive. It is now easily

seen that a suitable choice of z' is $z^* + \theta d$, where θ is a sufficiently small positive scalar. (Since $v(y^*) + v'(y^*)(z^* - y^*) = 0$, the linear independence hypothesis implies that $v'(y^*)$ has full row rank.)

Now partition the variable z into the variables s and t (with values s' and t' at z') so that the function $\overline{v}$ defined by $\overline{v}(s, t, y) = v(y) + v'(y)(z - y)$ has a nonsingular Jacobian with respect to s at the point $(z', y^*) = (s', t', y^*)$. It follows from the implicit function theorem that there exists a neighborhood N of (t', y^*) and a differentiable function h defined on N with the properties that $h(t', y^*) = s'$ and $\overline{v}(h(t, y), t, y) = 0$ for $(t, y) \in N$. Without loss of generality we can assume that N was chosen small enough so that all of the inequality constraints involved in defining $\Gamma m(y)$ are satisfied by $(h(t, y), t)$ when $(t, y) \in N$. (This follows from elementary continuity arguments.) Hence if $\{y_i\}$ is any sequence converging to y^*, it follows that for i sufficiently large (say $i \geq m$), we have $(t', y_i) \in N$, so that $\overline{v}(h(t', y_i), t', y_i) = 0$, and hence the equality constraints involved in defining $\Gamma m(y_i)$ are also satisfied at the point $(h(t', y_i), t') =: z_i$. The sequence $\{z_i\}$ so defined for $i \geq m$ thus has the property that $z_i \in \Gamma m(y_i)$ and $z_i \rightarrow z'$. To complete the proof of l.s.c. at y^* we must prove the existence of a similar sequence for each $z \in \Gamma m(y^*)$. In order to do this, we first note that $\Gamma m(y^*)$ is a convex set, so that given any $z \in \Gamma m(y^*)$, the line segment connecting z and z' lies in $\Gamma m(y^*)$. Moreover, since $z \in S \cap \{z \mid u(y^*) + u'(y^*)(z - y) \geq 0\}$ and

$$z' \in \text{ int } S \cap \{z \mid u(y^*) + u'(y^*)(z - y) > 0\} =: \overline{S}, \text{ it}$$

follows from a well-known theorem on convex sets (see, for example [2]) and a simple computation that all points on that line segment with the possible exception of z also lie in $\overline{S}$. But at each point in $\overline{S} \cap \Gamma m(y^*)$ we can construct the sequence required in the definition of l.s.c. by exactly the same method used for z'. Letting $z'_i := (1/i)\, z' + (1-1/i)z$ and performing such a construction for $i = 1, 2, \ldots,$ we obtain a sequence of sequences from which, by the preceding lemma, we can construct a sequence converging to z and satisfying the requirements in the definition of l.s.c. . This completes the proof of l.s.c. at y^*.

Now for y sufficiently to y^* we have previously noted that the point $(t', h(t', y))$ lies in $\Gamma m(y)$ and satisfies all of the inequality constraints strictly. Since there also exists a neighborhood of y^* in which the Jacobian $v'(y)$ has full row rank, it follows that for all y in some neighborhood of y^* the point $(t', h(t', y))$ has the same properties with respect to $\Gamma m(y)$ that z' had with respect to $\Gamma m(y^*)$. Hence the proof of l.s.c. of Γm at such y may be carried out in the same manner.

The next example illustrates that the compactness hypothesis cannot be deleted in theorem 1.6.

EXAMPLE: Let $G = R = [-2, -1\frac{1}{2}] \cup [0, \frac{1}{2}] \cup [2, +\infty)$,

$$\Omega(y) = \begin{cases} \{y\} & \text{if } y \in [-2, -1\frac{1}{2}] \\ \{-2\} & \text{if } y = 0 \\ \{-2+y, 1/y\} & \text{if } y \in (0, \frac{1}{2}] \\ \{1/2y\} & \text{if } y \in [2, +\infty) \end{cases},$$

$$f(z, y) = \begin{cases} z + 2 & \text{if } z \in [-2, -1\frac{1}{2}] \\ 0 & \text{if } z \in [0, \frac{1}{2}] \\ 1/z & \text{if } z \in [2, +\infty) \end{cases},$$

and

$$\varphi(y) = \begin{cases} y + 2 & \text{if } y \in [-2, -1\frac{1}{2}] \\ y + 1 & \text{if } y \in [0, \frac{1}{2}] \\ 1 + 2/3y & \text{if } y \in [2, +\infty) \end{cases}.$$

It is easily verified that with the above definitions the conditions stated prior to theorem 1.6 are satisfied, that Ω is continuous on G, and that f and φ may be extended to continuous functions on E^2 and E^1 respectively. Suppose that we choose $y_0 = \frac{1}{2}$. It may be verified that $M(y_0) = \{-1\frac{1}{2}, 2\}$, so that we can choose $y_1 = 2$. Since $\Omega(y_1) = \{\frac{1}{4}\}$, it follows that $y_2 = \frac{1}{4}$. Continuing in a similar fashion, the sequence of iterates $\{\frac{1}{2}, 2, \frac{1}{4}, 4, \frac{1}{8}, \ldots\}$ is obtained. However, the accumulation point 0 does not belong to $M(0) = \{-2\}$.

<u>PROOF OF THEOREM 2.3</u>: We will first show that all three algorithms have the strict monotonicity property. Using a second order Taylor expansion and the definition of R, we obtain for $z \in F$ the inequality $\varphi(z) \leq \varphi(y_i) + \varphi'(y_i)(z - y_i) + R \cdot \|z - y_i\|^2$. If $\|z - y_i\| \leq \delta \cdot K(y_i, z_i^*)$, this becomes

$$\begin{aligned}\varphi(z) &\leq \varphi(y_i) + \varphi'(y_i)(z - y_i) + R \cdot \delta^2 \cdot K^2(y_i, z_i^*) \\ &\leq \varphi(y_i) + \varphi'(y_i)(z - y_i) \\ &\qquad - \delta^2 \cdot \alpha \cdot \varphi'(y_i)(z_i^* - y_i) \cdot \|z_i^* - y_i\|^{-1} \cdot K(y_i, z_i^*).\end{aligned}$$

If we denote by y' the point generated by applying algorithm D to y_i, we have $y' - y_i = (z_i^* - y_i) \cdot \|z_i^* - y_i\|^{-1} \cdot K(y_i, z_i^*)$, and the inequality reduces to $\varphi(z) \leq \varphi(y_i) + \varphi'(y_i)(z - y_i) - \delta^2 \cdot \alpha \cdot \varphi'(y_i)(y' - y_i)$. Three cases will now be considered: (1) if $z = y'$, choose $\delta = 1$, and the inequality becomes $\varphi(z) \leq \varphi(y_i) + (1 - \alpha)\varphi'(y_i)(y' - y_i)$; (2) if z is generated by algorithm C, choose $\delta = 1$, and it follows from $\varphi'(y_i)z \leq \varphi'(y_i)y'$ that $\varphi(z) \leq \varphi(y_i) + \varphi'(y_i)(y' - y_i) - \alpha \cdot \varphi'(y_i)(y' - y_i) = \varphi(y_i) + (1 - \alpha)\varphi'(y_i)(y' - y_i)$; and (3) if $z = y_i + \omega \cdot (y' - y_i)$, where $0 \leq \omega \leq 1$, choose $\delta = \omega$, yielding

$$\begin{aligned}\varphi(z) &\leq \varphi(y_i) + \omega \cdot \varphi'(y_i)(y' - y_i) - \omega^2 \cdot \alpha \cdot \varphi'(y_i)(y' - y_i) \\ &= \varphi(y_i) + (1 - \omega\alpha) \cdot \omega \cdot \varphi'(y_i)(y' - y_i) \\ &\leq \varphi(y_i) + (1 - \alpha) \cdot \omega \cdot \varphi'(y_i)(y' - y_i).\end{aligned}$$

By the analysis in case (3), it is easily seen that in algorithm E the relation $\varphi(y_{i+1}) \leq \varphi(y_i) + (1 - \alpha) \cdot \rho^j \cdot \varphi'(y_i)(z_i^* - y_i)$, where $y_{i+1} = y_i + \rho^j \cdot (z_i^* - y_i)$

is satisfied if $\theta j \cdot \| z_i^* - y_i \| \leq K(y_i, z_i^*)$, proving that the algorithm is well-defined. Moreover, since $\beta \leq \theta$, the point z generated by algorithm E must satisfy $\varphi(z) \leq \varphi(y_i) + (1 - \alpha) \cdot \beta \cdot \varphi'(y_i)(y' - y_i)$. For all three algorithms then, $\{\varphi(y_i)\}$ is a non-increasing sequence.

Let y^* be an accumulation point of $\{y_i\}$ at which the point-to-set mapping Γ is continuous. Choose subsequences $\{y_{n_i}\}$, $\{y_{n_i+1}\}$, and $\{z^*_{n_i}\}$ such that $y_{n_i} \rightarrow y^*$ and the latter two are convergent with limit points $\bar{y}$ and z^* respectively. As a consequence of theorem 1.5, z^* is a solution of $L(y^*)$. If we now suppose that y^* does not solve $L(y^*)$, then $K(y^*, z^*) > 0$. For algorithms C and D we thus have

$$
\begin{aligned}
\varphi(\bar{y}) = \lim \varphi(y_{n_i+1}) &\leqq \lim [\varphi(y_{n_i}) + (1 - \alpha) \cdot \varphi'(y_{n_i})(y_{n_i+1} - y_{n_i})] \\
&= \varphi(y^*) + (1 - \alpha) \cdot \varphi'(y^*)(\bar{y} - y^*) \\
&= \varphi(y^*) + (1 - \alpha) \cdot \varphi'(y^*)(z^* - y^*) \cdot \| z^* - y^* \|^{-1} \cdot K(y^*, z^*) \\
&< \varphi(y^*) .
\end{aligned}
$$

This is impossible, however, since $\{\varphi(y_i)\}$ is a non-increasing sequence. By inserting the factor β in the appropriate places, we can prove the conclusion for algorithm E .

(Alternative proofs have been constructed (see Meyer [9]) for methods C and D by establishing the strict monotonicity property and the continuity of certain point-to-set mappings. In this way the conclusion is obtained as a direct consequence of theorem 1.6, but at the expense of increasing the complexity of the proof.)

PROOF OF ASSERTIONS IN SECTION 3.2:

By using the continuity of the terms involved, it is easily shown that Ω is everywhere u.s.c. To prove l.s.c. on S, we first observe that the point-to-set mapping defined by $\Omega'(y) = S \cap (B + y)$ is l.s.c. on S. This is seen by noting that interior points of $B + y$ that lie in S also lie in $B + \bar{y}$ for $\bar{y}$ sufficiently close to y, and that a boundary point of $B + y$ that lies in S is the limit of interior points of $B + y$ contained in S. Now let (v, z) be an arbitrary point of $\Omega(y)$ and let $\{y_n\}$ be a sequence of points in S converging to y. By the preceding argument, there exists a sequence of points $\{z_n\}$ with $z_n \in \Omega'(y_n)$ converging to z. It is clear that a sequence $\{v_n\}$ converging to v can now be chosen so that $(v_n, z_n) \in \Omega(y_n)$, completing the proof of l.s.c.

To simplify notation, we will drop the subscripts in the following proof of the monotonicity property asserted in section 3.2. Suppose that there exists a point $(v^*, z^*) \in \Omega(y)$ with $v^* < 0$. We will show that for sufficiently small positive λ, the point $\bar{z} = y + \lambda(z^* - y)$ belongs to F and satisfies $\varphi(\bar{z}) < \varphi(y)$. It is clear that $\bar{z} \in S \cap (B + y)$ for $\lambda \in [0,1]$, so to prove feasibility we need only show that $u(\bar{z}) \geq 0$. If $u_i(y) > 0$, then clearly $u(\bar{z}) > 0$ for λ sufficiently small; and if $u_i(y) = 0$, then $0 > v^* \geq -[u_i(y) + u_i'(y)(z^* - y)]$ implies $u_i'(y)(z^* - y) > 0$, and again it is true that $u_i(\bar{z}) > 0$ for sufficiently small positive λ. Since $H(y)$ is assumed positive semi-definite, $0 > v^* \geq \varphi(y)(z^* - y) + \frac{1}{2}(z^* - y)\, H(y)(z^* - y)$ implies $0 > \varphi'(y)(z^* - y)$, and the required result follows.

BIBLIOGRAPHY

1. F. Hausdorff, Set Theory, Chelsea, New York, 1962.

2. C. Berge, Topological Spaces, Macmillan, New York, 1963.

3. E. J. McShane and T. A. Botts, Real Analysis, Van Nostrand, Princeton, 1959.

4. M. Q. Jacobs, "Some Existence Theorems for Linear Optimal Control Problems," SIAM J. on Control, 5 (1967), pp. 418-437.

5. G. Debreu, Theory of Value, Cowles Foundation Monograph #17, John Wiley and Sons, New York, 1959.

6. J. B. Rosen, "Iterative Solution of Nonlinear Optimal Control Problems," SIAM J. on Control, 4 (1966), pp. 223-244.

7. W. I. Zangwill, "Convergence Conditions for Nonlinear Programming Algorithms," Working Paper No. 197, Center for Research in Management Science, University of California, Berkeley, 1966.

8. J. Ponstein, "Seven Kinds of Convexity," SIAM Review, 9 (1967), pp. 115-119.

9. R. R. Meyer, "The Solution of Non-convex Optimization Problems by Iterative Convex Programming," Ph.D. Thesis, The University of Wisconsin, Madison, 1968.

10. R. E. Griffith and R. A. Stewart, "A Nonlinear Programming Technique for the Optimization of Continuous Processing Systems," Management Science, 7 (1961), pp. 379-392.

11. M. Frank and P. Wolfe, "An Algorithm for Quadratic Programming," Naval Res. Logist. Quart., 3 (1956), pp. 95-110.

12. D. M. Topkis and A. F. Veinott, Jr., "On the Convergence of Some Feasible Direction Algorithms for Nonlinear Programming," SIAM J. on Control, 5 (1967), pp. 280-294.

13. G. Zoutendijk, Methods of Feasible Directions, Elsevier, New York, 1960.

14. O. L. Mangasarian, "Pseudo-convex Functions," SIAM J. on Control, 3 (1965), pp. 281-290.

15. V. F. Dem'yanov and A. M. Rubinov, "The Minimization of a Smooth Convex Functional on a Convex Set," SIAM J. on Control, 5 (1967), pp. 268-279.

16. J. E. Kelley, "The Cutting Plane Method for Solving Convex Programs," J. SIAM, 8 (1960), pp. 703-712.

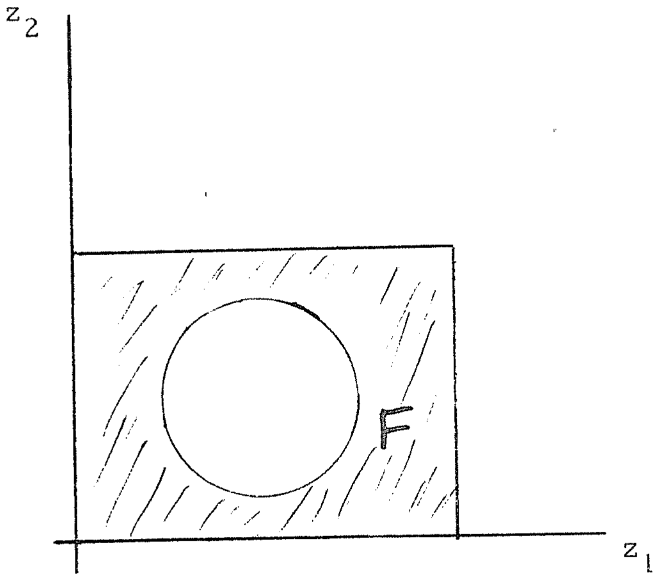

Figure 1 A Reverse-Convex Set

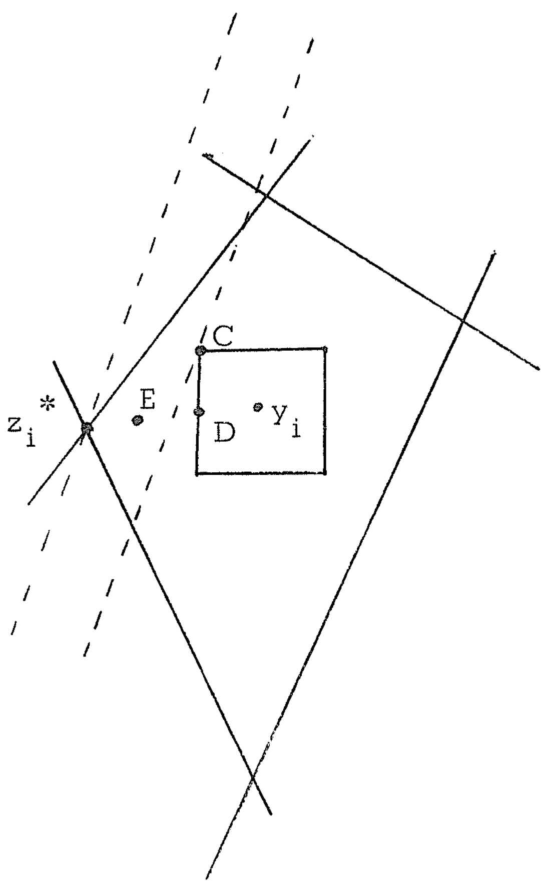

Figure 2 Successor Points