COMPUTER SCIENCES DEPARTMENT
University of Wisconsin-Madison

A Comparison of the Forcing Function and Point-to-Set Mapping Approaches to Convergence Analysis

by

R. R. Meyer

Computer Sciences Technical Report #259

October 1975

A Comparison of the Forcing Function and Point-to-Set Mapping Approaches to Convergence Analysis

by

R. R. Meyer

Computer Sciences Department
University of Wisconsin
Madison, Wisconsin 53706

Received 8/20/75

October 1975

Abstract

A "forcing function" approach is developed for the analysis of convergence properties of "monotonic" mathematical programming algorithms. This approach differs from rather more traditional analyses based on point-to-set mappings in that it does not require point-to-set mapping concepts. A comparison is given between the forcing function and point-to-set mapping approaches that indicates that they are essentially mathematically equivalent for two major categories of algorithms, but that only the forcing function approach is readily extended to a third category of algorithms involving anti-jamming parameters.

CONTENTS

Page

1. Introduction 1

2. The Forcing Function Approach. 3

3. Relationships between Forcing Functions and Point-to-Set Mappings. 14

4. Convergence Theorems for Point-to-Set Mappings 21

5. Methods Involving Anti-Jamming Parameters. 24

6. Conclusions. 28

References. 29

Appendix. 31

1. Introduction

The point-to-set mapping approach to the qualitative analysis of convergence of mathematical programming algorithms is well-known, having been considered and promoted in many papers and books. (The reader entirely unfamiliar with the field can get a "feel" for the area by consulting [5], [7], [8], [9], [13] or [17] or the excellent survey article by Hogan [4].) On the other hand, analyses of convergence that do <u>not</u> rely on properties of point-to-set mappings have also been numerous (see, for example, [2], [6], [10], [11], [15], [18]), but have not considered algorithms at the level of generality of the point-to-set mapping approach. This report proposes a "forcing function" approach to convergence analysis that not only is more general than the best-known point-to-set mapping results, but moreover, is often easier to apply and may be readily extended to handle a class of algorithms that do not appear to be amenable to analysis by a straightforward point-to-set mapping approach. (These are constrained optimization methods involving "anti-jamming" parameters.) The forcing function approach has the further pedagogical advantage of requiring only certain continuity properties of real-valued functions rather than of point-to-set mappings.

In the results to be obtained below, the sequence of iterates $\{x_i\}$ should be thought of as resulting from the application of an iterative algorithm. Such sequences will always be assumed to be contained in a <u>closed</u> set $G \subseteq \mathbb{R}^n$. Unless otherwise specified,

the domain of all functions and mappings considered below will be G. (For unconstrained optimization algorithms an appropriate G might be $\mathbb{R}^n$, whereas for constrained optimization, the set G is often taken as the feasible region. Many of the results below also remain valid when G is a closed subset of an arbitrary topological space, and this will be pointed out when it is the case.)

The notation $q_i \overset{I}{\to} q$ is to be understood to mean that I is an infinite increasing sequence of non-negative integers and that $\lim\limits_{\substack{i\to\infty \\ i\in I}} q_i$ is q.

2. The Forcing Function Approach

Roughly speaking, the <u>non-negative</u> function δ to be considered below will play the role of an "optimality indicator" in the class of algorithms to be described, in that an iterate x_i may be repeated only if $\delta(x_i) = 0$. When $\delta(x_i) > 0$, the next iterate, x_{i+1}, will be required to have a "value improvement" (in terms of a particular function ϕ to be introduced below) of at least $\delta(x_i)$.

The term "forcing function approach" is used since, on the one hand, δ "forces" an improvement if $\delta(x_i) > 0$, and, on the other hand, convergence of a sequence of function values of ϕ will be seen to "force" the sequence $\{\delta(x_i)\}$ to converge to 0. (The term "forcing function" is used by Ortega and Rheinboldt [11] to describe a related but slightly different property. Specifically, they define a mapping $\sigma : [0,\infty) \to [0,\infty)$ to be a forcing function if, for any sequence $\{t_k\} \subset [0,\infty)$, the property $\sigma(t_k) \to 0$ implies $t_k \to 0$; they then analyze a family of optimization methods in which $\delta(x_i)$ is equal to a forcing function of a certain scalar function of x_i. See the Appendix for further comparisons.)

Let $\delta : G \to R^1_+$ and define

(2.1) $\Omega^L \equiv \{x \mid \exists \{y_i\} \subset G \text{ with } y_i \to x, \delta(y_i) \to 0\}$,

(2.2) $\Omega^* \equiv \{x \mid x \in G, \delta(x) = 0\}$,

(2.3) $\Omega^+ \equiv \{x \mid x \in G, \delta(x) > 0\}$.

Since δ is assumed to be non-negative, note that $G = \Omega^* \cup \Omega^+$.

Thinking of δ as an "optimality indicator" that is 0 at points satisfying some optimality condition, it is clearly desirable for an iterative algorithm to have the property that its iterates must converge to a point in Ω^*. In order to achieve this strong result (given as Corollary 2.5 below) a number of hypotheses are needed. By assuming only some of the hypotheses of Corollary 2.5, however, weaker results that are of some interest in themselves are obtained, and we shall first develop these weaker results.

<u>Lemma 2.1</u>: Let $\phi : G \to R^1$ be a function that is <u>lower semi-continuous</u> (l.s.c.) on G, and let $\{x_i\}$ be a sequence with the property that

(2.4) $\phi(x_i) - \phi(x_{i+1}) \geq \delta(x_i) \quad (i = 0,1,2,\ldots)$

If $\{x_i\}$ has an accumulation point $\bar{x}$, then $\delta(x_i) \to 0$ and $\bar{x} \in \Omega^L$.

<u>Proof</u>: If $\bar{x}$ is an accumulation point of $\{x_i\}$, it follows from (2.4), the non-negativity of δ, and the l.s.c. of ϕ, that $\phi^* \equiv \lim \phi(x_i)$ exists and that $\phi(x_i) \geq \phi(\bar{x})$ for all i. Thus $\phi(x_i) - \phi(x_{i+1}) \to 0$, $\delta(x_i) \to 0$, and if I is the index set corresponding to the subsequence of $\{x_i\}$ converging to $\bar{x}$, then $x_i \overset{I}{\to} \bar{x}$ and $\delta(x_i) \overset{I}{\to} 0$, proving that $\bar{x} \in \Omega^L$. □

It might be noted that the sole continuity hypothesis, namely, l.s.c. of ϕ on G, could be replaced by the hypothesis that <u>ϕ is bounded from below on G</u>, since the l.s.c. of ϕ is only needed to establish convergence of $\{\phi(x_i)\}$. However, in most applications the corresponding ϕ is at least continuous, and often continuously differentiable, so the l.s.c. hypothesis appears preferable to the boundedness hypothesis. Since no special properties of $\mathbb{R}^n$ were used, not that Lemma 2.1 remains true if G is a subset of some topological space. The same observation also applies to Lemma 2.2 and Theorem 2.3 below.

It might also be noted that Lemma 2.1 is similar in some respects to some results of Eaves and Zangwill [3]. They develop a theory of <u>cutting plane</u> algorithms by assuming that the <u>distance</u> between an iterate and certain prior iterates is bounded from below by a non-negative "separator" function δ, that has the property that if $z_i \to z$ and $\delta(z_i) \to 0$, then z must be in what is termed "the goal set". In Lemma 2.1, δ is a lower bound for the <u>change</u> in an <u>arbitary</u> function ϕ, and Ω^L itself plays the role of the "goal" set, rather than being a subset of a "goal" set that is never characterized. Since convergence to a point outside of Ω^L is also impossible under the hypotheses made by Eaves and Zangwill, it seems inappropriate to describe any point outside of Ω^L as being a "goal" of the algorithm.

A comparison of Lemma 2.1 with the more closely related results

of Zangwill [17] and Polak [13] is given the Appendix. Again the conclusion that the accumulation points must lie in Ω^L turns out to be a sharper result than membership in a so-called "solution" or "desirable" set.

In most specific applications, the function δ turns out to have certain continuity properties that allow a strengthening of the conclusion of Lemma 2.1. In particular, the weak continuity property that we will now introduce turns out to be satisfied by most optimality indicators that arise in practice (the only significant exception seems to arise from optimality indicators associated with certain feasible direction methods; this point will be taken up in section 5):

A scalar-valued function is said to be <u>null-continuous</u> or C_0 <u>at a point z</u> if the existence of a sequence $\{y_i\}$ with $y_i \to z$ and $\omega(y_i) \to 0$ implies $\omega(z) = 0$. The function is said to be C_0 on a set if it is C_0 at each point of the set. (as will be seen, this continuity concept for scalar functions will replace point-to-set mapping continuity properties in the convergence theorems to be developed. From a pedogogical standpoint, it also appears preferable to introduce convergence analysis through the use of continuity properties of functions rather than continuity properties of point-to-set mappings, since students often have difficulties in obtaining a feeling for point-to-set mappings).

Note that null-continuity of δ on G is a weaker property than lower semi-continuity of δ on G (assuming $G \neq \Omega$), but a stronger property than lower semi-continuity on Ω^* . (It is, in fact, <u>equivalent</u> to the relation $\Omega^L = \Omega^*$ and also equivalent to l.s.c. of δ on Ω^L .) For many well-known constrained optimization algorithms, the corresponding optimality indicator is <u>not</u> lower semi-continuous on all of G , but is a function for which $\Omega^L = \Omega^*$ may be established.

(It might also be noted that if δ is non-negative and null-continuous on G , then there exists a function δ_L defined on G such that (1) $0 \leq \delta_L(x) \leq \delta(x)$ for all $x \in G$, (2) δ_L is l.s.c. on G , and (3) $\{x | \delta_L(x) = 0\} = \Omega^*$. In fact, δ_L may be defined at each $x \in G$ by the equation $\delta_L(x) \equiv \inf \{\theta | \exists \{y_i\}$, $y_i \to x$, $\delta(y_i) \to \theta\}$, and properties (1), (2), and (3) are easily verified. From the standpoint of application, however, null-continuity is a weaker requirement than l.s.c. and one that may often be verified more easily than l.s.c.)

As an immediate consequence of Lemma 2.1 and the property that $\Omega^L = \Omega^*$ when δ is null-continuous we have:

<u>Lemma 2.2</u>: Let ϕ be l.s.c. on G , let δ be a C_0 function on G , and let $\{x_i\}$ be a sequence satisfying (2.4). If $\{x_i\}$ has an accumulation point $\bar{x}$, then $\bar{x}$ Ω^* .

<u>Example 1</u>

As a simple example of the forcing function approach we will consider the method of steepest descent with an Armijo-type step-

size. (Examples of the application of this approach to constrained optimization methods may be found in [6] and in Chung [1], where an interesting application to exact penalty methods is made.) We will assume that ϕ is continuously differentiable on all of $\mathbb{R}^n$, that $G = \mathbb{R}^n$, and that, given the ith iterate x_i , the (i+1)st iterate x_{i+1} is uniquely determined by the relations:

$$x_{i+1} = x_i - \lambda_i \nabla\phi(x_i)^T \text{ , where}$$

$$\lambda_i = \max\{\lambda | \lambda = 2^{-i},\ i = 0,1,2,\ldots, \phi(x_i) - \phi(x_i - \lambda\nabla\phi(x_i)^T) \geq \tfrac{1}{2}\lambda \| \nabla\phi(x_i)\|^2\}$$

It is easily shown that a suitable optimality indicator for this example is obtained by setting

$$\delta(z) = L(z)\| \nabla\phi(z)\|^2 = \phi(x_i) - \phi(x_{i+1}) ,$$

where L has the property that if $y_i \to z$, then $\underline{\lim}\, L(y_i) > 0$. Thus, δ is non-negative and null-continuous on $\mathbb{R}^n$, since $y_i \to z$ and $\delta(y_i) \to 0$ imply $\nabla\phi(y_i) \to 0$, so that $\nabla\phi(z) = 0$. Note that $\Omega^L = \Omega^* = \{x | \nabla\phi(x) = 0\}$. Lemma 2.2 thus guarantees that if the sequence $\{x_i\}$ has an accumulation point $\bar{x}$, then $\nabla\phi(\bar{x}) = 0$. (Boundedness of $\{x_i\}$ may be guaranteed by appropriate level set compactness hypotheses on ϕ .) □

Of course, if $\{x_i\}$ has no accumulation point, then there is no guarantee that Ω^* is non-empty or that $\delta(x_i) \to 0$. Examples are easily constructed with these properties. Furthermore, even if $\{x_i\}$ is bounded, there is no guarantee that the $\{x_i\}$ will converge to a unique accumulation point, and, in fact, "oscillatory" behavior of the $\{x_i\}$ between a finite number of accumulation points may occur (see [8]). (Indeed, under certain weak hypotheses, the existence of a sequence satisfying the hypotheses of Lemma 2.2 yet displaying oscillatory behavior can be guaranteed, as will be shown in Theorem 4.3). From a practical point of view, however, oscillatory behavior is quite rare, suggesting that some additional properties are generally satisfied which prevent such bad behavior. Thus, we are led to consider additional hypotheses under which convergence of the sequence $\{x_i\}$ may be demonstrated. The most obvious (but least applicable) such hypothesis is given in Corollary 2.3 below, and a more useful approach is presented in Theorem 2.4 and Corollary 2.5. (In all of the remaining results of this section, the compactness of closed, bounded subsets of $\mathbb{R}^n$ is exploited, so a direct extension to more general spaces is not possible.)

<u>Corollary 2.3</u> Let the hypotheses of Lemma 2.2 hold, and assume in addition that ϕ is continuous, that $\{x_i\}$ is bounded, and that, for each $\theta \in R$, the set $\Omega^\# \cap \{x | x \in G, \phi(x) = \theta\}$ contains at most one element; then there exists an $x^\# \in \Omega^\#$ such that $x_i \to x^\#$.

Proof: If the result were false, then there would be two subsequences with index sets I and J such that $x_i \overset{I}{\to} x'$ and $x_i \overset{J}{\to} x''$ with $x' \neq x''$. By Lemma 2.2, x' and x'' are in $\Omega^\#$, and by the monotonicity of $\{\phi(x_i)\}$, $\phi(x') = \phi(x'')$, a contradiction. □

The disadvantage to the above approach to proving convergence is that it requires what amounts to a global uniqueness hypothesis. This type of hypothesis is usually not verifiable except under strict convexity hypotheses. On the other hand, since most algorithms perform only a local search at each iteration, their convergence properties are generally determined by the local behavior of the function to be minimized. To make these notions precise, we will introduce in Theorem 2.4 below a "stability" hypothesis (2.5) that, in effect, "damps" the step-length $\|x_i - x_{i+1}\|$ when x_i is near $\Omega^\#$.

<u>Theorem 2.4</u>: Let the hypotheses of Lemma 2.1 hold, let $\{x_i\}$ be bounded, and, <u>in addition</u>, assume that there exist functions ρ and μ such that $\mu : \mathbb{R}^n \to \mathbb{R}^1_+$, μ is a C_0 function defined on $\mathbb{R}^n$ with $\mu(x) > 0$ for $x \neq 0$, ρ is a function from G into $\mathbb{R}^1_+$ such that $\delta(x_i) \to 0$ implies $\rho(x_i) \to 0$, and for $i = 0,1,\ldots$

$$(2.5) \qquad \rho(x_i) \geq \mu(x_i - x_{i+1}) .$$

Then $\|x_i - x_{i+1}\| \to 0$, and the set of accumulation points of $\{x_i\}$ consists of a single point <u>or</u> a continuum.

<u>Proof</u>: Suppose that $\|x_i - x_{i+1}\| \not\to 0$. Then there exists an index set J such that $x_i \overset{J}{\to} x^\#$, $x_{i+1} \overset{J}{\to} x^{\#\#}$ with $x^\# \neq x^{\#\#}$. By Lemma 2.1 $\delta(x_i) \to 0$, so (2.5) implies $\mu(x_i - x_{i+1}) \to 0$, and thus $\mu(x^\# - x^{\#\#}) = 0$. However, $\|x^\# - x^{\#\#}\| > 0$ implies $\mu(x^\# - x^{\#\#}) > 0$, a contradiction.

Thus $\|x_i - x_{i+1}\| \to 0$, and because of the boundedness of $\{x_i\}$, the remaining conclusion is a well-known result of Ostrowski [12]. □

Example 2:

If the sequence $\{x_i\}$ was derived according to the procedure described in Example 1, then by taking $\mu(z) = \|z\|$ and $\rho(z) = \|\nabla\phi(z)\|$, it is easily seen that the iterates of Example 1 satisfy (2.5) and that ρ and μ satisfy the hypotheses of Theorem 2.4. □

The following Corollary is an immediate consequence of Lemma 2.2 and Theorem 2.4 and establishes sufficient conditions for convergence of the entire sequence $\{x_i\}$ to a point in Ω^* .

Corollary 2.5: Let $\{x_i\}$ be a bounded sequence satisfying:

(a) $\phi(x_i) - \phi(x_{i+1}) \geq \delta(x_i)$ $(i = 0,1,2,\ldots)$, where ϕ is l.s.c. on G and δ is C_0 on G, and

(b) $\rho(x_i) \geq \mu(x_i - x_{i+1})$ $(i = 0,1,2,\ldots)$, where $\rho(x) = 0$ for $x \in \Omega^*$ and ρ is continuous at each $x \in \Omega^*$, and μ is C_0 on R^n and satisfies $\mu(x) > 0$ for $x \neq 0$.

If for each $x^* \in \Omega^*$, there exists an open set $N(x^*)$ containing x^* such that $N(x^*) \cap \Omega^* = \{x^*\}$, then $\{x_i\}$ converges to a point of Ω^* .

Proof: By Theorem 2.4, either $\{x_i\}$ converges to a point in Ω^* or its accumulation points form a continuum contained in Ω^* . However, by hypothesis, Ω^* consists of isolated points and hence cannot contain

a continuum. □

Example 3:

Again let $\{x_i\}$ be as in Example 1, and suppose that $\phi \in C^2$ and that $\nabla^2\phi(x)$ is non-singular if $x \in \Omega^*$. Then for each $x \in \Omega^*$, there exists an open set $N(x)$ such that $N(x) \cap \Omega^* = \{x\}$. (For, otherwise, there would be an $x^* \in \Omega^*$ and a sequence $\{y_i\}$ with $y_i \to x^*$ and $\nabla\phi(y_i) = 0$. Without loss of generality, we may assume the sequence $\{(y_i - x^*)/\|y_i - x^*\|\}$ converges to d , where $\|d\| = 1$, and it is then easily shown that $\nabla^2\phi(x^*)d = 0$, contradicting the non-singularity of $\nabla^2\phi$ on Ω^* .) □

Note that Corollary 2.5 is a global convergence theorem, i.e., it guarantees convergence to a point in Ω^* from an arbitrary starting point x_0 of G provided that the monotonicity and localization hypotheses are satisfied by the iterates and by Ω^* . Point-of-attraction theorems establishing local convergence under somewhat weaker hypotheses as well as convergence theorems that make use of the properties of accumulation points of $\{x_i\}$ may be found in [8], but it should be recognized that for global convergence, global hypotheses are required. The main results of this section are summarized in Table 1, which also indicates the results to be obtained in sections 3 and 4.

Table 1

Properties of the Point-to-Set Mapping S	Properties of the Functions δ, ρ, μ	Convergence Results
Monotonicity	Non-negativity of δ	Accumulation points are in Ω^L
$x^* \in S(x^*)$ (x^* a GFP)	$\delta(x^*) = 0$	x^* can be an accumulation point
Monotonicity plus sequential monotonicity at non-GFP's	Non-negativity and null-continuity of δ	Each accumulation point $\hat{x}$ is a GFP and $\delta(\hat{x})=0$ ($\hat{x} \in \Omega^*$)
$\{x^*\} = S(x^*)$ (x^* an SFP)	$\delta(x^*) = 0$; $\rho(x^*) = 0$	If $x_0 = x^*$, then $x_i = x^*$ for all i
Monotonicity plus sequential monotonicity at non-SFP's plus u.s.c. at SFP's	Non-negativity and null-continuity of δ ; ρ is continuous and equals 0 at points x such that $\delta(x)=0$; μ is null-continuous and positive-definite	If $\{x_i\}$ is bounded and the SFP's do not form a continuum, then $\{x_i\}$ <u>converges</u> to an $\hat{x}$ such that $\delta(\hat{x}) = \rho(\hat{x}) = 0$ ($\hat{x}$ is an SFP)

3. Relationships between Forcing Functions and Point-to-Set Mappings

Given a pair of functions ϕ , δ defined on G , the <u>algorithm corresponding to ϕ and δ</u> is defined to be the algorithm given by:

(3.1) choose an arbitrary $x_0 \in G$, and

(3.2) given x_i , choose x_{i+1} such that

$$\phi(x_i) - \phi(x_{i+1}) \geq \phi(x_i) \qquad (i = 0,1,2,\ldots)$$

Clearly, this algorithm will be well-defined if and only if the set defined by

(3.3) $$S(x) \equiv \{y \mid y \in G,\ \phi(x) - \phi(y) \geq \delta(x)\}$$

is <u>non-empty</u> for <u>all</u> $x \in G$. Having so defined the point-to-set mapping S , this algorithm could also be thought of as <u>the algorithm corresponding to S</u> since (3.2) could be replaced by the statement

(3.4) given x_i, choose $x_{i+1} \in S(x_i)$ $\qquad (i = 0,1,2,\ldots)$.

Thus, we could attempt to analyze this algorithm <u>either</u> by considering the properties of ϕ and δ and applying the results of the previous section, <u>or</u> by considering properties of S , and applying point-to-set mapping convergence theorems. In this section we will discuss what properties of S are, in some sense, equivalent to certain properties of ϕ and δ , and develop point-to-set mapping convergence theorems analogous to the convergence theorems of section 2. (The results to be established in this section are summarized in Table 1.)

As shown in the previous section, some convergence properties can be proved if we merely assume that ϕ is l.s.c. on G and that δ is a non-negative C_0 function on G.

These properties of ϕ and δ, however, do <u>not</u> imply semi-continuity properties for S. S, for example, may fail to be upper semi-continuous (or "closed" as this property is sometimes described) as a result of discontinuities in ϕ and/or δ. (A point-to-set T from G into the subsets of G will be said to <u>u.s.c. at a point x</u> if $\{x_i\} \subseteq G$, $x_i \to x$, $y_i \in T(x_i)$, and $y_i \to y$ imply $y \in T(x)$, <u>and u.s.c. on a subset of G</u> if it is u.s.c. at every point in that subset.) In order to see what properties may be claimed for S in this case, some additional notation will be introduced.

Let T be a point-to-set mapping from G into the subsets of G. A point $x^{\#}$ is defined to be a <u>generalized fixed-point</u> (GFP) of T if $x^{\#} \in T(x^{\#})$, and a <u>strong fixed-point</u> (SFP) of T if $T(x^{\#}) = \{x^{\#}\}$. (Clearly every SFP is also a GFP, but the converse will not hold if $\{x^{\#}\}$ is a proper subset of $T(x^{\#})$. As will be seen below, there is a correspondence between the set of GFP's and the set of points on which a related optimality indicator vanishes, and a correspondence between the set of SFP's and the set of points on which both an optimality indicator and a certain distance majorant vanish.)

T is said to be <u>monotonic</u> on G w.r.t. a function $\omega : G \to R^1$ if $y \in T(x)$ implies $\omega(y) \leq \omega(x)$. T will be said to be <u>sequentially monotonic</u> w.r.t. ω on a set $M \subseteq G$ if $x \in M$, $x_i \to x$, $y_i \in T(x_i)$, $\omega(x_i) \to \omega^{\#}$, and $\omega(y_i) \to \bar{\omega}$ imply $\bar{\omega} < \omega^{\#}$. (Note that if $x^{\#}$ is a GFP or SFP of T, then $x^{\#}$ <u>cannot</u> be in the set M on which T is sequentially monotonic, because we may take $x_i \equiv y_i \equiv x^{\#}$, and the strict inequality in the definition of sequential monotonicity is not satisfied. In the convergence results below, the mappings considered will be sequentially monotonic at all points other than GFP's or SFP's. Note also that if $x \in M$ and $y \in T(x)$, then $\omega(y) < \omega(x)$, but that sequentially monotonicity is a <u>stronger</u> property than the simple requirement that $\omega(y) < \omega(x)$ whenever $x \in M$ and $y \in T(x)$.)

Our first result using these definitions indicates the properties induced by requiring δ to be a non-negative C_0 function on G.

<u>Theorem 3.1</u>: If δ is a non-negative C_0 function on G, then (a) S is <u>monotonic</u> on G w.r.t. ϕ, and (b) S is <u>sequentially monotonic</u> on $G/\Omega^{\#}$ w.r.t. ϕ, and (c) $\Omega^{\#}$ is the <u>set of GFP's</u> of S.

<u>Proof</u>: Properties (a) and (c) follow directly from the definitions, so we will exhibit the proof for (b) only. Let $x \in \Omega^{+}$, $x_i \to x$, $y_i \in S(x_i)$, $\phi(x_i) \to \phi^{\#}$, $\phi(y_i) \to \bar{\phi}$. By the non-negativity of δ, $\bar{\phi} \leq \phi^{\#}$. If $\bar{\phi} = \phi^{\#}$, then $\delta(x_i) \to 0$ and thus $x \in \Omega^{\#}$, a contradiction, so $\bar{\phi} < \phi^{\#}$. □

Let the functions ρ and μ be defined on G, and $\mathbb{R}^n$ respectively, and define the point-to-set mapping

$$(3.5)\qquad \hat{S}(x) \equiv \{y \mid y \in S(x),\ \rho(x) \geq \mu(x-y)\}$$

Note that S is a restriction of $\hat{S}$, by which we mean that $\hat{S}(x) \subseteq S(x)$ for all $x \in G$. As a restriction of S, it is easily seen that conclusions (a) and (b) of Theorem 3.1 must continue to hold when S is replaced by $\hat{S}$. We will now show that conclusion (c) may be strengthened in a useful manner if appropriate properties are assumed for ρ and μ. (These properties are essentially those used in Theorem 2.4.)

Theorem 3.2: Let δ be a non-negative C_0 function on G, let ρ be a non-negative function on G such that ρ is continuous on Ω^* and $\rho(x^*) = 0$ if $x^* \in \Omega^*$, and let μ be a non-negative C_0 function on R^n with $\mu(z) \geq 0$ if $z \neq 0$. Then (a) $\hat{S}$ is sequentially monotonic on G/Ω^* w.r.t. ϕ, and (b) S is u.s.c. on Ω^*, which is the set of SFP's of $\hat{S}$.

Proof: As noted previously, conclusion (a) follows from the observation that $\hat{S}$ is a restriction of S; we thus need only prove (b). If $x^* \in \Omega^*$, then $\rho(x^*) = 0$, so the inequality $\rho(x^*) \geq \mu(x^*-y)$ and the positive-definite property of μ force $y = x^*$ when $y \in \hat{S}(x^*)$, so x^* must be a SFP of $\hat{S}$. Conversely, if x^* is a SFP of $\hat{S}$, then $x^* \in \hat{S}(x^*)$ implies $\delta(x^*) = 0$. Now suppose also that $z_i \to x^*$, $y_i \in \hat{S}(z_i)$, $y_i \to y^*$. Since $\rho(x^*) = 0$ and ρ is continuous on Ω^*, $\rho(z_i) \to 0$, and thus $\mu(z_i-y_i) \to 0$. Since μ is a C_0 function, $\mu(x^*-y^*) = 0$, and thus $y^* = x^* \in S(x^*)$. □

Having derived properties of S and $\hat{S}$ that are induced by properties of the functions appearing in their definitions, we will now take the opposite point of view, and, given a point-to-set mapping T with certain properties, we will show that related functions δ, ρ, and μ with the properties introduced in section 2 may be constructed.

Let T be a point-to-set mapping from G into its subsets. If T is monotonic on G w.r.t. a function ϕ, we define the non-negative optimality indicator corresponding to T and ϕ to be

$$(3.6)\qquad \delta^*(x) = \inf_{y \in T(x)} (\phi(x)-\phi(y)) .$$

Also, we define the distance majorant associated with T to be the extended real-valued function on G defined by

$$(3.7)\qquad \rho^*(x) = \sup_{y \in T(x)} \|y-x\| .$$

The set of GFP's of T is denoted by Γ^* and the set of SFP's of T is denoted by Γ^{**}.

Our first result employing these definitions gives sufficient conditions for the optimality indicator to be C_0 on G.

Theorem 3.3: Let T be monotonic on G w.r.t. ϕ, and let T be sequentially monotonic on G/Γ^* w.r.t. ϕ. If ϕ is continuous or if ϕ is l.s.c. and bounded from above on G, then δ^* is a non-negative C_0 function on G, and $\{x \mid \delta^*(x) = 0\} = \Gamma^*$.

<u>Proof</u>: Clearly if $x^{\#} \leq \Gamma^{\#}$, then $\delta^{\#}(x^{\#}) = 0$. On the other hand, if $\delta^{\#}(\bar{x}) = 0$, then by choosing $x_i = \bar{x}$ for all i, letting $\{y_i\}$ be such that $\{y_i\} \subseteq T(\bar{x})$ and $\phi(y_i) \to \phi(\bar{x})$, and exploiting the sequential monotonicity property of T on $G/\Gamma^{\#}$, we conclude that $\bar{x} \notin G/\Gamma^{\#}$, so $\bar{x} \in \Gamma^{\#}$.

If $\{z_i\} \subseteq G$ and $z_i \to \hat{x}$ then because of the hypotheses on ϕ, there exists a subsequence of $\{z_i\}$ such that $\phi(z_i)$ is convergent. So if $x_i \to x$ and $\delta^{\#}(x_i) \to 0$, without loss of generality we may assume that there exists a $\phi^{\#}$ such that $\phi(x_i) \to \phi^{\#}$ and a sequence $\{y_i\}$ such that $y_i \in T(x_i)$ for each i and $\phi(y_i) \to \phi^{\#}$. Thus $x \notin G/\Gamma^{\#}$ and $\delta^{\#}$ is a C_0 function on G. □

The following theorem shows that if T is sequentially monotonic on $G/\Gamma^{\#\#}$, u.s.c. on $\Gamma^{\#\#}$, and has a weak boundedness property "near" $\Gamma^{\#\#}$, then $\rho^{\#}$ is continuous at each point of $\Gamma^{\#\#}$.

<u>Theorem 3.4</u>: Let T be sequentially monotonic on $G/\Gamma^{\#\#}$ w.r.t. ϕ, and let T be u.s.c. on $\Gamma^{\#\#}$. If for each $x^{\#\#} \in \Gamma^{\#\#}$ there exist positive constants K and K' such that $\|x-x^{\#\#}\| \leq K$ implies that, for some $y \in T(x)$, $\|y-x^{\#\#}\| \leq K'$, then the distance majorant $\rho^{\#}$ defined by (3.7) vanishes and is continuous at each point of $\Gamma^{\#\#}$.

<u>Proof</u>: If $x^{\#\#} \in \Gamma^{\#\#}$, then by definition $\rho^{\#}(x^{\#\#}) = 0$. Let $\{x_i\} \subseteq G$ with $x_i \to x^{\#\#}$. For i sufficiently large, $\rho^{\#}(x_i) \leq K'$, so choose

an index set I and a sequence $\{y_i\}$ such that $y_i \in T(x_i)$ for all i, $y_i \xrightarrow{I} y$, and $\|y_i-x_i\| \xrightarrow{I} \overline{\lim_{i\to\infty}}\, \rho^{\#}(x_i)$. But by the u.s.c. of T on $\Gamma^{\#\#}$, $y \in T(x^{\#\#})$ so $y = x^{\#\#}$ and $\overline{\lim_{i\to\infty}}\, \rho^{\#}(x_i) = 0$. Since $\overline{\lim}\, \rho^{\#}(x_i) \geq \underline{\lim}\, \rho^{\#}(x_i) \geq 0$, we have $\overline{\lim}\, \rho^{\#}(x_i) = \underline{\lim}\, \rho^{\#}(x_i) = 0$ and thus $\rho^{\#}$ is continuous at $x^{\#\#}$. □

The following Corollary summarizes the results of the previous two theorems:

<u>Corollary 3.5</u>: Let T satisfy the hypotheses of Theorems 3.3 and 3.4. Then there exist functions δ, ρ, and μ such that $y \in T(x)$ implies $\phi(x) - \phi(y) \geq \delta(x)$ and $\rho(x) \geq \mu(x-y)$, where δ and ρ are non-negative on G, δ is C_0 on G, ρ vanishes and is continuous at each point of $\Gamma^{\#\#}$, and μ is C_0 and positive-definite on R^n.

<u>Proof</u>: Let $\delta \equiv \delta^{\#}$, $\rho \equiv \rho^{\#}$, and $\mu(x-y) \equiv \|x-y\|$, and apply Theorems 3.3 and 3.4. □

4. Convergence Theorems for Point-to-Set Mappings

By using the convergence results of section 2 and Theorems 3.3 and 3.4, it is possible to develop convergence theorems for algorithms based on point-to-set mappings. However, rather than constructing proofs for such theorems by applying previous theorems, it turns out to be somewhat simpler and more illuminating to give direct proofs.

The first result of this type is the analog of Lemma 2.2 suggested by Theorem 3.3.

<u>Theorem 4.1</u>: Let T be a point-to-set mapping from G into the <u>non-empty</u> subsets of G, and let T be <u>monotonic</u> on G w.r.t. ϕ on G/Γ^* continuous function ϕ, and <u>sequentially monotonic</u> w.r.t. ϕ on G/Γ^* (where Γ^* is the set of GFP's of T). If a sequence generated by the algorithm corresponding to T has an accumulation point x^*, then $x^* \in \Gamma^*$.

<u>Proof</u>: We will suppose that $x^* \notin \Gamma^*$, and show a contradiction. Since ϕ is assumed l.s.c., $\phi^* \equiv \lim \phi(x_i) \geq \phi(x^*)$. Let I be such that $x_i \xrightarrow{I} x^*$. Then $\phi(x_i) \xrightarrow{I} \phi^*$ and $\phi(x_{i+1}) \xrightarrow{I} \phi^*$, contradicting the sequential monotonicity property at x^*. □

A similar analog of Theorem 2.4 may be stated, but for the sake of brevity we will state and prove the analog of Corollary 2.5 suggested by Theorem 3.5.

<u>Theorem 4.2</u>: Let T satisfy the hypotheses of Theorem 4.1, and let $\{x_i\}$ be generated by the algorithm corresponding to T. If (a) $\{x_i\}$ is bounded, (b) the set of SFP's of T coincides with Γ^* and does <u>not</u> contain a continuum, and (c) T is u.s.c. at each SFP, then $x_i \to x^*$, where x^* is an SFP of T.

<u>Proof</u>: Let x^* be an accumulation point of $\{x_i\}$. By the previous theorem $x^* \in \Gamma^*$ so x^* is an SFP of T by hypothesis (b). We will assume $\|x_{i+1} - x_i\| \not\to 0$ and show a contradiction. If $\|x_{i+1} - x_i\| \not\to 0$, there exists an I and a $\delta > 0$ such that $\|x_{i+1} - x_i\| \geq \delta$ for $i \in I$ and $x_i \xrightarrow{I} x'$, $x_{i+1} \xrightarrow{I} x''$. Since x' must be an SFP, $x'' = x'$ and thus $\|x_{i+1} - x_i\| \xrightarrow{I} 0$, a contradiction. Since $\{x_i\}$ is bounded and $\|x_{i+1} - x_i\| \to 0$, if $\{x_i\}$ did not converge, its accumulation points would form a continuum contained in Γ^*, which is impossible. □

The crucial role of the SFP's in the preceding theorem is illustrated by the following result, which shows that, if Γ^* is finite but contains no SFP's, it is always possible to generate a <u>divergent</u> sequence by using the algorithm corresponding to T.

<u>Theorem 4.3</u>: Let the hypotheses of Theorem 4.1 hold. If Γ^* is a finite set containing <u>no</u> SFP's, then there exists an x_o and a corresponding sequence $\{x_i\}$ generated by the algorithm corresponding to T such that $\{x_i\}$ does <u>not</u> converge.

Proof: Let $\Phi \equiv \{x' | x' \in \Gamma^{*}, \phi(x') \leq \phi(x) \text{ for all } x \in \Gamma^{*}\}$ and let $x_0 \in \Phi$. Given x_i, choose $x_{i+1} \neq x_i$. (This is always possible, for otherwise x_i would be an SFP.) We will suppose that $x_i \to x^{*}$, and show a contradiction. By the previous theorem, $x^{*} \in \Gamma^{*}$, and since $\phi(x^{*}) \leq \phi(x_0)$, we have $x^{*} \in \Phi$ and thus $\phi(x^{*}) = \phi(x_i)$ for all i. Thus $\delta(x_i) = 0$ for all i and $x_i \in \Phi$ for all i. Since Φ is a finite set, the relations $\{x_i\} \subseteq \Phi$ and $x_i \to x^{*}$ imply $x_i = x^{*}$ for all i sufficiently large, contradicting the fact that $x_{i+1} \neq x_i$. □

5. Methods Involving Anti-Jamming Parameters

We now wish to extend the convergence analysis approach developed in previous sections to algorithms for which the optimality indicator depends not only on x but also on a scalar parameter ε. This extension allows the analysis of constrained optimization methods employing a so-called "anti-jamming" parameter. The conditions to be given below were previously stated in [10], but here they will be presented as a natural extension of the results in sections 2-4.

The simplest approach to an appropriate extension of the previously developed theory is to replace the relation $\phi(x_i) - \phi(x_{i+1}) \geq \delta(x_i)$ by

$$(5.1) \qquad \phi(x_i) - \phi(x_{i+1}) \geq \hat{\delta}(\varepsilon_i, x_i)$$

in order to reflect the dependence of the change in ϕ on the i^{th} value of the parameter, ε_i. Note that by defining the composite variable $z = (\varepsilon, x)$ and the function $\hat{\phi}(z) \equiv \phi(x)$, the relation (5.1) can be written so that the same variable appears on both sides, i.e., in the form $\hat{\phi}(z_i) - \hat{\phi}(z_{i+1}) \geq \hat{\delta}(z_i)$. If $\hat{\delta}$ is non-negative on $R^1 \times G$ and ϕ is l.s.c. on G (so that $\hat{\phi}$ will also be l.s.c. on $R^1 \times G$), then Lemma 2.1 may be applied to establish properties of the accumulation points of the sequence $\{z_i\}$. (More generally, any set of relations of the form $f(u_i) - f(u_{i+1}) \geq g(v_i)$, where the u_i are in some space U and the v_i are in some space V may be converted in an obvious fashion to a new set of relations in which variables from $U \times V$ appear on both sides and the functions involved have the same continuity and non-negativity properties as f and $\mathcal{G}$, so that the results of section 2 may be applied.) As in section 2, we would then like to go a bit further and exploit additional properties of $\hat{\delta}$ in order to sharpen the characterization of the accumulation points obtained from Lemma 2.1. Unfortunately, while an analog of Lemma 2.2 may be established if $\hat{\delta}$ is C_0 on $R^1 \times G$, for many well-known feasible direction methods the corresponding function $\hat{\delta}$ is <u>not</u> null-continuous on $R^1_+ \times G$ (for a simple example of this phenomenon see p. 24 of [10].)

Thus the properties of $\hat{\delta}$ that we will exploit are weaker than null-continuity, but will nonetheless be strong enough to guarantee that the accumulation points satisfy an optimality condition. These properties will also, of course, be such that they are satisfied by the well-known feasible direction methods. The appropriate additional properties of $\hat{\delta}$ are as follows:

(5.2) $\quad \hat{\delta}(\varepsilon_i, x_i) \geq \delta_2(\omega_i, x_i) \cdot \min\{\omega_i, \|x_i - x_{i+1}\|\}$, where

(5.3) $\quad \omega_i \equiv \delta_3(\varepsilon_i, x_i)$, and

(5.4) $\quad \|x_i - x_{i+1}\| \geq \delta_1(\min\{\varepsilon_i, \omega_i\}, x_i)$,

where δ_1 and δ_2 are non-negative and have the <u>generalized forcing function property</u> (for $j = 1,2$) on $R^1_+ \times G$:

(5.5) $\quad \delta_j(\eta_i, y_i) \to 0$ and $y_i \to \bar{y}$ imply $\eta_i \to 0$

and δ_3 is non-negative and null-continuous on $R^1_+ \times G$. By assuming (5.1)-(5.5) and making an assumption on the relationship of $\{\varepsilon_i\}$ to $\{\omega_i\}$, the following Theorem shows that δ_3, which plays the role of an optimality indicator, vanishes at the accumulation points of $\{(\varepsilon_i, x_i)\}$.

<u>Theorem 5.1</u>: Let (5.1)-(5.5) hold, and let δ_3 be null-continuous on $R^1_+ \times G$. Let $\{\varepsilon_i\}$ be such that the existence of a <u>subsequence</u> of $\{\varepsilon_i\}$ converging to 0 implies (1) that $\varepsilon_i \to 0$ and (2) that $\{\omega_i\}$ also contains a subsequence converging to 0. If $\{x_i\}$ contains an accumulation point $\bar{x}$, then $\delta_3(0,\bar{x}) = 0$.

<u>Proof</u>: This result is proved in Theorem 1, p. 7, of [10]. □

It is shown in [10] that this theorem may be applied to the analysis of the feasible direction methods of Zoutendijk [18], Topkis and Veinott [15], and Mangasarian [6], and that it also suggests new and, in some cases, more efficient parameter generation schemes. While it is thus possible to extend the forcing function approach to algorithms with an anti-jamming parameter, it does not appear possible to similarly extend the point-to-set mapping approach in a natural way to cover this situation, since the convergence proof depends on the special structure of $\hat{\delta}(\varepsilon_i, x_i)$. It might be noted that Zangwill [17] also establishes the convergence of the feasible direction methods that he considers by a direct argument rather than by the application of his general point-to-set mapping convergence theorems.

The interested reader may also refer to [10] for extensions of Theorem

5.1 that give sufficient conditions for the convergence of the full sequence $\{x_i\}$ to a point x^* such that $\delta_3(0,x^*) = 0$. These conditions are analogous to those assumed in Corollary 2.5.

6. Conclusions

A general convergence theory for monotonic mathematical programming algorithms has been developed via the forcing function approach. This approach has the pedagogical advantage of avoiding the use of point-to-set mappings, but is nevertheless shown to be equivalent to a development relying on point-to-set mapping properties for two of the three classes of algorithms considered. The forcing function appears to have some advantages in terms of providing a framework for the analysis of the third class of algorithms, feasible direction methods. On the other hand, there are classes of algorithms involving contraction mappings [8], cyclic or "restart" policies ([9]), and linearization procedures [14] for which a point-to-set mapping approach appears quite suitable whereas the forcing function approach would be somewhat unnatural. The point-to-set mapping approach also offers geometric insights not as easily obtained from the forcing function approach.

References

1. S. M. Chung, Globally and Superlinearly Convergent Algorithms for Nonlinear Programming. Ph.D. Thesis, University of Wisconsin Computer Sciences Dept., 1975.

2. J. W. Daniel, The Approximate Minimization of Functionals, Prentice-Hall, Englewood Cliffs, N.J., 1971.

3. B. C. Eaves and W. I. Zangwill, Generalized Cutting Plane Algorithms, SIAM J. Control, 9 (1971), pp. 529-542.

4. W. W. Hogan, Point-to-Set Maps in Mathematical Programming, SIAM Review, 15(1973), pp. 591-603.

5. P. Huard, Optimization Algorithms and Point-to-set Maps, Mathematical Programming, 8 (1975), pp. 308-331.

6. O. L. Mangasarian, Dual Feasible Direction Methods, in Techniques of Optimization, ed. by A. V. Balakrishnan, Academic Press, New York, 1972.

7. R. R. Meyer, The Validity of a Family of Optimization Methods, SIAM J. Control, 8 (1970), pp. 41-54.

8. R. R. Meyer, Sufficient Conditions for the Convergence of Monotonic Mathematical Programming Algorithms, forthcoming in Journal of Computer and System Science.

9. R. R. Meyer, On the Convergence of Algorithms with Restart, University of Wisconsin Computer Sciences Technical Report # 225, Madison, Wisconsin, 1974.

10. R. R. Meyer, A Convergence Theory for a Class of Anti-Jamming Strategies, University of Wisconsin Mathematics Research Center Report #1481, Madison, Wisconsin, 1975.

11. J. M. Ortega and W. C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.

12. A. M. Ostrowski, Solution of Equations and Systems of Equations, Academic Press, New York, 1966.

13. E. Polak, Computational Methods in Optimization, Academic Press, New York, 1971.

14. S. M. Robinson and R. R. Meyer, Lower Semicontinuity of Multivalued Linearization Mappings, SIAM J. Control, 11 (1973), pp. 525-533.

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

16. Philip Wolfe, On the Convergence of Gradient Methods Under Constraint, IBM Research Paper RC-1752, Yorktown Heights, N.Y., 1967.

17. W. I. Zangwill, Nonlinear Programming, Prentice Hall, Englewood Cliffs, N.J., 1969.

18. G. Zoutendijk, Methods of Feasible Directions, Elsevier, Amsterdam, 1960.

Appendix

In order to obtain globally convergent mathematical programming algorithms, it is customary in practice to introduce step-size procedures that guarantee a "sufficient decrease" in some function. In terms of the theory described above, "sufficient decrease" means that the function δ determining a lower bound for the decrease should be null-continuous and that the set of points on which δ vanishes should coincide with the set of points satisfying an appropriate optimality condition. For purposes of comparison with the results of Zangwill and Polak, however, we must allow for the possibility of a "worsening" (or increase) in the value of ϕ, or an empty set of successors, even though for nonlinear minimization algorithms used in practice it is always possible to let $x_{i+1} = x_i$ if computations at the i^{th} iteration have not yielded a point with a smaller objective value.

A Comparison with Polak's Theorem 1.3.10

Polak's basic algorithm and his corresponding Theorem 1.3.10 read as follows:

Algorithm (Polak): Let A be a point-to-set mapping from G into the non-empty subsets of G.

Step 0 Compute a $z_0 \in G$

Step 1 Set $i = 0$

Step 2 Compute a point $y \in A(z_i)$

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

Step 4 If $\phi(z_{i+1}) \geq \phi(z_i)$ stop; else, set $i = i+1$ and go to step 2.

Theorem 1.3.10 (Polak)

Suppose that (i) ϕ is either continuous at all nondesirable points or ϕ is bounded from below on G;

(ii) for every $z \in G$ which is not desirable, there exist an $\varepsilon(z) > 0$ and a $\bar{\delta}(z) < 0$ such that $\phi(z'') - \phi(z') \leq \bar{\delta}(z) < 0$ for all $z' \in T$ such that $\|z'-z\| \leq \varepsilon(z)$ and for $z'' \in A(z')$

Then, either the sequence $\{z_i\}$ constructed by the algorithm is finite and its next to last element is desirable, or else it is infinite and every accumulation point of $\{z_i\}$ is desirable.

In order to compare our approach with that of Polak, we first define

$$\delta(x) \equiv \max\{0, \inf\{\phi(x) - \phi(x') \mid x' \in A(x)\}\} . \tag{A.1}$$

Note that δ is non-negative on G. We will now show how Lemma 2.1 may be used to obtain a strengthened version of Polak's theorem.

Lemma A.1: Let ϕ be either l.s.c. or bounded from below on G, and let hypothesis (ii) of Polak's theorem hold. If the set of z_i constructed by Polak's algorithm is finite, then its next-to-last element is in Ω^*. If the set of z_i has an accumulation point $\bar{x}$, then $\bar{x} \in \Omega^L$, which is contained in the set of desirable points.

Proof: In the finite termination case, the conclusion is obvious. In the infinite case, since $\phi(z_i) - \phi(z_{i+1}) \geq \delta(z_i)$ for all i, Lemma 2.1 applies (recall that the proof requires only that ϕ is bounded from below). To see that Ω^L is a subset of the desirable points, suppose that $y_i \to x$ and $\delta(y_i) \to 0$, but that $x \in T'$, the set of non-desirable points. Since $\delta(y_i) \to 0$, there exists a sequence $\{y_i'\}$, with $y_i' \in A(y_i)$ such that $\overline{\lim}[\phi(y_i) - \phi(y_i')] \leq 0$, contradicting Polak's hypothesis (ii). □

Note that Lemma A.1 yields a stronger result than Polak's theorem, since Ω^L may be a proper subset of the desirable points, as the following example shows:

Example: Let $G \equiv \{1/n | n = 1,2,\ldots\} \cup \{0\} \cup \{-1\}$, $\phi(x) \equiv x$, $T' \equiv \{1/n | n = 2,3,\ldots\}$, and

$$A(x) \equiv \begin{cases} \{-1\} & \text{if } x = -1, \text{ or } 0 \\ \{1/(n+1)\} & \text{if } x = 1/n,\ n = 1,2,\ldots \end{cases} .$$

Then all of the hypotheses of Polak's theorem are satisfied, and $G/T' = \{-1,0,1\}$, $\Omega^+ = \{-1,0\}$, and $\Omega^* = \{-1\}$. Note that although the point 1 has been classed as a "desirable" point, the algorithm can neither terminate at 1 nor converge to 1. In this case, then, Lemma A.1 is a sharper result than Polak's, since it restricts the terminal and accumulation points to smaller sets.

The difference between the sharpness of the two results is essentially a result of the fact that Ω^+ and Ω^* are completely determined by A and ϕ , whereas Polak's "desirable set" is not uniquely determined. In fact, if the hypotheses of Polak's theorem are satisfied when the desirable set

is taken to be a particular set T^* , then the hypotheses remain satisfied when the desirable set is taken to be any larger set, i.e., any subset of G containing T^* (it could, for example, be taken to be G itself).

Comparison with Zangwill's Theorem A

In Theorem A the algorithm is given a point z_1 and generates the sequence $\{z_k\}_1^\infty$ by use of the recursion $z_{k+1} \in A(z_k)$.

Convergence Theorem A (Zangwill)

Let the point-to-set map $A : G \to G$ determine an algorithm that given a point $z_1 \in G$ generates the sequence $\{z_k\}_1^\infty$. Also let a solution set $\Omega \subset G$ be given.

Suppose

(1) All points z_k are in a compact set $X \subset G$

(2) There is a continuous function $Z : G \to E^1$ such that:

(a) if z is not a solution, then for any $y \in A(z)$

$$Z(y) > Z(z)$$

(b) If z is a solution, then either the algorithm terminates or for any $y \in A(z)$

$$Z(y) \geq Z(z)$$

and

(3) The map A is closed at z if z is not a solution.

Then either the algorithm stops at a solution, or the limit of any convergent subsequence is a solution.

The statement of Zangwill's theorem is a bit unclear, since the suggestion of the possibilities that "the algorithm terminates" or that "the algorithm stops" cannot be reconciled with the hypothesis that the algorithm generates an infinite sequence $\{z_i | i = 1,2,\dots\}$. On the basis of some of Zangwill's other results we will assume that the statement that "the algorithm terminates" at x_i is equivalent to $A(x_i)$ being empty. Hence, we will again define δ to take this into account, and apply Lemma 2.1 to obtain a strengthened result. Let $\phi \equiv -Z$ and let δ be defined as follows on G :

$$\delta(x) \equiv \begin{cases} \inf\{\phi(x) - \phi(y) | y \in A(x)\}, \text{ if } A(x) \text{ is non-empty} \\ 0 \quad \text{if } A(x) \text{ is empty} . \end{cases} \tag{A.4}$$

(Note that δ is non-negative on G .)

<u>Lemma A.2</u>: Let (2) and (3) of Theorem A hold, and let $\{z_i \; i \in I\}$ be a set of points generated by Zangwill's algorithm. If I is finite, then the last element $z_i \in \Omega^* \cap \Omega$. If I is infinite and the sequence $\{z_i\}$ has an accumulation point $\bar{x}$, then $\bar{x} \in \Omega^L$. If I is infinite and the sequence $\{z_i\}$ is bounded, then each accumulation point $x^* \in \Omega^L \cap \Omega$.

Proof: The proof of the first two conclusions is analogous to the proof of Lemma A.1. If the sequence $\{z_i\}$ is bounded, then by Zangwill's theorem the accumulation points belong to Ω , and by the second conclusion of the Lemma they also belong to Ω^L . □

It should be noted that Ω^L may contain points not in Ω and vice-versa, as the following example shows:

<u>Example</u>:

Let $N \equiv \{n | n = 1,2,\dots\}$,

$G \equiv N \cup \{1/n | n = 1,2,\dots\} \cup \{0\} \cup \{-1\}$,

$$A(x) \equiv \begin{cases} -1 \text{ if } x = -1 \text{ or } 0 \\ (1/x) + 1 \text{ if } x = 1/n;\ n = 2,3,\dots \\ 1/(x+1) \text{ if } x = n;\ n = 1,2,3,\dots \end{cases}$$

$$\phi(x) \equiv \begin{cases} x \text{ if } x \le 1 \\ 1/x \text{ if } x > 1 \end{cases}$$

$\Omega \equiv \{-1,1\}$

Note that hypotheses (2) and (3) of Zangwill's theorem are satisfied, and that $\Omega^L = \{-1,0\}$ and $\Omega^* = \{-1\}$. Thus, by Lemma A.2, if the algorithm terminates in a finite number of steps, it must terminate at the point -1; if the algorithm yields an infinite set of iterates, then either of -1 or 0 could be accumulation points; and, if it yields an infinite set of iterates contained in a bounded set, then they must converge to -1 . Note, in fact, that these results are in this case the best possible, since -1 will be the unique accumulation point if the algorithm starts with $z_1 = -1$ or 0 , and 0 will be the unique accumulation point for any other starting point in G . By comparison, Zangwill's Theorem A does not apply to the case in which the z_i are unbounded (which occurs unless $z_1 = -1$ or 0), and in the bounded iterate case, Theorem A narrows the candidates for accumulation points to the set $\{-1,0\}$. Analogs of the comments made regarding the sharpness of Polak's theorem apply here, since Ω is not uniquely determined by A and Z , and may be taken to be much larger than is really necessary, whereas Ω^* and Ω^L are uniquely determined by A and Z . Of course, in applying Theorem A one would like to choose the set Ω as small as possible, i.e., as the intersection of all Ω for which hypotheses (2) and (3) were satisfied. This approach has the disadvantage of providing a rather complex definition of the "solution set" Ω (at least as compared to the definition of Ω^*), as well as a set that does not contain all the possible accumulation points in the more general unbounded iterate case.

Other Forcing Function Approaches

In the approach to convergence analysis used by Ortega and Rheinboldt [11] (a similar approach is also used by Daniel [2]), two main hypotheses are made regarding the decrease in ϕ at each iteration:

$$\phi(x_i) - \phi(x_{i+1}) \geq \sigma_1(|\nabla\phi(x_i)p_i|/\|p_i\|) \ , \quad \text{and}$$

$$|\nabla\phi(x_i)p_i|/\|p_i\| \geq \sigma_2(\|\nabla\phi(x_i)\|) \ ,$$

where p_i is a search direction and the σ_j have the property that, for $j = 1$ or 2 , $\lim_{k\to\infty} \sigma_j(t_k) = 0$ implies $\lim_{k\to\infty} t_k = 0$ for any non-negative

sequence $\{t_k\}$. (By considering the above two inequalities, the step-size and the direction-generation techniques of an algorithm may be analyzed separately, so that the potential independence of those two techniques is emphasized.) If we assume in addition that σ_1 is monotone non-decreasing (a hypothesis that is satisfied by all the algorithms considered in Ortega and Rheinboldt [11], where, in fact, in most cases $\sigma_1(t) = Mt^2$ for some $M > 0$), then we have

$$\phi(x_i) - \phi(x_{i+1}) \geq \sigma_1)(\sigma_2(\|\nabla\phi(x_i)\|)) \ .$$

Letting $\delta(x_i) \equiv \sigma_1(\sigma_2(\|\nabla\phi(x_i)\|))$ and assuming that ϕ is continuously differentiable and $\sigma_1(0) = \sigma_2(0) = 0$, we may conclude that δ is C_0 and that the set of points on which δ vanishes is the set of points on which $\nabla\phi$ vanishes. Thus, under these hypotheses Lemma 2.2 may be applied, and we may conclude that $\nabla\phi$ will vanish at the accumulation points of $\{x_i\}$. Finally Ortega and Rheinboldt also establish, for each algorithm they consider, conditions analogous to (2.5) to guarantee $\|x_i - x_{i+1}\| \to 0$ (these are generally of the form $\bar{M}|\nabla\phi(x_i)p_i|/\|p_i\| \geq \|x_i - x_{i+1}\|$, where $\bar{M} > 0$) , or prove the relation $\|x_i - x_{i+1}\| \to 0$ by utilizing properties of the step-size techniques together with a hypothesis (hemivariateness) on ϕ .