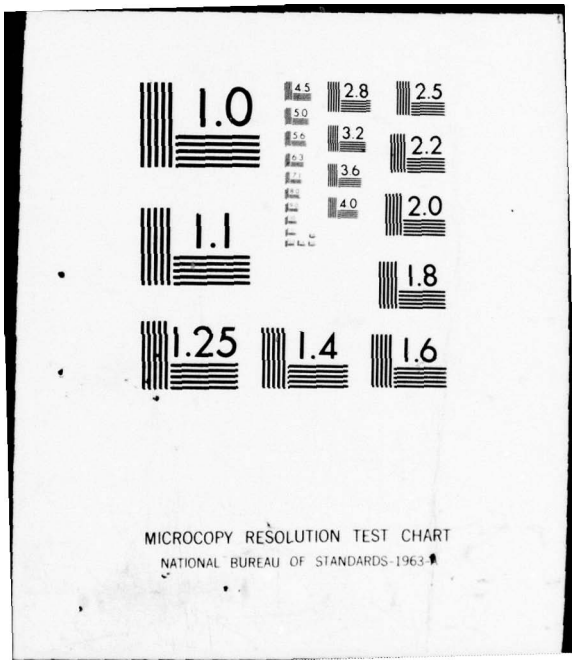
AD-A053 200 STANFORD UNIV CALIF SYSTEMS OPTIMIZATION LAB F/G 12/1
ON THE SOLUTION OF NONLINEAR EQUATIONS BY PATH METHODS.(U)
OCT 77 T R ELKEN N00014-75-C-0267
UNCLASSIFIED SOL-77-25 NL

1 OF 2
AD A053200


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

12

ON THE SOLUTION OF NONLINEAR EQUATIONS BY PATH METHODS

BY

THOMAS R. ELKEN

TECHNICAL REPORT SOL 77-25

OCTOBER 1977

Systems Optimization Laboratory

Department of Operations Research

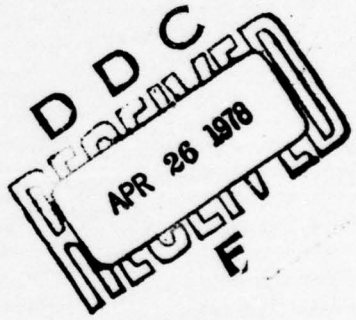
Stanford University

Stanford California 94305

ON THE SOLUTION OF NONLINEAR EQUATIONS BY PATH METHODS

by

Thomas R. Elken



TECHNICAL REPORT SOL 77-25

October 1977

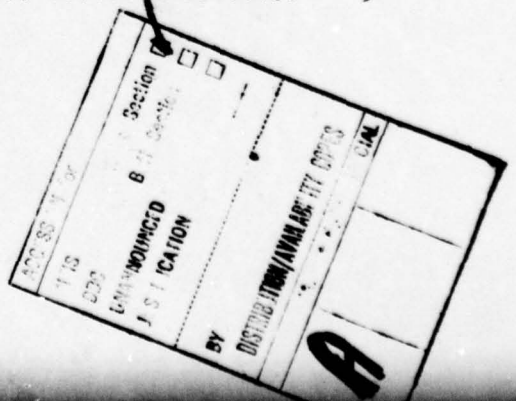
SYSTEMS OPTIMIZATION LABORATORY

DEPARTMENT OF OPERATIONS RESEARCH

Stanford University
Stanford, California

Research and reproduction of this report were partially supported by the U.S. Energy Research and Development Administration Contract EY-76-03-0326 PA #18; The Office of Naval Research Contracts N00014-75-C-0267 and N00014-75-C-0865; the National Science Foundation Grants MCS76-81259 and MCS76-20019.

TABLE OF CONTENTS



NOTATION AND NUMBERING

We define some notation which may not be completely standard. All points and sets are in $\mathbb{R}^n$ unless indicated otherwise.

1) $\mathbb{R}^n_+ \equiv \{x \in \mathbb{R}^n | x \geq 0\}$

2) For $x, y \in \mathbb{R}^n$, denote their <u>inner product</u> by $\langle x,y \rangle = \sum_{i=1}^{n} x_i y_i$.

3) Let $\|x\| = \langle x,x \rangle^{1/2}$ be the <u>norm</u> of x.

4) Denote by $d(\cdot,\cdot)$ the distance function defined as

$$d(A,B) = \inf_{\substack{x \in A \\ y \in B}} \|x-y\|$$

5) Let the <u>ball about</u> A <u>of radius</u> ϵ, where A may be a point or a set, be

$$B(A,\epsilon) = \{x \in \mathbb{R}^n | d(x,A) \leq \epsilon\}$$

6) For any positive integer m, let $\underline{m} \equiv \{1,2,\ldots,m\}$. If $m = 0$, then $\underline{m} \equiv \emptyset$.

7) $\text{conv}[A_1,A_2,\ldots,A_n]$ is the convex hull of the sets (or points) $A_1, A_2, \ldots, A_n$.

8) $\mathbb{R}^{n \times m}$ is the space of all real-valued matrices with n rows and m columns.

9) If $\sigma \subset \underline{n}$ and $\mu \subset \underline{m}$ are index sets, then, for $A \in \mathbb{R}^{n \times m}$,

$A_{\sigma,\cdot}$ consists of the rows of A indexed by σ, and

$A_{\cdot,\mu}$ consists of the columns of A indexed by μ

$A_{\sigma,\mu}$ is the submatrix of elements whose indices are in $\sigma \times \mu$.

10) For two sets A and B,

$$A \setminus B = \{x | x \in A,\ x \notin B\}$$

and

$$A - B = \{z | z = x-y \text{ for some } x \in A, \text{ and } y \in B\}.$$

11) The boundary of a set C is called ∂C, and the interior of set C is called C^0.

Numbering

The chapters are numbered by Roman numerals and the sections are numbered consecutively within each chapter, i.e., II.1, II.2, etc. All theorems, lemmas, and examples are numbered consecutively within each section. Equations are numbered separately and are identified by being enclosed in parentheses. Equation (V.1.2) is the second equation in Section V.1, for instance. The chapter numeral is omitted for results or equations referred to within the same chapter.

ON THE SOLUTION OF NONLINEAR EQUATIONS BY PATH METHODS

ABSTRACT

The problem considered is that of finding a solution to a system of nonlinear equations subject to some auxiliary constraints. The methods studied here are called path methods, also referred to as "continuation" or "global Newton" methods, for solving equations. A general theory is developed which unifies the results from several papers and allows new methods to be analyzed easily. The new methods are shown to converge under more general boundary and monotonicity conditions than those assumed for the existing methods. A rigorous proof of convergence is given for an algorithm which implements a general path method.

ON THE SOLUTION OF NONLINEAR EQUATIONS BY PATH METHODS

CHAPTER I

GENERAL

I.1. Introduction and Summary

The problem we consider is that of finding a solution to a system of nonlinear equations subject to some auxiliary constraints. The objective is to find methods for the solution of such systems which are convergent under rather general conditions and are efficient computationally.

The predominate efforts in studying this problem in recent years have made use of "piecewise linear" or "simplicial methods" for solving systems of equations (cf. Scarf [1967a], Kuhn [1968], Eaves and Saigal [1972], and Merrill [1972]). But a different approach to the equation-solving problem has been revived recently. Kellogg, Li and Yorke [1976] and S. Smale [1976] have proposed "continuation" or "global Newton" methods for solving systems of equations. The history of "continuation" methods goes back to Lahaye [1934], [1938] in the univariate case and Davidenko [1953a,b] in the multivariate case. We hope to elucidate the common features of these methods and to propose some new methods which converge under more general assumptions.

This report contains approximately the first half of the author's dissertation, The Computation of Equilibria by Path Methods.

Many of the theorems and procedures presented here are applied to the development of algorithms for the economic equilibrium problem. When we refer to results or chapters in the report which contains the rest of the dissertation (Elken [1977]), we will refer to them as being in Part 2.

In the remainder of this chapter, we will discuss the equation solving problem further and motivate the development of path methods for solving systems of equations.

In Chapter II we review some basic results from differential topology and introduce a unified theoretical framework for dealing with path methods. The reader familiar with the differential topology may want to skip the first section of this chapter. The theory of subdivided complexes which follows helps to provide new (and easier) proofs for some theorems which demonstrate the existence of paths from a known starting point to a solution point for a system of equations.

In Chapter III results are derived which can be applied to any path method. The primary concern is to develop results which show that we can determine a consistent orientation for a path and that a discrete algorithm can be devised which is guaranteed to follow a path from one boundary of a bounded set to another boundary. This convergence proof is a new result.

Chapter IV presents the existing path methods of Kellogg, Li, and Yorke [1976] and Smale [1976] in a homotopy framework and new proofs that the paths have the desired characteristics. Also new path methods are developed which are motivated by existing path methods or fixed point algorithms. These are shown to be convergent under more general conditions than the existing path methods.

I.2. The Motivation for Path Methods

We are concerned with proving theorems which demonstrate the existence of a path from a known starting point $x^0 \in D \subset \mathbb{R}^n$ to a solution x^* of the problem

$$f(x) = 0 , \qquad x \in D \tag{2.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ and D satisfy certain conditions. The existence of a differentiable path from x^0 to x^* can motivate any number of algorithms suitable for computation of x^*. We will present several. First, however, we discuss other approaches to studying the problem (2.1).

The history of the study of the existence question and algorithms for finding a solution of (2.1) is long and cannot be reported in full here. The most celebrated existence theorem is the Brouwer fixed point theorem (Milnor [1965]): Let $g : D \to D$ be continuous and D a compact convex subset of $\mathbb{R}^n$. Then there is a point $x \in D$ with $g(x) = x$. With a preliminary definition we can state an equivalent theorem in terms of (2.1). We say that $f(x)$ <u>points into</u> D for $x \in \partial D$ if there is an $\alpha > 0$ such that for $\beta \in (0,\alpha)$, $x + \beta f(x) \in D$. The equivalent theorem is: Let $f : D \to \mathbb{R}^n$ be continuous and D be compact and convex. If $f(x)$ points into D for each $x \in \partial D$, then there is an x such that $f(x) = 0$. The equivalence is easily seen when one notes that solving $g(x) = x$ is equivalent to solving $f(x) = 0$ when $f(x) = g(x) - x$ and that solving $f(x) = 0$ is equivalent to finding a fixed point of g when $g(x) = x + \beta f(x)$ for some $\beta > 0$.

These theorems indicate that in some sense fixed point problems and equation solving problems are equivalent. In this thesis the discussion will be concerned with problems of the form (2.1) because the formulas involved in the theorems and algorithms are, in general, easier to state in this framework.

The algorithms for solving (2.1) fall into three general classes:

1) Iterative methods such as Newton's method and its variants.
2) Piecewise linear methods. Those methods generally utilize a triangulation of D or some set containing D.
3) Path methods, also described as "continuation methods," or "global Newton methods."

Iterative methods are primarily of the following form:

$$\text{given } x^0, \text{ let } x^{k+1} = x^k - B^k f(x^k), \quad k = 0,1,... \tag{2.2}$$

where B^k is either $f'(x^k)^{-1}$, an approximation thereof, or a matrix which behaves like $f'(x)^{-1}$ on some subspace. The advantages of these methods are that they generally provide rapid convergence to a solution when x^k is in some neighborhood of a solution x^* to (2.1). On the other hand the only convergence theorems with reasonable conditions on f are local results: if S is a small enough neighborhood of x^* then $x^0 \in S$ implies that (2.2) is a convergent sequence.

The piecewise linear methods are also referred to as "complementary pivot methods," "fixed point methods," or "simplicial methods." The pioneering work in this field was due to Lemke [1965], Scarf [1967a], and Kuhn [1968]. The theory has had a host of contributors including Eaves,

Saigal, Merrill, and Todd. An excellent unification of the theory is contained in Eave's "A Short Course in Solving Equations with P. L. Homotopies" [1976]. This method solves problems of the form (2.1) by considering a piecewise linear approximation $g:D \to \mathbb{R}^n$ of f, then adding a parameter $\theta \in [0,T]$, $0 < T \leq +\infty$ and a fixed vector $d \in \mathbb{R}^n$ so that the problem

$$F(x,\theta) = g(x) + \theta d = 0 \tag{2.3}$$

has a trivial and unique solution for $\theta = T$ (or θ sufficiently large). If the collection of pieces $\mathcal{M}$ on which g is linear and $g(x)$ satisfy certain properties, then one can follow a piecewise linear path from (x^0,T) to $(x^*,0)$ defined by $(x,\theta) \in F^{-1}(0)$. Then $g(x^*) = 0$ and x^* is an approximation to the solution of $f(x) = 0$ in the sense that $\|f(x^*)\|$ is small.

The advantage of these methods is that they do not require that f be differentiable. In fact f can be a point to set mapping which has a closed graph. Thus, Kakutani fixed point problems (Kakutani, [1947]) can be solved. This is important because often it is convenient, if not necessary, to formulate some equilibrium and some optimization problems as Kakutani fixed point problems (see Scarf [1973] and Eaves [1971a]). The disadvantages are that often the triangulation schemes are difficult to arrive at, and for accurate approximation, the grid size may need to be so small that the convergence of the algorithm becomes too slow.

The path methods, which we shall concentrate on in this work, have quite a recent history also. They are very similar, in principle, to the piecewise linear methods in that one considers a system of equations such as

$$F(x,\theta) = f(x) + \theta d = 0 , \qquad x \in S \times [0,T] ,$$

for example. F is defined so that the solution is trivial when $\theta = T$, and under certain conditions $F^{-1}(0) = \{(x,\theta) | F(x,\theta) = 0\}$ defines a path, a set homeomorphic to an interval in R^1, from (x^0,T) to $(x^*,0)$. The primary difference in the requirements of path methods and piecewise linear methods is that in the path methods we require that f be twice continuously differentiable (C^2) rather than merely a closed point-to-set map.

The contributors to this theory fall into two groups. The work before 1975 (Lahaye [1943], [1948], Davidenko [1953a,b], Freudenstein and Roth [1963], Meyer [1968], to name a few) was confined to studying continuation methods which were usually based upon a homotopy $H:D \times [0,1] \to \mathbb{R}^n$, where $H(x,1)$ has a trivial solution and $H(x,0)$ is a solution to $f(x) = 0$, $x \in D$. A typical formulation would let

$$H(x,t) = f(x) + tf(x^0) = 0, \qquad \text{for some } x^0 \in S.$$

The conditions on f which these authors imposed were generally quite strong. A typical assumption would be that f is norm coercive ($\|x^k\| \to \infty$ implies that $\|f(x^k)\| \to \infty$) and that $f'(x)$ is nonsingular everywhere. These conditions would imply that $H^{-1}(0)$ is connected and monotone in t. This assumption is quite difficult to check and is unlikely to be satisfied in many problems.

In 1976 papers were published by Kellogg, Li, and Yorke, S. Smale, and R. Wilson all proposing methods for defining paths which could be followed from a known starting point to a solution. The first two

methods are for general equation solving problems, while Wilson's is for solving a piecewise linear approximation of a Walrasian general equilibrium model. The advantages of the more recent approaches are that no connectedness or monotonicity are required in $H^{-1}(0)$, and the structure of the set S is used to give conditions for success which are easier to verify than those for most continuation methods.

The favorable attributes of the path methods are that global convergence can be guaranteed for a large class of problems, no artificial triangulations need be defined, and efficient iterative equation solving routines can easily be implemented when the solution is nearby. Drawbacks include the requirement that f be at least C^2 and that it is a difficult problem to design a computer algorithm which follows a nonlinear curve in $\mathbb{R}^n$. This last difficulty is not as overwhelming as one might think. An algorithm which is guaranteed to stay near a particular curve will probably be slow, in general, but algorithms have been developed which approximately follow a path and converge to the solution quite quickly. Kellogg, Li, and Yorke [1976] give such an algorithm with numerical results and Chapter VIII shows that path methods provide efficient means to calculating equilibrium points as well.

Another similarity in the papers of Kellogg, Li, and Yorke, Smale, and Wilson was the use of Sard's Theorem, and other results from the theory of differential topology to prove their theorems. Since the reader may not be familiar with these results, we present them in the next chapter.

CHAPTER II

DIFFERENTIAL TOPOLOGY AND SUBDIVIDED COMPLEXES

II.1. Background in Differential Topology

In order to prove global convergence theorems for path methods, it is very useful to use some of the results from the theory of differential topology. In contrast to analysis, differential topology deals with the properties of sets and functions "in the whole" as well as locally. This branch of mathematics explores the relationship of functions $f:M \to N$ and the sets $f^{-1}(y)$ for values $y \in N$. The definitions and results in this section are from Milnor [1965] and Hirsch [1976]. The theorems stated here can be found in either work.

A map $f:M \to N$, where $M \subset \mathbb{R}^m$ and $N \subset \mathbb{R}^n$, is of class C^r if all r^{th} order partial derivatives exist and are continuous. A map f is smooth if it is of class C^r for every $r \geq 1$. A map $f:M \to N$ is called a diffeomorphism (resp., C^r-diffeomorphism) if f carries M homeomorphically onto N and both f and f^{-1} are smooth (of class C^r).

Definition 1.1. A subset $M \subset \mathbb{R}^k$ is called a smooth manifold (C^r-manifold) of dimension m iff each $x \in M$ has a neighborhood $W \cap M$, where W is an open subset of $\mathbb{R}^k$, which is diffeomorphic (C^r-diffeomorphic) to an open subset $U \subset \mathbb{R}^m$.

We shall refer to a manifold of dimension m as an m-manifold.

Definition 1.2. A circle is the set $\{x \in \mathbb{R}^2 \mid \|x\| = 1\}$. An interval is any convex set in $\mathbb{R}^1$ containing more than one point.

Theorem 1.3. Any C^r 1-manifold is a disjoint union of connected sets each of which is C^r-diffeomorphic to either a circle or an interval of real numbers, $r \geq 1$.

The usefulness of this result is that if we know that a set $f^{-1}(u)$ is a C^r 1-manifold, and x^0 is a boundary point of that manifold, then, if one could follow that connected component of $f^{-1}(u)$ with x^0 as a boundary point, one would never return to a point which has already been visited. Thus, the component of $f^{-1}(u)$ must either have one other boundary point x^* or it must be diffeomorphic to a ray. The importance of this property for the convergence proofs presented below is analogous to the importance of the celebrated "Lemke argument" (Lemke [1965], Scarf [1973], Eaves [1976]) in the theory of solving piecewise linear equations.

To define the notion $f'(x)$ for a C^r map $f : M \to N$ of C^r manifolds, $r \geq 1$, we first associate with each $x \in M \subset \mathbb{R}^k$ a linear subspace $TM_x \subset \mathbb{R}^k$ of dimension m called the tangent space of M at x. Then $f'(x)$ will be a linear mapping from TM_x to TM_y; where $y = f(x)$.

For an open set $U \subset \mathbb{R}^k$, $TU_x = \mathbb{R}^k$. For any C^r map $f : U \to V$ the Jacobian

$$f'(x) = \mathbb{R}^k \to R^\ell$$

is defined as the matrix of partial derivatives, i.e.,

$$f'(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}$$

for $i \in \underline{\ell}$, $j \in \underline{k}$.

Now let us define the tangent space TM_x for an arbitrary, smooth, m-dimensional manifold $M \subset \mathbb{R}^k$. Choose a parametrization

$$g : U \to M \subset \mathbb{R}^k$$

of a neighborhood $g(U)$ of $x \in M$, with $g(u) = x$. Here U is an open subset of $\mathbb{R}^m$. Thinking of g as a mapping from U to $\mathbb{R}^k$, the derivative

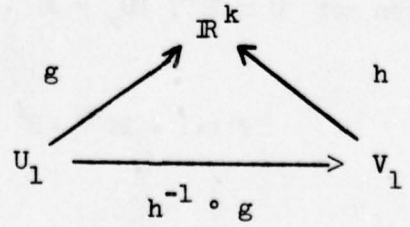
$$g'(u):\mathbb{R}^m \to \mathbb{R}^k$$

is defined. Set TM_x equal to the image of $g'(u)$.

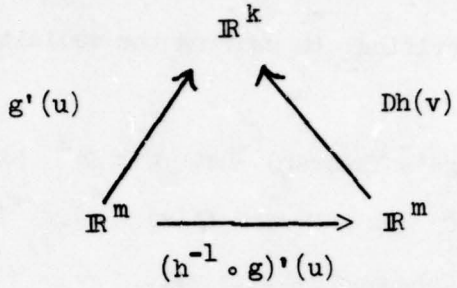
We must prove that this construction does not depend upon the particular choice of parametrization g. Let $h: V \to M \subset \mathbb{R}^k$, $x \in h(X) \subset M$, $v = h^{-1}(x)$,

$$h^{-1} \circ g:U_1 \ni u \to V_1 \ni v \ .$$

The commutative diagram of maps



gives rise to the commutative diagram of linear maps.



It follows immediately that

$$\text{Image}(g'(u)) = \text{Image}\ (h'(v)) \ .$$

Thus TM_x is well defined.

Consider a C^r map $f:M \to N$ from an m-manifold to an n-manifold. Let C be the set of all $y \in M$ such that the jacobian matrix

$$f'(y):\mathbb{R}^m \to \mathbb{R}^n$$

has rank less than n. Then C will be called the set of <u>critical points</u>, $M \setminus C$ the set of regular points, $f(C)$ the set of critical <u>values</u>, and $N \setminus f(C)$ the set of <u>regular values</u> of f.

The following <u>regular value theorem</u> is often used to define manifolds (or in particular, paths).

<u>Theorem 1.4</u>. Let $f = M \to N$ be a C^r map, $r \geq 1$. If $y \in N$ is a regular value, then the set $f^{-1}(y) \subset M$ is a C^r manifold of dimension m-n.

The following result makes precise the statement, if y is in the range of f, then "in general" $f^{-1}(y)$ is a C^r manifold. This theorem is critical in proving the validity of path methods.

Theorem 1.5. (Sard's Theorem) Let $U \subset \mathbb{R}^m$ be open, $f:U \to \mathbb{R}^n$ a C^r map and $C = \{x \in U | \text{rank } f'(x) < n\}$. If $r > \max\{0, m-n\}$ then $f(C)$ has Lebesgue measure zero.

Since we are concerned primarily with 1-manifolds defined by $f^{-1}(y)$ for some $y \in N$, the weakest differentiability condition that can be specified for f, and still be able to apply Sard's Theorem, is that f be C^2.

Definition 1.6. Consider the half-space $H^m = \{(x_1,\ldots,x_m) \in \mathbb{R}^m | x_n \geq 0\}$. The boundary ∂H^m is $\mathbb{R}^{m-1} \times 0 \subset \mathbb{R}^m$. A subset $X \subset \mathbb{R}^k$ is called a C^r manifold with boundary if each $x \in X$ has a neighborhood $U \cap X$ C^r-diffeomorphic to an open subset of H^m. The boundary ∂X is the set of all points in X which correspond to points of ∂H^m under such a diffeomorphism.

Now consider a C^r map $f: X \to N$, $r \geq 1$, from an m-manifold with boundary to an n-manifold where $m > n$.

Theorem 1.7. If $y \in N$ is a regular value for both f and for the restriction $f|\partial X$, then $f^{-1}(y) \subset X$ is a C^r $(m-n)$-manifold with boundary. Furthermore, $\partial(f^{-1}(y))$ is precisely equal to $f^{-1}(y) \cap \partial X$.

This is the form of the regular value theorem which will be most useful in proving the global characteristics of certain paths defined in the chapters below.

II.2. Orientation

In this section we present some standard notions of orientation which will be useful to use when we prove results concerning the orientation of paths in Section III.2. The following definitions are again an amalgamation of those in Milnor [1965] and Hirsch [1976].

Definitions. An orientation for a finite dimensional real vector space V is an equivalence class of ordered bases as follows: The ordered basis $(b_1,\ldots, b_n) = B$ determines the same orientation as the basis $(b'_1,\ldots, b'_n) = B'$ if $B' = BA$ where A is an $n \times n$ matrix with $\det A > 0$. It determines the opposite orientation if $\det A < 0$. The vector space $\mathbb{R}^n$ has a standard orientation corresponding to the basis $(e_1,\ldots, e_n)$ where e_i is the i^{th} unit vector denoted by ω_n.

In the case of a zero dimensional vector space it is convenient to define an "orientation" as the symbol +1 or -1.

We shall denote an orientation ω of V by $\omega = [b_1,\ldots, b_n]$ as the equivalence class of bases with the same orientation as V. (V,ω) is an oriented vector space. $-\omega$ denotes the opposite orientation to ω.

If $L\ V \to W$ is an isomorphism of vector spaces and $\omega = [e_1, \ldots, e_n]$ is an orientation of V then $L(\omega) = [Le_1, \ldots, Le_n]$ is the induced orientation of W. Given (V,ω) and (W,ω'), an isomorphism $L:V \to W$ is called orientation preserving if $L(\omega) = \omega'$; otherwise L is orientation reversing.

An oriented manifold consists of a V smooth manifold M together with a choice of orientation ω_x for each tangent space TM_x. If $m \geq 1$ these are required to fit together as follows: For each $x \in M$ there should exist a neighborhood $U \subset M$ and a diffeomorphism h mapping U onto an open subset of $\mathbb{R}^m$ such that $h'(x)(\omega_x) = (\omega_m)$ for each $x \in U$. If M is connected and orientable then it has precisely two orientations.

If M has a boundary then we can distinguish three kinds of vectors in the tangent space TM_x at a boundary point:

1) There are the vectors tangent to the boundary, forming an $(m-1)$-dimensional subspace $T(\partial M)_x \subset TM_x$.
2) There are the "outward" vectors, forming an open half space bounded by $T(\partial M)_x$; and
3) There are the "inward" vectors forming a complementary half-space.

Each orientation ω for M determines an orientation for ∂M as follows. For $x \in \partial M$ choose a basis B such that $[b_1, \ldots, b_n] = \omega(M)$ and in such a way that $b_1, b_2, \ldots, b_{m-1}$ are contained in $T(\partial M)_x$ and so that b_m is an outward vector. Then $[b_1, \ldots, b_{m-1}] = \omega_x(\partial M)$, the induced orientation of ∂M at x.

If dim M = 1, then each boundary point x is assigned the orientation -1 or +1 according as a positively oriented vector at x points inward or outward (Figure 2.1).

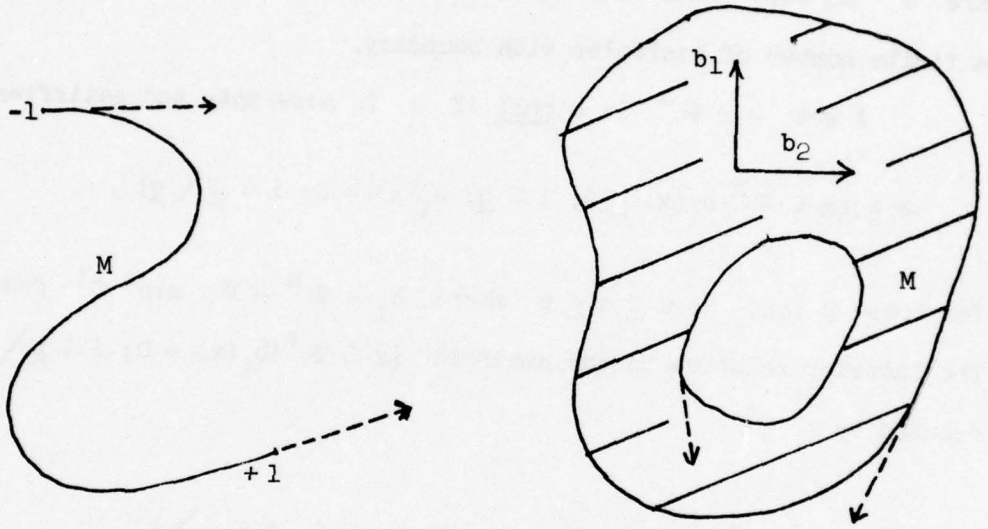


FIGURE 2.1

II.3. Subdivided Complexes

In this section we will essentially be extending the piecewise linear topology developed by Eaves [1976] to the case when the functions are C^2 on each "cell" and the cells are defined by an intersection of a finite number of manifolds with boundary.

A set $\sigma \subset \mathbb{R}^n$ is a cell if σ is non-empty and satisfies

$$\sigma = \{x \in \mathbb{R}^n | b_i(x) \leq 0,\ i \in \underline{q};\ b_i(x) = 0;\ i \in \underline{p} \setminus \underline{q}\}\ ,$$

for some p and q, $0 \leq q \leq p$ where $b_i = \mathbb{R}^n \to \mathbb{R}$ are C^2 functions. The interior relative to the manifold $\{x \in \mathbb{R}^n | b_i(x) = 0;\ i \in \underline{p} \setminus \underline{q}\}$ denoted by

$$\overset{0}{\sigma} \equiv \{x \in \mathbb{R}^n | b_i(x) < 0,\ i \in \underline{q};\ b_i(x) = 0,\ i \in \underline{p} \setminus \underline{q}\}$$

is also assumed to be non-empty. Let

$$A(x) = \{i | b_i(x) = 0,\ i \in \underline{p}\} \qquad \text{for } x \in \sigma$$

be the set of binding constraints of x relative to σ.

We shall always assume that, for $x \in \sigma$, that $\{\nabla b_i(x) | i \in A(x)\}$ is a linearly independent set of vectors (if $A(x) \neq \emptyset$). This is a commonly used constraint qualification in nonlinear programming.

A cell $\sigma \subset \mathbb{R}^n$ is called a cell of dimension m, or an m-cell, if each $x \in \overset{0}{\sigma}$ has a neighborhood $W \cap \sigma$ which is diffeomorphic to an open subset U of $\mathbb{R}^m$.

Suppose the index set $B \subset \underline{p}$ is equal to $A(x)$ for some $x \in \sigma$. Then

$$\beta \equiv \{x | b_i(x) \leq 0,\ i \in q \setminus B,\ b_i(x) = 0,\ i \in B\}$$

is a <u>face</u> of σ. Notice that $B \supset p \setminus q$. B is called the index set associated with β. Also note that β could be all of σ or a single point.

<u>Lemma 3.1</u>. If a face $\beta \subset \sigma$ has an associated index set B, then let $r = |B|$. Then if $\sigma \subset \mathbb{R}^n$, $r \leq n$ and β is a cell of dimension n-r.

<u>Proof</u>. Let $S \supset \sigma$ be a smooth manifold. Define the C^2 map

$$h : S \to \mathbb{R}^r \quad \text{as} \quad h(x) = \begin{pmatrix} \vdots \\ b_i(x) \\ \vdots \end{pmatrix}_{i \in B}.$$

Then, by definition of σ, for any $x \in \beta$, $\{\nabla b_i(x) | i \in B\}$ is a linearly independent set, or equivalently, $h'(x)$ is of rank r for any $x \in \beta$. Thus, 0 is a regular value of h, and by Theorem 2.4, $h^{-1}(0)$ is a C^2 manifold of dimension of dimension n-r. Hence, $\beta = h^{-1}(0) \cap \sigma$ is an (n-r)-cell. ∎

An (m-1)-face of an m-cell is called a <u>facet</u> of the cell.

Let $\mathcal{M} \neq \varphi$ be a finite or countable collection of m-cells in some Euclidean space. Let $\mathcal{M}^i$ for $i = 0, \ldots, m$ be the set of i-faces of the elements of $\mathcal{M}$; we call members of $U_{i=0,\ldots,m} \mathcal{M}^i$ and $\mathcal{M}^0$ cells and vertices of $\mathcal{M}$, respectively.

Definition 3.2. Let $M = \bigcup_{\sigma \in \mathcal{M}} \sigma$. We call $(M, \mathcal{M})$ a subdivided m-complex if

a) any two m-cells of $\mathcal{M}$ are disjoint or meet in a common face,

b) each (m-1)-cell of $\mathcal{M}$ lies in at most two m-cells, and

c) each point of M has a neighborhood meeting only finitely many m-cells of $\mathcal{M}$.

Example 3.3. Let σ be an m-cell, then $(M, \mathcal{M}) = (\sigma, \{\sigma\})$ is a subdivided m-complex.

The next result will be useful in the definition of a path method in Section IV.3.

Proposition 3.4. If $\mathcal{M} = \{\tau_i | i \in q\}$ is the collection of facets, of an m-cell σ, then $(\bigcup_{i \in q} \tau_i, \mathcal{M})$ is a subdivided (m-1)-complex.

Proof. For simplicity consider a cell σ which is defined only by inequality constraints: $\sigma = \{x \in \mathbb{R}^m | b_i(x) \leq 0, \ i \in q\}$. The cells τ_i are defined as

$$\tau_i = \{x \in \mathbb{R}^m | b_i(x) = 0, \ b_j(x) \leq 0, \ \forall \ j \neq i\} .$$

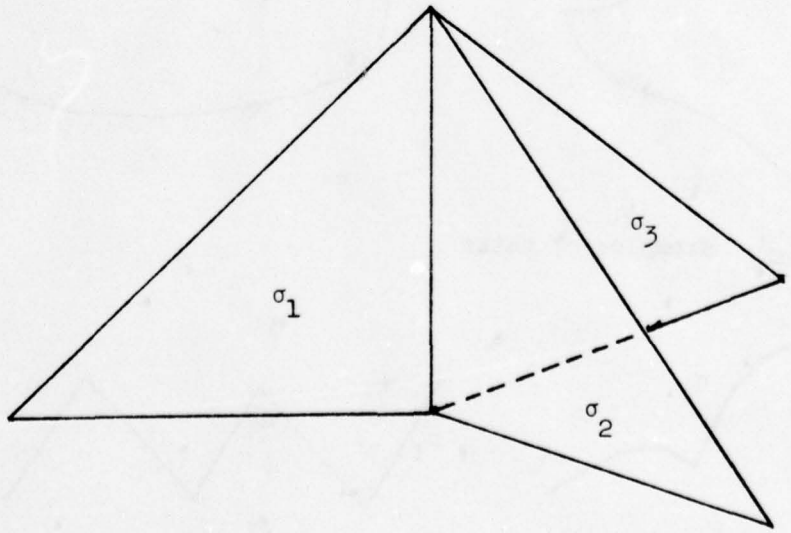
a) To prove that the first part of Definition 3.2 holds, it is sufficient to show that $\tau_j^0 \cap \tau_k^0 \neq \varphi$. Suppose $y \in \tau_j^0 \cap \tau_k^0$ for some $j, k \in q$. Then $b_j(x)$ and $b_k(x)$ are identical in some (m-1)-dimensional

neighborhood $B(y,\epsilon) \cap \tau_j^0$. This implies that $\nabla b_j(y) = \alpha \nabla b_k(y)$ for some $\alpha > 0$, which contradicts the constraint qualification in the definition of an m-cell.

b) Next we show that each (m-2)-cell lies in at most two (m-1)-cells of $\mathcal{M}$. Suppose that an (m-2)-cell γ lies in $\tau_1, \tau_2, \tau_3 \in \mathcal{M}$. Then $\gamma \subset \delta = \{x \in \mathbb{R}^m | b_i(x) = 0, i \in \underline{3}, b_i(x) \leq 0, i \in \underline{q} \setminus \underline{3}\}$. At some point $y \in \gamma^0$ there is an $\epsilon > 0$ such that $\gamma \cap B(y,\epsilon)$ is an (m-2)-manifold. But by Lemma 3.1, δ is a cell of dimension n-3, and we have a contradiction.

c) This last requirement follows from the fact that the number of cells in $\mathcal{M}$ is finite. ∎

Example 3.5. Consider three 2-cells in $\mathbb{R}^3$ which meet in a common facet. Then $(\cup_{i=1,2,3}\sigma_i, \{\sigma_1, \sigma_2, \sigma_3\}$ is not a subdivided 2-complex.



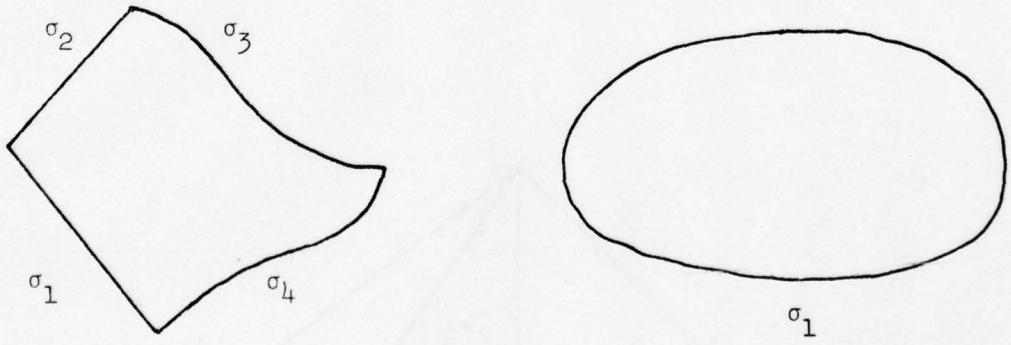
If $(M, \mathcal{\eta})$ is a subdivided m-complex for some subdivision $\mathcal{\eta}$, we call M an m-complex; M may have many possible subdivisions.

The following lemma is trivial to prove, but it is very important for the convergence theorems that follow.

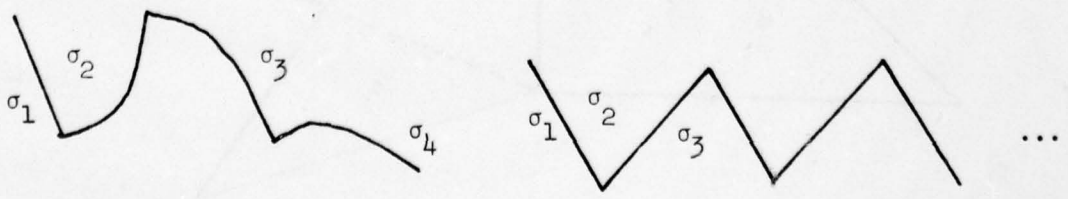
Lemma 3.6. A connected 1-complex is homeomorphic to either a circle or an interval.

For the two cases of 3.6 we call the connected 1-complex a loop or a path. We shall use path and curve synonymously.

Example 3.7. Examples of loops



Example 3.8. Examples of paths



Let M be an m-complex subdivided by $\mathcal{M}$. By the boundary of M, ∂M, we mean the union of all (m-1)-cells of $\mathcal{M}$ which lie in exactly one m-cell of $\mathcal{M}$.

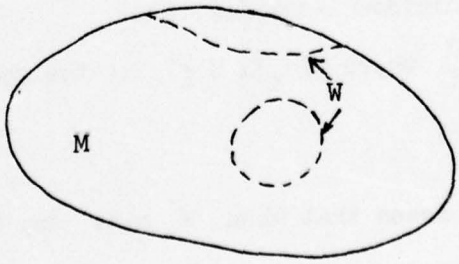
Lemma 3.9. The boundary of a complex is closed in the complex.

Proof. Let $\{x^k\} \subset \partial M$ be such that $\lim_k x^k = x$. If infinitely many of the x_k lie in one m-cell of $\mathcal{M}$, then x is in σ and hence in M. Otherwise, x is not in M due to condition (c) in the definition of a subdivided complex. ∎

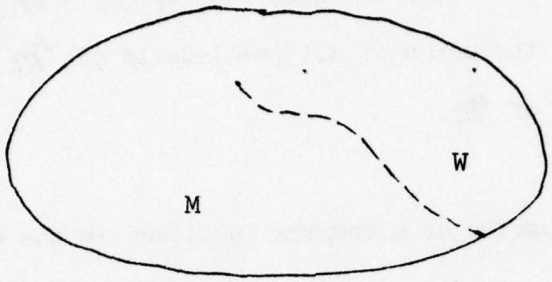
We are interested in the behavior of 1-complexes which are contained in m-complexes for $m > 1$. When a path has certain desirable properties we shall say that it is "neat." Let M be an m-complex and W a 1-complex contained in M. If W is closed in M and $\partial W = W \cap \partial M$ we say that W is neat in M. This terminology is used by Hirsch [1976] with reference to manifolds with boundary. Some examples illustrating the definition are below.

Example 3.10. The 1-complex W (dashed line) is neat in M in the first case, not neat in M for the others.

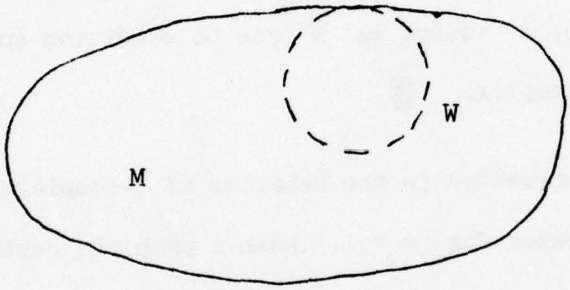
1.



2.



3.



Next we extend the definition of "neat" 1-complexes to deal with subdivisions of M.

Definition 3.11. Suppose that W is a 1-complex contained in M and $\mathcal{M}$ is a subdivision of M. Let $\mathcal{W}$ be composed of loops and paths $\gamma \subset \sigma \cap W$ for each $\sigma \subset \mathcal{M}$. We say that W is neat in $(M, \mathcal{M})$ if

a) $W \cap \sigma$ is neat in σ for any $\sigma \in \mathcal{M}$,

b) $(W, \mathcal{W})$ is a subdivided 1-complex, and

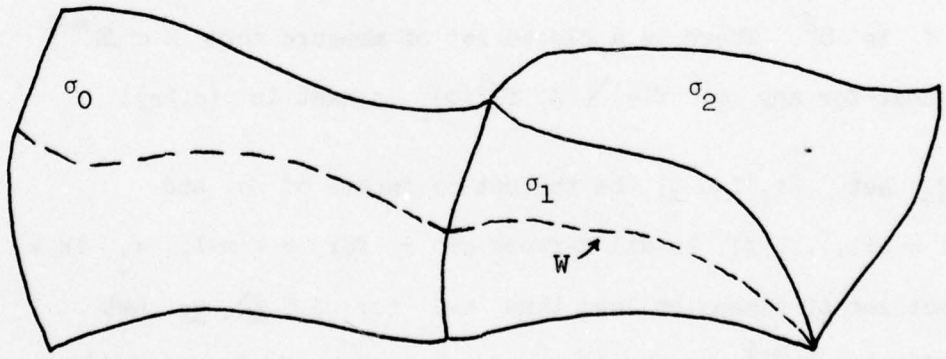
c) $\partial(W \cap \sigma) \subset \bigcup_{i=1}^{\ell} \tau_i^0$ where $\{\tau_i \mid i \in \underline{\ell}\}$ is the set of (m-1)-cells of $\mathcal{M}$ ($\mathcal{M}^{m-1}$).

Condition c) means that when W hits $\partial\sigma$, for any cell σ, it hits only one facet of σ.

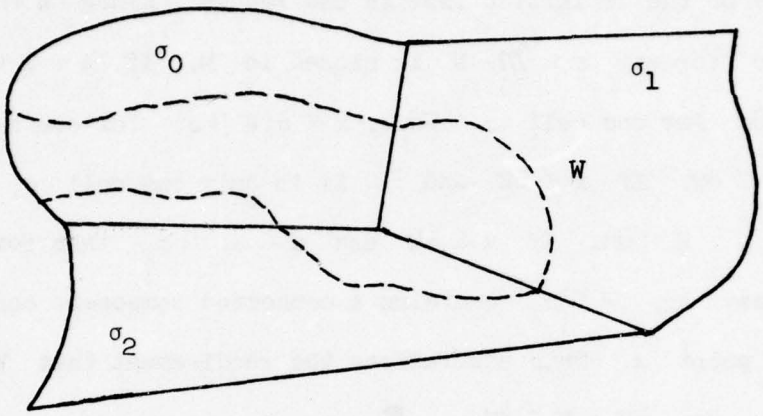
Lemma 3.12. If W is neat in $(M, \mathcal{M})$ then W is neat in M.

Proof. Part a) of the definition implies the result. Since $W \cap \sigma$ is closed in σ for any $\sigma \in \mathcal{M}$, W is closed in M. If $x \in W \cap \partial M$, then $x \in W \cap \partial\sigma$ for one cell σ. Thus, $x \in \partial(W \cap \sigma)$ for one cell, and we have $x \in \partial W$. If $x \in \partial W$ and x is in only one cell σ, it is clear that $x \in W \cap \partial M$. If $x \in \partial W$ and $x \in \sigma_1 \cap \sigma_2$, then for one of the cells, say σ_2, $W \cap \sigma_2$ contains a connected component consisting of the point x. This contradicts the requirement that $W \cap \sigma_2$ is a 1-complex. So, $\partial W = W \cap \partial M$. ∎

Example 3.13. W is not neat in $(M, \mathcal{M})$: $W \cap \sigma_i$ is not in σ_i, $i = 1,2$.



Example 3.14. W is neat in $(M, \mathcal{M})$



Next we prove the theorem which says that for almost all values φ, $f^{-1}(\varphi)$ is neat in σ.

Theorem 3.15. Consider $f:\sigma \to \mathbb{R}^{m-1}$ where σ is a bounded m-cell and f is C^2. There is a closed set of measure zero $Z \subset \mathbb{R}^{m-1}$ such that for any $\varphi \in f(\sigma) \setminus Z$, $f^{-1}(\varphi)$ is neat in $(\sigma, \{\sigma\})$.

Proof. Let $\{\tau_i \mid i \in \underline{q}\}$ be the set of facets of σ, and $\{\tau_i \mid i = q+1, \ldots, \ell\}$ be all r-faces of σ for $r < m-1$. τ_j is a compact set of dimension less than m-1 for $j \in \underline{\ell} \setminus \underline{q}$. Let $\nu_j \equiv f(\tau_j) \subset \mathbb{R}^{m-1}$ be the image of τ_j under f for $j \in \underline{\ell} \setminus \underline{q}$; ν_j is a closed set of dimension less than m-1. Hence, $\nu \equiv \bigcup_{j \in \underline{\ell} \setminus \underline{q}} \nu_j$ is a closed set of measure zero in $\mathbb{R}^{m-1}$, and if $\varphi \notin \nu$, then $f^{-1}(\varphi) \cap \partial\sigma \subset \bigcup_{i=1}^{q} \tau_i^0$.

Since τ_i, $i \in \underline{q}$ is contained in $\{x \in \mathbb{R}^n | b_i(x) = 0, b_j(x) = 0, j \in \underline{p} \setminus \underline{q}\}$, a C^2-(m-1)-manifold, there is a compact C^2-manifold with boundary $X_i \cup \partial X_i$ such that $\tau_i \subset \partial X_i$ and $\sigma \subset X_i$. By Theorem 2.5 (Sard's) and Theorem 2.7, the set $C_i \subset \mathbb{R}^{m-1}$ of critical values for $f|X_i$ and for $f|\partial X_i$ has measure zero for each $i \in \underline{q}$. Also, for $\varphi \in \mathbb{R}^{m-1} \setminus C_i$, $f^{-1}(\varphi)$ is a 1-manifold neat in $X_i \cup \partial X_i$.

To show that C_i is closed, let $\mathcal{J}$ be a convergent sequence $\{\varphi^k\} \subset C_i$ together with its limit point $\bar{\varphi}$. $f^{-1}(\mathcal{J})$ must contain a sequence $\{x^k\}$ such that $f(x^k) = \varphi^k$ for every k. Since $X_i \cup \partial X_i$ is compact, $\{x^k\}$ has a limit point $\bar{x}$. We know that $f'(x)$ is continuous and that every minor of size m-1 is zero for $x \in f^{-1}(\mathcal{J})$. Hence, $\bar{x}$ is a critical point, and, by continuity, $f(x) = \bar{\varphi}$ and $\bar{\varphi}$ is in C_i.

Since $i \in \underline{q}$ was arbitrary, $\cup_{i \in \underline{q}} C_i = C$ is a closed set of measure zero. We now have that $Z \equiv C \cup \nu$ is a closed set of measure zero. All that is left to show is that for $\varphi \in f(\sigma) \setminus Z$, $\partial(f^{-1}(\varphi)) = f^{-1}(\varphi) \cap \partial\sigma$.

For $x \in \partial\sigma \cap f^{-1}(\varphi)$, there is some i such that $x \in \tau_i^0$. Thus, there is some neighborhood $\mathcal{O}$ of x such that $\mathcal{O} \cap \sigma = \overline{\mathcal{O}} = \mathcal{O} \cap (X_i \cup \partial X_i)$. Since $\varphi \notin C$, $x \in f|_{X_i}^{-1}(\varphi) \cap \partial X_i = \partial f|_{X_i}^{-1}(\varphi)$. f is identical to $f|_{X_i}$ on $\overline{\mathcal{O}}$ so $x \in \partial f^{-1}(\varphi)$, and we have shown $(\partial\sigma \cap f^{-1}(\varphi)) \subset \partial f^{-1}(\varphi)$.

Suppose $x \in \partial f^{-1}(\varphi)$. If $x \in \partial\sigma$, an argument similar to the one above would show that $x \in \partial\sigma \cap f^{-1}(\varphi)$. If $x \notin \partial\sigma$, then for any

X_i, $x \in X_i$ but $x \notin \partial X_i$. Thus, $x \notin \partial f|_{X_i}(\varphi)$. But, on σ, $f|_{X_i}$ is identical to f, and we have $x \notin f^{-1}(\varphi)$, a contradiction. Therefore, the case $x \in \partial\sigma$ is the only one, and we are done. ∎

We will call values φ which are in $f(\sigma) \setminus Z$, as defined in the last theorem, good values of f with respect to σ.

In the theorems that follow, the assumption is commonly made that 0 is a good value of f with respect to σ. Theorem 3.11 implies that if this assumption is true, then there is a neighborhood of 0 consisting of good values. This is important from a computational point of view because in this case if the algorithm stays in $B(f^{-1}(0),\epsilon)$, for some $\epsilon > 0$, no singularities or branch points of a path will be encountered.

Another useful fact is that if 0 is a critical value, then any neighborhood $\mathcal{O}$ of 0 contains a regular value φ. However, it is possible that there are no good values in $\mathcal{O}$. For our applications of this theory, however, it will always be true that for some facet τ, $f|_\tau(x) = 0$ will have a unique solution x^0, and $f|'_\tau(x^0)$ will be of full rank. By the inverse function theorem, any sufficiently small neighborhood $\mathcal{O}$ of 0 will be in the range of f, and, hence, we can conclude that $\mathcal{O}$ contains a good value $\hat{\varphi}$ of f with respect to σ. Thus, when we make the assumption that 0 is a good value for f in the sequel, we make it with the understanding that if this is not true we can deal with a perturbation $\hat{\varphi}$ of 0.

A more satisfactory resolution of this regularity problem would be to use some deeper transversality results of differential topology (Hirsch, Section 3.2, [1976]). With these results we could perturb f to $\bar{f}$ if 0 was not a good value of f w.r.t. σ. We have avoided this course because of the excessive amount of preliminaries which it would require.

It is important to note that if 0 is a good value of f, then $x \in f^{-1}(0) \cap \partial\sigma_1$ means that x is in only one facet of σ_1. Hence, there is at most one adjacent cell σ_2 in which one could continue to follow $f^{-1}(0)$.

We say that φ is a good value of f with respect to (w.r.t.) $(M, \mathcal{M})$ if φ is a good value of f w.r.t. σ for every $\sigma \in \mathcal{M}$. Notice that if the subdivision of $\mathcal{M}$ is finite, then Theorem 3.15 immediately yields the following

Corollary 3.16. Let $(M, \mathcal{M})$ be a bounded subdivided m-complex and $|\mathcal{M}|$ be finite. If $f: M \to \mathbb{R}^{m-1}$ is C^2, then there is a closed set of measure zero $Z \subset \mathbb{R}^{m-1}$ such that the set of good values Φ of f w.r.t. $(M, \mathcal{M})$ can be written

$$\Phi = f(M) \setminus Z.$$

The next theorem is our main result concerning good values.

Theorem 3.17. Let $f : M \to \mathbb{R}^m$ be a continuous map such that $f|_\sigma$ is C^2 on each $\sigma \subset \mathcal{M}$, where $(M, \mathcal{M})$ is a subdivided $(m+1)$-complex, and each $\sigma \subset \mathcal{M}$ is bounded. If φ is a good value of f w.r.t. $(M, \mathcal{M})$ then $f^{-1}(\varphi)$ is a 1-complex neat in $(M, \mathcal{M})$.

Proof. Since φ is a good value for $f|_\sigma$ for any $\sigma \in \mathcal{M}$, $f^{-1}(\varphi) \cap \sigma$, is neat in $(\sigma, \{\sigma\})$ for each $\sigma \in \mathcal{M}$, by Theorem 3.14. Thus, parts 1) and 3) of Definition 3.10.5 are demonstrated. Let $W = f^{-1}(\varphi)$ and $\mathcal{W}$ be composed of the connected components $\gamma \subset (\sigma \cap W)$ for each $\sigma \in \mathcal{M}$. All that is left to prove is that $(W, \mathcal{W})$ is a subdivided 1-complex. We now show that the three conditions of Definition 3.3. are satisfied.

a) Suppose that $\lambda_1 \cap \lambda_2 \neq \emptyset$ for $\lambda_1, \lambda_2 \in \mathcal{W}$, then $\lambda_1 \subset \sigma_1 \cap f^{-1}(\varphi)$ and $\lambda_2 \subset \sigma_2 \cap f^{-1}(\varphi)$, by the definition of segments. It is clear that $\lambda_1^0 \subset \sigma_1^0$ and $\lambda_2^0 \subset \sigma_2^0$ because $\partial\lambda_i \subset \partial\sigma_i$, $i = 1,2$. From $\sigma_1^0 \cap \sigma_2^0 = \emptyset$ we have that λ_1 and λ_2 must meet in a boundary point.

b) If y is a boundary point of distinct $\mu_i \in \mathcal{W}$ for $i = 1,2,3$ then $y \in \tau^0 \subset \sigma_i$ for $i \in \underline{3}$, where τ is a facet of $\sigma_i \supset \mu_i$, $i \in \underline{3}$. This contradicts the fact that $\mathcal{M}$ is a subdivision of M.

c) For any $y \in \lambda_1 \subset f^{-1}(\varphi)$, if $y \in \sigma_1$ for only one $\sigma_1 \in \mathcal{M}$ then there is a neighborhood $\mathcal{O}$ of y for which $\mathcal{O} \cap f^{-1}(\varphi)$ meets only λ_1. If $y \in \sigma_2$ also, then b) implies that y is a boundary point of exactly two segments λ_1 and λ_2. Since φ is a regular value for $f|_{\sigma_1}$, $f|_{\sigma_2}$, $f|_{\sigma_1 \cap \sigma_2}$, one can find a neighborhood of y containing only λ_1 and λ_2. ∎

The following proposition is of critical importance in applications of the theory of subdivided manifolds.

Proposition 3.18. If the subdivision of $\mathcal{M}$ is finite and M is bounded, then $f^{-1}(\varphi)$ contains an even number of boundary points.

Proof. Let $|C|$ denote the number of connected components in the set C. To show that the subdivision $\mathcal{W}$ of $f^{-1}(\varphi)$ is finite, all we need show is that $|f^{-1}(\varphi) \cap \sigma|$ is finite for any $\sigma \in \mathcal{M}$. Since σ is closed and M is bounded, σ must be compact. Suppose there are an infinite number of distinct connected components of $f^{-1}(\varphi) \cap \sigma$, $\{\lambda_i\}$, $i = 1,2,\ldots$. Then pick a point $x^i \in \lambda_i$, $i = 1,2,\ldots$. Since σ is compact, we can, by choosing a subsequence if necessary, assume that $\lim_{i \to \infty} x^i = \bar{x} \in \sigma$, and $f(\bar{x}) = \varphi$ by continuity. By making a C^2 extension of f to an open set $C \subset \sigma$ we can use the fact that φ is a regular value of f to apply the implicit function theorem [Ortega and Rheinboldt (1970), p. 128] and conclude that there is a *neighborhood* $\mathcal{O}$ *of* $\bar{x}$ and a C^1 curve $x(t):(t_1,t_2) \to C$, $t_1 < 0 < t_2$, $x(0) = \bar{x}$, such that $f^{-1}(\varphi) \cap \mathcal{O} = \{x(t) | t \in (t_1,t_2)\}$ which contradicts the fact that $\bar{x}$ was a cluster point of $\{x^i\}$. Thus the subdivision of $\mathcal{W}$ is finite.

Consider any connected component $\gamma \subset f^{-1}(\varphi)$. If γ is a loop it has no boundary points. If γ has one boundary point $x^0 \subset \lambda^0 \subset \gamma$, then it must contain another: The subdivision of γ must be finite $= \{\lambda^0, \lambda^1, \ldots, \lambda^q\}$. $x^q \in \lambda^{q-1} \cap \lambda^q$ is a boundary point of λ^q.

If $\lambda^q \subset \sigma$, then $\sigma \cap f^{-1}(\varphi)$ is neat in σ and the fact that λ^q is closed in the compact set σ imply that there is another point x^{q+1} of λ^q and $x^{q+1} \subset \tau^0$ where τ is a facet of σ. x^{q+1} is in no other cell σ', otherwise λ^q would not be the last segment in the connected set γ. Thus x^{q+1} is the other boundary point of γ. Since this argument could be repeated for each of the finite number of components $\gamma \subset f^{-1}(\varphi)$, the number of boundary points of $f^{-1}(\varphi)$ is even. ∎

Corollary 3.19. If x^0 is a boundary point of $f^{-1}(\varphi)$ where φ is a good value and M is compact, $x^0 \in \partial M$ and the other boundary point x^* of the path γ containing x^0 is in the boundary of M.

CHAPTER III

INTRODUCTION TO PATH FOLLOWING ALGORITHMS

III.1. Objectives

In this chapter we present some of the possible algorithms for following $F^{-1}(0)$ for some deformation $F : M \to N$ which is defined in terms of f where $M \subset \mathbb{R}^{n+1}$, $N \subset \mathbb{R}^{n}$. We will not discuss the exact manner in which F is defined--that is the subject of the next chapter. We will merely be assuming that an initial point (x^0, θ^0) is available which is in $\partial M \cap F^{-1}(0)$, and if we can follow the component $\gamma \subset F^{-1}(0)$ containing (x^0, θ^0) until we reach the other point $(x^*, \theta^*) \in \gamma \cap \partial M$, we will have found a solution x^* to

$$f(x) = 0 , \qquad x \in S . \tag{1.1}$$

We will assume that γ is a C^1-diffeomorphism of a closed interval in $\mathbb{R}^1$ (which will be true if 0 is a good value for F w.r.t. M). Hence, there is some differentiable parametrization $\varphi = [0,T] \to M$ such that $\gamma = \{(x,\theta) \,|\, (x,\theta) = \varphi(t) \text{ for some } t \in [0,T]\}$ and $\varphi(0) = (x^0, \theta^0)$. Clearly then, we have

$$F(\varphi(t)) = 0 , \qquad t \in [0,T] . \tag{1.2}$$

Differentiating (1.2) with respect to t we get

$$F'(\varphi(t)) \cdot \frac{d\varphi(t)}{dt} = 0 .$$

Define $\dot{\varphi}(t) = d\varphi/dt$. We can completely specify $\varphi(t)$ as the solution to the initial value problem

$$F'(\varphi(t))\ \dot{\varphi}(t) = 0\ , \tag{1.3a}$$

$$\varphi(0) = (x^0, \theta^0)\ , \tag{1.3b}$$

$$\varphi(t) \in M \quad \text{for} \quad t > 0\ , \tag{1.3c}$$

$$\|\dot{\varphi}(t)\| = 1\ . \tag{1.3d}$$

(cf. Kellogg, Li, and Yorke [1976]).

The condition (1.3c) fixes the sign of the initial tangent vector $\dot{\varphi}(0)$, and requires it to point into M. The condition (1.3d) implies that the parameter t is the arc length along the curve. Under some rather strong conditions, numerical integration techniques such as Euler's method or the Runge-Kutta technique can be used to find the end-point $(x^*, \theta^*) = \varphi(T)$ (Ortega and Rheinboldt, pp. 339-340 [1970]). Kellogg, Li, and Yorke [1976] describe a similar algorithm under more general conditions, but do not give any convergence proof.

We will briefly describe an algorithm similar to the one described in Kellogg, Li, and Yorke, so that we can contrast it with the one for which we will prove a convergence theorem. This algorithm takes short steps in the direction of $\dot{\varphi}(t)$ as described in (1.3) and shortens the steplength if $\|F(x, \theta)\|$ changes too much. An assumption of this algorithm is that $\theta^0 = 0$ and θ is monotonically increasing along γ. The algorithm is as follows:

Algorithm 1.1.

0. Pick $(x^0,0) \in M$ such that $F(x^0,0) = 0$, and pick numbers $h > 0$, $\epsilon > 0$. Set $i = 0$.

1. Compute $F'(x^i,\theta^i)$.

2. Compute $(\dot{x},\dot{\theta})$ such that

$$F'(x^i,\theta^i)\begin{pmatrix}\dot{x}\\ \dot{\theta}\end{pmatrix} = 0 \quad \text{and} \quad \dot{\theta} > 0 .$$

3. Replace $(\dot{x},\dot{\theta})$ by $(\dot{x},\dot{\theta})/\|(\dot{x},\dot{\theta})\|$.

4. Set $(x^{i+1},\theta^{i+1}) = (x^i,\theta^i) + h(\dot{x},\dot{\theta})$

5. If $\|F(x^{i+1},\theta^{i+1}) - F(x^i,\theta^i)\| > \epsilon$, replace h by $h/2$ and go to Step 4. Otherwise, go to next step.

6. If $(x^{i+1},\theta^{i+1}) \notin M^0$, exit with (x^{i+1},θ^{i+1}) as the approximate solution to the problem of finding $F^{-1}(0) \cap \partial M$. Otherwise, replace i by $i+1$ and return to Step 1.

Clearly the approximate solution derived from this algorithm could be used as a starting point for some iterative technique such as Newton's method to find a more accurate solution. A possible course of Algorithm 1.1 is given in Figure 1.2.

Notice that after the third iteration the stepsize was cut in half due to Step 5. In this case the algorithm performed well and resulted in a good approximation to (x^*,θ^*).

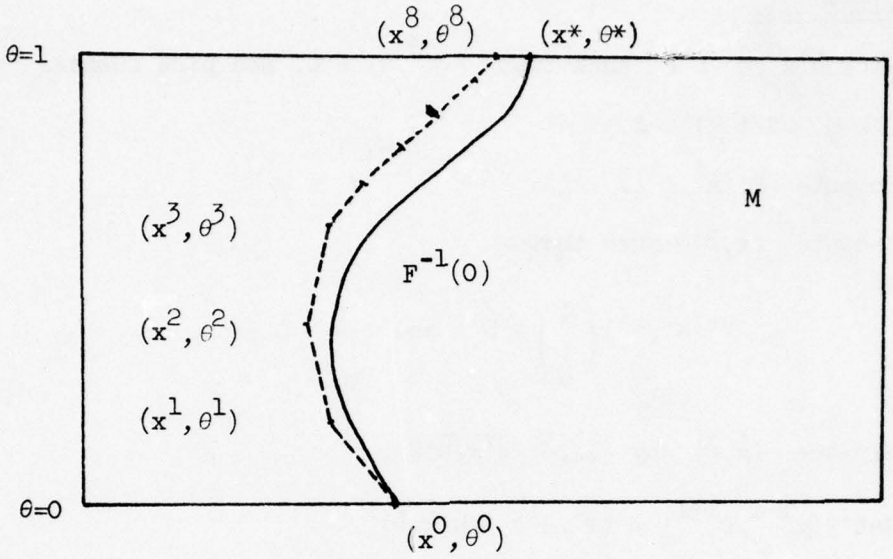


FIGURE 1.2

Next we illustrate two possible shortcomings of Algorithm 1.1.

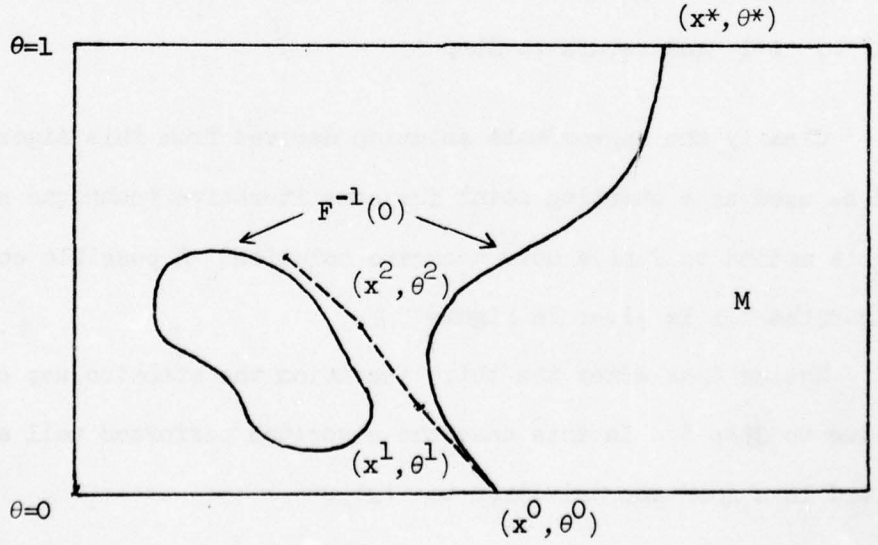
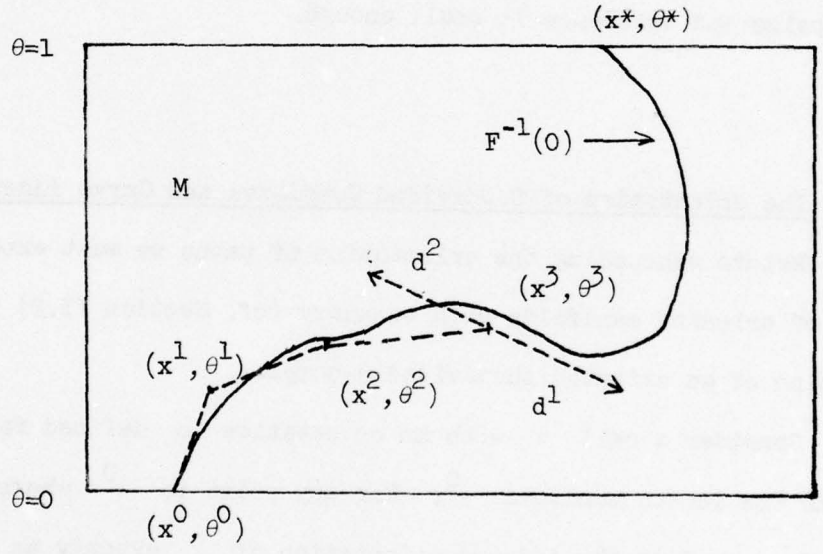


FIGURE 1.3

After the jump from (x^1, θ^1) to (x^2, θ^2), $F(x^2, \theta^2)$ is still near zero but (x^2, θ^2) is near a new component of $F^{-1}(0)$, a loop. The algorithm will fail to reach (x^*, θ^*). Even with a very small step size h and tolerance ϵ it appears difficult to prove that this type of behavior will never occur.

The next difficulty appears even more disadvantageous. It is caused by the naive method for choosing the sign of $(\dot{x}, \dot{\theta})$ in Step 2.



At the point (x^3, θ^3), the algorithm would choose d^2 as the direction for $(\dot{x}, \dot{\theta})$ while d^1 would be the direction in which to continue to produce a consistent movement along $F^{-1}(0)$. Any time $d\theta/dt = 0$ along $F^{-1}(0)$, Algorithm 1.1 is going to have difficulties. What is needed is a way to determine orientation which is not dependent upon a notion of monotonicity in any variable.

In the next section we will prove the necessary result concerning orientation. It says that the sign of $\dot{x}_i$, for some $i \in \underline{n}$ (or $\dot{\theta}$) is determined by the sign of the determinant of the matrix formed by removing the column corresponding to x_i (or θ) from $F'(x,\theta)$.

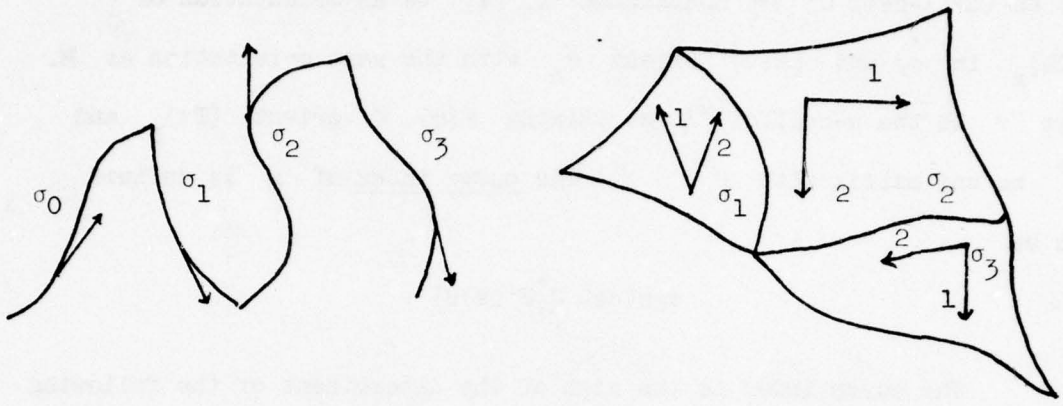
In the following section we shall describe an algorithm which uses the orientation result of Section 2 and returns to the curve $F^{-1}(0)$ periodically to prevent the situation of Figure 1.2 from occurring if the stepsize and tolerance is small enough.

III.2. The Orientation of Subdivided Complexes and Curve Index

Before discussing the orientation of paths we must extend the notion of oriented manifolds with boundary (cf. Section II.2) to a definition of an oriented subdivided m-complex.

Consider a cell σ with an orientation ω defined for each point in the smooth manifold σ^0. For any point in τ^0 where τ is a facet of σ, define the induced orientation of τ exactly as if τ was contained in $\partial\Sigma$ where $\Sigma \supset \sigma^0$ is a C^2-manifold with the same orientation as σ.

Let $(M, \mathcal{M})$ be a subdivided m-manifold with $m \geq 1$, and each m-cell of $\mathcal{M}$ is oriented. Let τ be a facet contained in two m-cells σ_1 and σ_2. If τ receives an opposite orientation from σ_1 and σ_2, we say that $(M, \mathcal{M})$ is oriented. When we say that $(M, \mathcal{M})$ is positively oriented we shall refer to the orientation of each m-cell as being positive.



Example 2.1. Oriented subdivided complexes.

Notice that if $(M, \mathcal{M})$ is a subdivided 2-complex that a "clockwise" or "counterclockwise" property of the ordered bases determine the orientation of a cell. If, in rotating from b_1 to b_2 in the shortest direction, one rotates in a clockwise direction for all $\sigma \in \mathcal{M}$, then $(M, \mathcal{M})$ is oriented.

If M is an m-complex in $\mathbb{R}^m$ we shall by convention give it the standard orientation ω_m. That is, each m-cell is oriented by ω_m. It is easy to verify that $(M, \mathcal{M})$ is oriented in this case.

Next we discuss the important concept of curve index. The development in this section is, again, a generalization of the results on curve index in Eaves [1976]. It is the last result in this section, Proposition 2.3, which allows us to decide which direction to move in a discrete path following algorithm.

Let $F : M \to N$ by a map from the oriented $(n+1)$-complex $(M, \mathcal{M})$ to the oriented n-complex $(N, \mathcal{N})$. Let y be a good value, $(W, \mathcal{W})$ be the 1-complex $F^{-1}(y)$, and $(W, \mathcal{W})$ have an orientation. Let σ

be an (n+1)-cell of $\mathcal{M}$ containing x, [v] be an orientation of $(TW)_x$ in σ, and [B,v] orient σ_x with the same orientation as M. Let τ be the n-cell of $\mathcal{N}$ containing $F(\sigma)$, C orient $(T\tau)_y$ and C^+ be any matrix with $C^+C = I$. The <u>curve index</u> of x is defined to be

$$\operatorname{sgn}(\det C^+F'(x)B) .$$

The curve index is the sign of the determinant of the following set of maps

$$\begin{array}{ccc} & F'(x) & \\ T\sigma_x & \longrightarrow & T\tau_y \\ \uparrow B & & \downarrow C^+ \\ \mathbb{R}^n & & \mathbb{R}^n \end{array}$$

if $M \subset \mathbb{R}^q$ and $N \subset \mathbb{R}^p$ then $C^+ \in \mathbb{R}^{n\times p}$, $F'(x) \in \mathbb{R}^{p\times q}$, and $B \in \mathbb{R}^{q\times n}$.

<u>Lemma 2.2</u>. The curve index is well-defined, nonzero and constant on any oriented path.

<u>Proof.</u> Part 1) shows that the curve index is independent of the choice of B and C. Part 2) shows that it is independent of $x \in W$. First we note that for points a in $T\tau_y$, $a = C\lambda$ for some $\lambda \in \mathbb{R}^p$, and

$$C^+a = C^+C\lambda = \lambda ,$$

so

$$CC^+a = C\lambda = a . \tag{2.1}$$

If we let G also satisfy $GC = I$, we have

$$\det C^+f'(x)B = \det GCC^+Df(x)B = \det G\, f'(x)B , \tag{2.2}$$

the last equality following from (2.1). Next, let (B_i,v_i) and C_i represent orientations of $T\sigma_x$ and $T\sigma_y$, respectively, for $i = 1,2$, where $(B_2,v_2) = (B_1,v_1)A$, $C_2 = C_1E(\det A, \det B > 0)$ and $[v_1] = [v_2] = (TW_x)$. Clearly,

$$\operatorname{sgn}\det C_2^+ F'(x)B_2 = \operatorname{sgn}\det C_1^+F'(x)B_2 ,$$

using $E^{-1}C_1^+C_2 = I$, $\det E^{-1} > 0$, and (2.2).

Note that since F is constant along W, $F'(x)v_1 = F'(x)v_2 = 0$. Hence,

$$(F'(x)B_1D_1,0) = F'(x)(B_1,v_1)D = F'(x)(B_2,v_2)$$

$$= (F'(x)B_2,0)$$

where

$$D = \begin{pmatrix} D_1 & d_2 \\ d_3 & d_4 \end{pmatrix} ,$$

and D_1 is $n \times n$. But $F'(x)B_1d_2 = 0$ implies $d_2 = 0$ ($F'(x)B_1$ has full column rank). $[v_1] = [v_2]$ implies $v_1 = \alpha v_2$ for $\alpha > 0$ so we have $d_4 > 0$. Hence, $\det D_1 > 0$ and $\operatorname{sgn}\det C_2^+F'(x)B_2$ $= \operatorname{sgn}\det C_1^+F'(x)B_1$ so the curve index is well defined at x.

Next we show that the curve index is constant for any $x \in W \cap \sigma^0$. We can choose the basis B as a continuous function of x, $B(x)$, along W because $B(x)$ is a choice of a linearly independent basis in $T\sigma^0_x$ and σ^0 is a smooth manifold. C^+ can be constant, hence $d(x) = \det C^+F'(x)B(x)$ is a continuous function of x. Since we have shown that the curve index is nonzero at some point $\bar{x} \in W \cap \sigma^0$. Suppose some point $\hat{x} \in W$ has opposite curve index to $\bar{x}$. Then

$$\det C^+F'(\bar{x})\, B(\bar{x}) \cdot \det C^+F'(\hat{x})B(\hat{x}) < 0.$$

Using a parametrization $\varphi(t)$ of W, there must be some $t^0 \in (\bar{t},\hat{t})$, where $\bar{x} = \varphi(\bar{t})$ and $\hat{x} = \varphi(\hat{t})$, such that

$$\det C^+F'(\varphi(t^0))\, B(\varphi(t^0)) = 0 .$$

C^+ has row rank n and $B(\varphi(t^0))$ has column rank n by definition, so $F'(\varphi(t^0))$ has rank less than n. This contradicts the fact that y is a regular value of F and $F(\varphi(t^0)) = y$.

If $\sigma \ni x$ is unique we have shown that the curve index is constant on $\omega \cap \sigma^0$. If x lies in more than one cell, then it lies in exactly two $(n+1)$-cells of $\mathcal{M}$, σ_1 and σ_2, because y is a good value for F w.r.t. $(M,\mathcal{M})$. Let $\tau = \sigma_1 \cap \sigma_2$ and choose an orientation $[B]$ of $T\tau_x$. Let $v_1 \in T(W \cap \sigma_1)_x$ and $v_2 \in T(W \cap \sigma_2)_x$ be chosen consistently with the orientation of (W,ω). Then, since 0 is a regular value for $F|_\tau$, (B,v_1) span $T\sigma_{1\,x}$, and B can be chosen so that $[B,v_1]$ is the orientation of σ_1. Due to the definition of

an oriented subdivided manifold, B has the opposite orientation to that induced on τ by σ_2. But, since v_2 is an "inward" vector for σ_2, it follows that the curve index is constant and well defined on oriented 1-complexes. ∎

The notion of curve index relates the behavior of the curve W to the Jacobian of map F which defines W. The next Theorem will be used repeatedly for the definition of algorithms and for proving theoretical results.

Suppose that $M \subseteq \mathbb{R}^{n+1}$, y is a good value for F w.r.t. $(M, \mathcal{M})$ and $\gamma \subseteq F^{-1}(y)$ is a path oriented so that its index is $i = \pm 1$. Let $(\dot{x}, \dot{\theta}) = \dot{\phi}(t)$ be the direction of γ at (x, θ), and let $F'(x) = (E, e)$ where $E \in \mathbb{R}^{n \times n}$.

Theorem 2.3. $\operatorname{sgn} \dot{\theta} = i \cdot \operatorname{sgn} \det E$.

Proof. $F'(x) \begin{pmatrix} \dot{x} \\ \dot{\theta} \end{pmatrix} = 0$, so if $\dot{\theta} = 0$, $E\dot{x} = 0$, and $\dot{x} \neq 0$ implies that $\det E = 0$.

If $\dot{\theta} > 0$, then, by our convention, M has the standard orientation. To represent this orientation, we use

$$(B, b) = \begin{pmatrix} I & \dot{x} \\ 0 & \dot{\theta} \end{pmatrix}$$

and $C = C^+ = I$ is the orientation for $N = \mathbb{R}^n$. Hence,

$$i = \operatorname{sgn} \det I(E, e)B = \operatorname{sgn} \det E ,$$

and we have $i \cdot \operatorname{sgn} \det E = 1 > 0$.

If $\dot{\theta} < 0$, then let

$$(B,b) = \begin{pmatrix} \tilde{I} & \bar{x} \\ 0 & \dot{\theta} \end{pmatrix} ,$$

where $\tilde{I} = (e^1, \ldots, e^{n-1}, -e^n)$ so that (B,b) represents the standard orientation in $\mathbb{R}^{n+1}$. We have

$$i = \operatorname{sgn} \det \tilde{I}(E,e)B = - \operatorname{sgn} \det(E,e)B = - \operatorname{sgn} \det E ,$$

and

$$i \cdot \det E < 0 . \qquad \blacksquare$$

Thus, we have the important property that the sign of $\dot{\theta}$ changes as the sign of the determinant of E. Clearly, a similar statement holds for any coordinate of (x,θ) and not just the last one.

The curve index is essentially arbitrary, so we must choose it in such a way that our algorithms will move in the desired direction along the path γ. We are always given an initial point $(x^0,\theta^0) \in \partial\sigma$ for some cell $\sigma \equiv \{x | b_i(x) \leq 0,\ i \in \underline{m}\}$ and we want to follow $F^{-1}(0)$ into the cell σ. Thus if we are (x,θ) in the null space of $F'(x^0,\theta^0)$ we know which sign to give $(\dot{x},\dot{\theta})$ so that it points into σ. Knowledge of $\dot{\theta}^0$ and $\operatorname{sgn} \det(\partial_1 F(x^0,\theta^0))$ along with Theorem 2.3 let us define i, the curve index in the appropriate way,

$$i = \operatorname{sgn} \dot{\theta} \cdot \operatorname{sgn} \det(\partial_1 F(x^0,\theta^0)) .$$

After i is determined at $(x^0, \theta^0) \in \partial M$, Theorem 2.3 can be used to determine the appropriate sign for $(\dot{x}, \dot{\theta})$.

Next we give explicit directions for implementing Theorem 2.3 in a path following algorithm. The numerically sophisticated reader may find the exposition above to be a sufficient description and may want to skip to the next section.

In this section to avoid singling out any component we consider a problem of following $F^{-1}(0)$ across a cell $\sigma \equiv \{x | b_i(x) \leq 0, \ i \in m\}$ where $\sigma \subset \mathbb{R}^{n+1}$ and F maps σ into $\mathbb{R}^n$. Assume that $x^0 \in F^{-1}(0) \cap \partial\sigma$ satisfies

$$b_1(x^0) = 0 , \qquad b_i(x^0) < 0 , \quad i \neq 1.$$

Let $F'_j(x)$ be the matrix $F'(x)._{\underline{n+1} \setminus \{j\}}$.

First we will discuss the problem of finding a nonzero element of the null space of $F'(x)$. We are assuming that 0 is a regular value for F, so that in some neighborhood of $F^{-1}(0)$, $F'(x)$ has rank n. Thus, some $n \times n$ submatrix of $F'(x)$ has full rank. To find a $v \neq 0$ such that $F'(x)v = 0$, first try to solve

$$F'_j(\bar{x})y = - F'(\bar{x})._j \tag{3.1}$$

for $j = n+1$. If $F'_{n+1}(\bar{x})$ is singular, let $j = j-1$ and again try to solve (3.1). Eventually, a j will be found such that $F'_j(\bar{x})$ is nonsingular. Then v defined by

$$v_i = \begin{cases} y_i , & \text{for } i < j \\ 1 , & \text{for } i = j \\ y_{i-1}, & \text{for } i > j \end{cases}$$

satisfies $F'(\bar{x})v = 0$. In the algorithm below the phrase, "find v such that $F'(\bar{x})v = 0$," means that a procedure such as the above is to be followed.

The easiest way to calculate sgn det $F'_j(\bar{x})$ is to keep track of certain things as the system (3.1) is being solved. A common (and efficient) method for solving a system of equations is to develop a LU decomposition of $F'_j(\bar{x})$ using Gaussian elimination with partial pivoting. In such an algorithm one can determine sgn det $F'_j(\bar{x})$ = sgn det LU by the following procedure

0. let $k = 1$
1. each time two rows of U are interchanged let $k = -k$
2. each time a negative diagonal element U_{ii} is calculated let $k = -k$.

The result k will be the $\operatorname{sgn}(\det F'_j(\bar{x}))$ because L is lower triangular with ones along the diagonal and U is upper triangular.

Thus, in the sequel when we say "compute sgn det E," we mean that a procedure similar to that above should be followed and no real work need be done.

We summarize this process in algorithmic form. We first show how the initial unit tangent to $F^{-1}(0)$ pointing into σ is found.

0. Given x^0 such that $x^0 \in \sigma = \{x | b_i(x) \leq 0,\ i \in m\}$ and $b_i(x^0) = 0$.

1. Find v such that $F'(x^0)v = 0$ by solving a system such as (3.1) and let $i' = \text{sgn}(\det F'_j(x^0))$.

2. Set $\delta = -\text{sgn}\langle \nabla b_1(x^0), v\rangle$.

3. Let $v = \delta \cdot \frac{v}{\|v\|}$,

4. Save $i = i' \cdot \delta$.

Now we have the curve index i and the initial tangent to $F^{-1}(0)$ for a curve following algorithm to move along. At some other point x^k near the path $F^{-1}(0)$ the procedure is as follows.

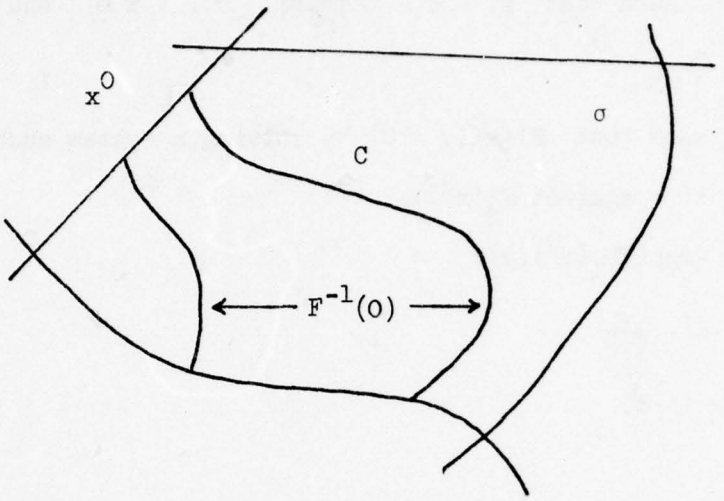
1. Find v such that $F'(x^k)v = 0$ and let $i' = \text{sgn}(\det F'_j(x^0))$.

2. Let $\delta = i \cdot i'$.

3. Let $v = \delta \cdot \frac{v}{\|v\|}$.

III.3. Proving Convergence for a Path Following Algorithm

In this section we define an algorithm which will follow a path across a compact cell $\sigma \subset \mathbb{R}^{m+1}$. Let C be the connected component of $F^{-1}(0)$ which contains a boundary point x^0 of $F^{-1}(0)$, where $F : \sigma \to \mathbb{R}^m$ is C^2. The algorithm will stay near the curve C and converge to the opposite boundary point of C, $x^* \in \partial\sigma$. A 2-dimensional example is given below.



Example 3.1.

Let the cell be defined as $\sigma \equiv \{x \in \mathbb{R}^{m+1} | b_i(x) \leq 0,\ i \in \underline{q}\}$. We assume that 0 is a good value for F (see Theorem II.3.15).

The algorithm is essentially a procedure for moving a fixed distance along a tangent to a curve near C. Then a hyperplane normal to the tangent is used to get a fully determined system of equations. Newton's method is used to solve this system approximately until we again have a point sufficiently close to C. This subroutine is repeated until the opposite boundary has been reached. At this point the constraint which is violated is augmented to F so that Newton's method can be used to determine the boundary point x^* to the desired tolerance.

A more explicit description of the algorithm is given below. The theoretical and computational question of interest is the determination of $\tau > 0$, the steplength along the tangential approximation, and $\epsilon > 0$ the termination criterion for the Newton subroutines.

Path-Following Algorithm 3.2.

0. The initial information consists of positive scalars τ, ϵ and the initial point $x_{0,0} \in F^{-1}(0) \cap \partial\sigma$. $k := 0$.

1. Determine $u(x_{k,0})$, the tangent to $F^{-1}(0)$ at $x_{k,0}$ so that $\|u(x_{k,0})\| = 1$ and the direction of $u(x_{k,0})$ is consistent with the direction of movement along $F^{-1}(0)$.

   Let $x_{k,1} = x_{k,0} + \tau u(x_{k,0})$.

2. Define $G_k: \mathbb{R}^{m+1} \to \mathbb{R}^{m+1}$ to be F augmented with the function defining the hyperplane normal to $u(x_{k,0})$ at $x_{k,1}$.

$$G_k(x) = \begin{pmatrix} F(x) \\ \langle u(x_{k,0}),\ x_{k,1} - x \rangle \end{pmatrix}$$

$$\ell := 1$$

3. Use Newton's method to solve $G_k(x) = 0$: Calculate

$$x_{k,\ell+1} = x_{k,\ell} - G_k'(x_{k,\ell})^{-1} G_k(x_{k,\ell}), \quad \ell = 1,2,3, \ldots$$

   until $\|G_k(x_{k,\ell+1})\| < \epsilon$.

4. If $x_{k,\ell+1} \in \sigma$, let $x_{k+1,0} := x_{k,\ell+1}$, $k := k+1$, and return to Step 1. Otherwise,

5. Let $\hat{i}$ be the index such that $b_{\hat{i}}(x_{k,\ell+1}) > 0$, $x_{k+1,0} := x_{k,\ell+1}$ $k := k+1$, and $\ell := 1$. Define

$$G_k(x) \equiv \begin{pmatrix} F(x) \\ b_{\hat{i}}(x) \end{pmatrix}$$

6. Use Newton's method to solve $G_k(x) = 0$: Iterate

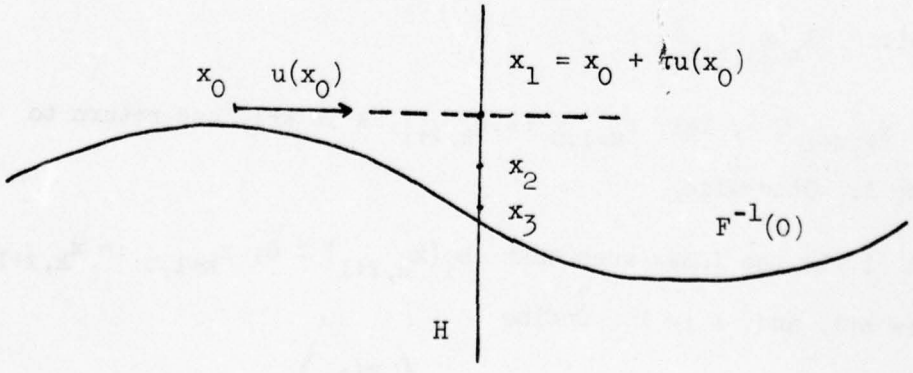
$$x_{k,\ell+1} = x_{k,\ell} - G_k'(x_{k,\ell})^{-1} G_k(x_{k\ell}) , \quad \ell = 1,2,\ldots ,$$

until $\|G(x_{k,\ell+1})\| < \epsilon$. Stop, $x_{k,\ell+1}$ is the approximate endpoint x^*.

We shall show that for some choice of positive scalars τ and ϵ, the Jacobians $G_k'(\cdot)$ are of full rank and after a finite number of Newton iterations, an approximate solution will be found.

Notice that it is assumed in Step 5 that there is only one constraint $\hat{i}$ which is violated by the first $x_{k,0}$ which is not in σ. We shall show that it is possible to choose τ and ϵ small enough so that this will be the case. However, in any practical implementation, strategies are available which should be used to designate a particular constraint if more than one constraint is violated by $x_{k,0}$. Such strategies were not included in the algorithm in order to make the convergence proof simpler.

Example 3.3. The subproblem in Steps 1-4: (the subscript k is dropped)



$$H \equiv \{x \mid \langle u(x_0),\ x_1 - x\rangle = 0\}$$

Since the algorithm uses Newton's method to calculate a point in the intersection of $F^{-1}(0) \cap H$ and $x_1 \in H$, it can be shown that x_ℓ, $\ell = 2,3, \ldots$ will all be in the hyperplane H.

We shall first prove that for any $x_0 \in B(C,\epsilon)$, $x_1 = x_0 + \tau u(x_0)$ is in a neighborhood of $y = F^{-1}(0) \cap H$ small enough so that the Newton sequence $x_1, x_2, x_3, \ldots$ converges to y. First, we require some definitions and preliminary results.

To help us determine when certain matrices are invertible we need the following lemmas whose proofs are in Ortega and Rheinboldt [1970, p. 45-46].

Lemma 3.4. (Perturbation Lemma). Let $A, C \in \mathbb{R}^{n \times n}$ and assume that A is invertible, with $\|A^{-1}\| \leq \alpha$ ($\|A\| = \sup_{\|x\|=1} \|Ax\|$). If $\|A-C\| \leq \beta$ and $\beta\alpha < 1$, then C is invertible, and

$$\|C^{-1}\| \leq \alpha/(1-\alpha\beta) .$$

Lemma 3.5. Suppose that the mapping $A: D \subset \mathbb{R}^m \to R^{n \times n}$ is continuous at a point $x^0 \in D$ for which $A(x^0)$ is invertible. Then there is a $\delta > 0$ and a $\gamma > 0$ so that $A(x)$ is invertible, and

$$\|A(x)\|^{-1} \leq \gamma, \quad \text{for any } x \in D \cap B(x^0,\delta) .$$

We also use the following mean value theorem:

Theorem 3.6. Let $F: D \subset \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable on a convex set $D_0 \subset D$ and suppose that for constants $\alpha \geq 0$ and $p \geq 0$, F' satisfies

$$\|F'(u) - F'(v)\| \leq \alpha \|u-v\|^p , \qquad \forall\, u,\ v \in D_0 .$$

Then for any $x,\ y \in D_0$

$$\|F(y) - F(x) - F'(x)(y-x)\| \leq [\alpha/(p+1)]\ \|y-x\|^{p+1} .$$

Proof. (Ortega-Rheinboldt [1970], p. 73).

The next result is contained in the proof of Kellogg, Li and Yorke's Theorem 2.1 [1976].

Lemma 3.7. There is an open set V containing $F^{-1}(0)$ and a continuous vector function $u(x) \neq 0,\ x \in V$, such that $F'(x)\, u(x) = 0,\ x \in V$.

For our purposes, since σ and C are compact, we can find a $\zeta > 0$ such that we can let $V \supset B(C,\zeta) \supset C$, and for $x \in B(C,\zeta)$, $u(x) \neq 0$ is defined and $F'(x)\, u(x) = 0,\ x \in B(C,\zeta)$

Definiton 3.8. Let $\mathbb{R}^n$ be the product space $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_p}$, where $n_1 + \cdots + n_p = n$, and denote the elements of $\mathbb{R}^n$ by $x = (x^1, \ldots, x^p)$ with $x^i \in \mathbb{R}^{n_i}$, $i \in \underline{p}$. Let $F = D \subset \mathbb{R}^n \to \mathbb{R}^m$ and, for given $x = (x^1, \ldots, x^p) \in D$, set

$$D_i = \{y \in \mathbb{R}^{n_i} | (x^1, \ldots, x^{i-1}, y, x^{i+1}, \ldots, x^p) \in D\},$$

and define $F_i = D_i \to \mathbb{R}^m$ by

$$F_i(y) = F(x^1, \ldots, y, \ldots, x^p), \quad y \in D_i .$$

Then F has a <u>partial F-derivative</u>

$$\partial_i F(x) \equiv F_i'(x^i)$$

at x with respect to $\mathbb{R}^{n_i}$ if $x^i \in \text{int } D_i$ and F_i has an F-derivative at x^i.

The implicit function theorem is given for the case of a continuous partial derivative since we usually assume that F is C^2.

<u>Implicit Function Theorem 3.9</u>. Suppose that $F: D \subset \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$ is continuous on an open neighborhood $D_0 \subset D$ of a point (x^0, y^0) for which $F(x^0, y^0) = 0$. Assume that $\partial_1 F$ is continuous and nonsingular at (x^0, y^0) and that it exists on D_0. Then there exist open neighborhoods $S_1 \subset \mathbb{R}^n$ and $S_2 \subset \mathbb{R}^p$ of x^0 and y^0, respectively, such that, for any $y \in \bar{S}_2$, the equation $F(x,y) = 0$ has a unique solution $x = H(y) \in \bar{S}$, and the mapping $H = S_2 \to \mathbb{R}^n$ is continuous. Moreover, if $\partial_2 F$ exists at (x^0, y^0), then H is F-differentiable at y^0 and

$$H'(y^0) = -[\partial_1 F(x^0, y^0)]^{-1} \partial_2 F(x^0, y^0).$$

<u>Proof</u>. (Ortega and Rheinboldt [1970], pp. 128-129.)

The following lemma is concerned with the existence of a solution for each of the subproblems in Steps 1-4.

Lemma 3.10. Consider the function $h:\sigma \times [0,\tau] \to \mathbb{R}^n$ such that if

$$G(v,x,t) = \begin{pmatrix} F(v) \\ \langle u(x), (x + tu(x)) - v \rangle \end{pmatrix}$$

then $h(x,t)$ satisfies $G(h(x,t),x,t) = 0$, for $t \in [0,\tau]$, $x \in B(c,\epsilon)$.

There exists scalars $\tau' > 0$ and $\epsilon' > 0$ such that h exists and is continuous and $\partial_1 G(h(x,t),x,t)$ is invertible for $x \in B(C,\epsilon')$ and $t \in [0,\tau']$.

Proof. Consider the point $(x^0, x^0, 0)$, $x^0 \in C$.

$$\partial_1 G(x^0,x^0, 0) = \begin{pmatrix} F'(x^0) \\ u(x^0) \end{pmatrix} .$$

Since $u(x)$ is continuous, and F is C^2, $\partial_1 G$ is continuous at $(x^0, x^0, 0)$ and it is nonsingular because $F'(x^0)y = 0$ implies $y = \alpha u(x^0)$ for some $\alpha \in \mathbb{R}$. The implicit function theorem 5.8, then gives us that the equation $G(v,x,t)$ has a unique solution $v = h(x,t)$ for (x,t) in a neighborhood $S^1 \times T^1$ of $(x^0, 0)$ and $h:S \times I \to \mathbb{R}^n$ is continuous. We also know that since $h(x^0,0) = x^0$, $\partial_1 G$ is continuous, and $\partial_1 G(x^0, x^0, 0)$ is nonsingular that there is a neighborhood $S^2 \times I^2$ of $(x^0, 0)$ such that $\partial_1 G(h(x,t),x,t)$ is nonsingular for

$(x,t) \in S^2 \times I^2$. Let $S \times I = (S^1 \times I^1) \cap (S^2 \times I^2)$.

Since $x^0 \in C$ was arbitrary, we can index $S \times I$ by $x \in C$ and conclude that

$$\bigcup_{x \in C} (S \times I)_x \supset C \times \{0\}$$

is an open cover for $C \times \{0\}$. Since $C \times \{0\}$ is compact in $\mathbb{R}^n \times \mathbb{R}$. there is a finite subcover $K = \{(S \times I)_i \mid i = 1,\ldots,p\}$ of $C \times \{0\}$. Let $d = \inf\{d((x,0),(y,s)) \mid x \in C,\ (y,s) \notin K\}$ then d is positive because otherwise by the continuity of d, there is some $(\bar{x},0) \in C \times \{0\}$ which is a cluster point for a sequence of points which are not in K, but $(\bar{x},0)$ is in $(S \times I)_i$ for some $i \in \underline{p}$ and we have a contradiction.

Since d is positive, we can choose an $\epsilon' > 0$ and $\tau' > 0$ such that $h(x,t)$ is continuous on $B(C,\epsilon') \times [-\tau',\tau']$ and $\partial_1 G(h(x,t),x,t)$ is nonsingular for $(x,t) \in B(C,\epsilon') \times [-\tau,\tau]$. ∎

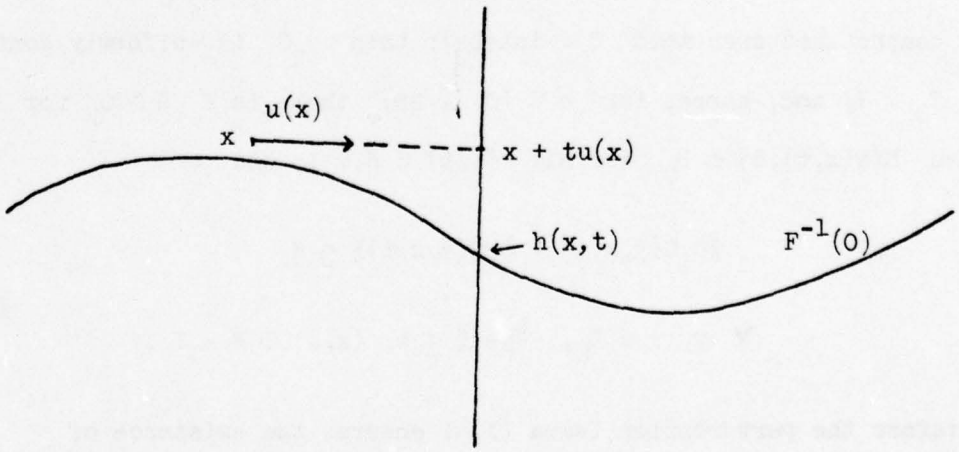


FIGURE 3.11. A Depiction of $h(x,t)$

Next we prove the main result of this section, which allows us to move along the curve any finite distance.

Theorem 3.12: ϵ^2 and $\tau^2 > 0$ can be chosen so that in Algorithm 1, after a finite number of repetitions of Step 3, there is a k such that $x_{k,0} \notin \sigma$.

Proof. By Lemma 3.10, $\partial_1 G(h(x,t),x,t)$ is continuous and nonsingular for all $(x,t) \in B(c,\epsilon') \times [0,\tau']$ $(\cong B \times I)$. By Lemma 3.5 $\partial_1 G(h(x,t),x,t)$ is continuous for $(x,t) \in B \times I$, and, since $B \times I$ is compact, there is a $\beta < +\infty$ such that

$$\|\partial_1 G(h(x,t),x,t)\| \leq \beta \qquad \forall\, (x,t) \in B \times I\ .$$

Let D be an open set containing σ such that the extension of F to D, $\overline{F}$, is C^2 and 0 is a regular value for $\overline{F}$. Let $D_0 \subset D$ be any compact set such that $C \subset \mathrm{int}(D_0)$; then $\partial_1 G$ is uniformly continuous on $D_0 \times I$, and, hence, for $e \in (0, 1/2\beta)$ there is a $\delta > 0$ for which $B(h(x,t),\delta) \subset D_0$ for all $(x,t) \in B \times I$ and

$$\|\partial_1 G(y,x,t) - \partial_1 G(z,x,t)\| \leq e$$

$$\forall\ y,\ z \in D_0,\ \ \|y-z\| \leq \delta,\ (x,\epsilon) \in B \times I\ . \tag{3.2}$$

Therefore the perturbation lemma (3.4) ensures the existence of $\partial_1 G(y,x,t)$ for each $(x,t) \in B \times I$ and $y \in B(h(x,t),\delta)$. Moreover, we have

$$\|\partial_1 G(y,x,\epsilon)\| \leq \beta/[1-\beta e]$$

$$\forall\ y \in B(h(x,t),\delta),\ (x,t) \in B \times I \tag{3.3}$$

For any fixed $(x,t) \in B \times I$, consider, now, the Newton process

$$y^{k+1} = y^k - \partial_1 G(y^k,x,t)^{-1}\, G(y^k,x,t), \quad k = 0,1,\ldots \tag{3.4}$$

with $y^0 \in B(h(x,t),\delta)$.

Then we will show that

$$\|y^k - h(x,t)\| \leq \alpha^k \delta, \qquad k = 0,1,\ldots \tag{3.5}$$

where $\alpha = \beta e/(1 - \beta e) < 1$.

(3.5) is true by assumption for $k = 0$, and if it holds for some $k \geq 0$, then $y^k \in B(h(x,t),\delta)$, and, hence, by (3.4), (3.2), (3.3) and Theorem 3.6

$$\begin{aligned}
\|y^{k+1} - h(x,t)\| &= \|y^k - h(x,t) - \partial_1 G(y^k,x,t)^{-1}\, G(y^k,x,t)\| \\
&\leq \|\partial_1 G(y^k,x,t)\| \cdot \|G(h(x,t),x,t) - G(y^k,x,t) \\
&\qquad - \partial_1 G(y^k,x,t)(h(x,t) - y^k)\| \\
&\leq [\beta/(1-\beta e)]e\, \|h(x,t) - y^k\| \\
&\leq \alpha^{k+1}\, \delta\ .
\end{aligned} \tag{3.6}$$

(Note that Theorem 3.6 was applied with $\alpha = e$ and $p = 0$.)

Therefore the bound (3.5) is demonstrated and the Newton sequence (3.4) remains in $B(h(x,t),\delta)$ and converges to $h(x,t)$.

It remains to specify an ϵ^2 and a τ^2 positive so that for any $x \in B(c,\epsilon)$, $x + \tau u(x) = \bar{x} \in B(h(x,\tau),\delta)$. Choose $\epsilon \in (0,\epsilon^1)$ so that for any $x \in B(C,\epsilon^2)$, $d(x,h(x,0)) < \delta$. This is clearly possible because for $x \in C$, $d(x,h(x,0)) = 0$ and C is a compact set and both d and h are continuous functions.

Next we must specify τ. Consider the following family of sets parametrized by t :

$$A_t = \{x \in B^0(C,\epsilon^1) \mid D(x,s) < \delta,\ 0 \leq s \leq t\}\ , \qquad t \in (0,\tau')$$

where $D(x,t) = d(x + tu(x),\ h(x,t))$ is well-defined and continuous (Lemma 3.10 shows that $h(x,t)$ is well-defined and continuous) for $x \in B^0(C,\epsilon^1)$, $t \in (0,\tau')$. A_t is an open set by the continuity of $D(\cdot,t)$. For any $x' \in B(C,\epsilon^2)$, $D(x,0) < \delta$ by the definition of ϵ. So by the continuity of $D(x,\cdot)$ there is an interval $(0,t')$ such that for $s \in (0,t')$, $D(x,s) < \delta$. Thus $\cup_{t\in(0,\tau')} A_t \supset B(C,\epsilon^2)$. By compactness there is a finite subcover $\cup_{i=1}^{k} A_{t_i}$ of $B(C,\epsilon^2)$. Let $\tau = \min\{t_i \mid i \in \underline{k}\}$, then $A_\tau \supset B(C,\epsilon^2)$ because $A_\tau \supset A_{t_i}$ for any $t_i > \tau$ by the definition of A_{t_i}.

Now all that is left to do is show that we can choose a termination tolerance $\epsilon^3 > 0$ so that $\|F(x)\| \leq \epsilon^3$ and $x \in B(C,\epsilon^1)$ implies that $d(x,C) \leq \epsilon^2$ or $x \in B(C,\epsilon^2)$.

Define the function $\gamma: \mathbb{R}^1_+ \to \mathbb{R}^1_+$ as

$$\gamma(\alpha) = \max\{d(x,C) \mid x \in B(C,\epsilon^1) \text{ and } \|F(x)\| \leq \alpha\} \tag{3.7}$$

γ is well defined because $d(\cdot,C)$ is continuous and the constraint set is compact. Clearly $\varphi(0) = 0$ because $\|F(x)\| = 0$ and $x \in B(C,\epsilon^1)$ means that $x \in C$ because ϵ^1 was chosen so that $B(C,\epsilon^1) \cap F^{-1}(0) = C$. If we can show that γ is right continuous, then we can choose an $\epsilon > 0$ so that $\gamma(\epsilon^3) \leq \epsilon^2$ and we will be done.

Let $\{\alpha^k\}$ be a sequence such that $\lim_k \alpha^k = \bar{\alpha}$ and $\alpha^k > \alpha^{k+1}$. Let

$$x^k = \text{argmax}\{d(x,C) \,|\, x \in B(C,\epsilon^1),\ \|F(x)\| \leq \alpha^k\}\ ,\quad k = 1,2,\ \ldots\ . \tag{3.8}$$

Then since $x^k \in B(C,\epsilon^1)$, a compact set, there is some subsequence $\{x^{k_\ell}\}$ which converges to x. Let $\bar{x}$ solve (3.7) for $\bar{\alpha}$, then since $\alpha^{k_\ell} > \bar{\alpha}$ for all ℓ,

$$d(x^{k_\ell},C) \geq d(\bar{x},C)\ ,\qquad \ell = 1,2,\ \ldots\ ,$$

and

$$d(x',C) \geq d(\bar{x},C)$$

by the continuity of $d(\cdot,C)$. But

$$\|F(x')\| \leq \bar{\alpha} \quad \text{and} \quad x' \in B(C,\epsilon^1)$$

so

$$d(x',C) \leq d(\bar{x},C)$$

and

$$d(x',C) = d(\bar{x},C)\ . \tag{3.9}$$

Since $\{\gamma(\alpha^{k_\ell})\}$ is a subsequence of $\{\gamma(\alpha^k)\}$ and $\gamma(\alpha^k) \geq \varphi(\alpha^{k+1})$ by (3.7), we have

$$\lim_k \varphi(\alpha^k) = \lim_\ell \varphi(\alpha^{k_\ell})$$
$$= \lim_\ell d(x^{k_\ell}, C)$$
$$= d(x', C)$$
$$= d(\bar{x}, C)$$
$$= \varphi(\bar{\alpha}) \ .$$

γ is right continuous at zero, in particular, so we can choose $\epsilon^3 > 0$ so that $\gamma(\epsilon^3) \leq \epsilon^2$.

Now we must show that we really are making progress along the curve. Since C is compact we can parametrize it by path length. If $C = \{x | x = \varphi(t),\ t \in [0,T],\ \|\dot{\varphi}(t)\| = 1\}$. Then we say that C has length T.

We can measure our movement along C during the k^{th} subproblem by $d(h(x_{k-1,0}, \tau),\ h(x_{k,0}, \tau))$. For simplicity let $y_{k+1} \equiv h(x_{k,0}, \tau)$ and $x_k \equiv x_{k,0}$.

Since $\|u(x_k)\| = 1$, $d(x_k, \bar{x}_k) = \tau$, where $\bar{x}_k \equiv x_k + \tau u(x_{k-1})$. Since $y_{k+1} \in \{x | \langle u(x_k),\ \bar{x}_k - x\rangle = 0\} \equiv H(\bar{x})$ and $\bar{x}_k$ is the solution of $\min_{x \in H(\bar{x}_k)} d(x_k, x)$, we have $d(x_k,\ y_{k+1}) \geq \tau$. Also, by the termination criterion,

$$d(x_k, y_k) \leq \epsilon \ ,$$

so

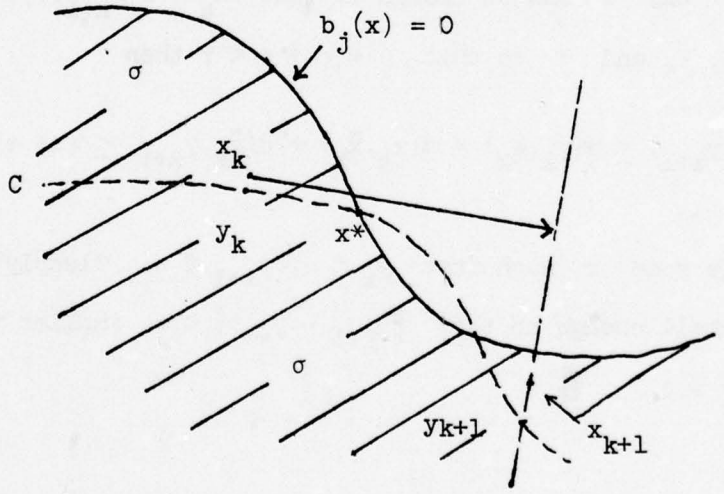
$$d(y_k, y_{k+1}) \geq d(x_k, y_{k+1}) - d(x_k, y_k)$$
$$\geq \tau - \epsilon \ .$$

So choose $\epsilon = \min(\epsilon^3, \tau/2)$ and $d(y_k, y_{k+1}) \geq \tau/2$. If $y_0 = x^0$, then for an integer $K > T \cdot 2/\tau$, $d(x^0, y_{k+1}) \geq T$, implying that $y_k \notin C$. By definition $y_k \in F^{-1}(0)$. In order to conclude that there is some k such that $x_k \in \sigma$ and $x_{k+1} \notin \sigma$ we must guarantee that there is some $y_{k+1} \notin \sigma$.

Suppose that $A(x') = j$, then since 0 is a good value

$$\begin{pmatrix} F'(x') \\ b_j(x') \end{pmatrix}$$

has rank $n+1$. We are concerned with avoiding the following situation.



Thus if $\varphi(T) = x^*$, then $b_j(\varphi(t))$ is a C^1 function from $\mathbb{R}^1$ into $\mathbb{R}^1$ and $\frac{d}{dt} b_j(\varphi(t))\Big|_{t=T} > 0$ by the nonsingularity of

$$\begin{pmatrix} F'(x) \\ b_j(x) \end{pmatrix}$$

at x' and the fact that $b_j(\varphi(T')) < 0$. Thus there is an interval $(T, T + \gamma)$ over which $b_j(\varphi(t)) > 0$ and hence $\varphi(t) \notin \sigma$ for $t \in (T, T + \gamma)$.

To insure that there is some $y_k \in \{y | y = \varphi(t),\ t \in (T, T + \gamma)\}$ we first need to get an upper bound on $d(y_k, y_{k+1})$. Remember $y_{k+1} = h(x_k, \tau)$ and δ is the radius of the ball around y_k which $\bar{x}_k = x_k + \tau u(x_k)$ must be in. We have shown that no matter how small $\delta > 0$ is, ϵ and τ can be chosen so that $\bar{x}_k \in B(y_{k+1}, \delta)$. So if we choose δ, ϵ, and τ so that $\delta + \epsilon + \tau < \gamma$ then

$$d(y_k, y_{k+1}) \leq d(y_k, x_k) + d(x_k, \bar{x}_k) + d(\bar{x}_k, y_{k+1}) \leq \epsilon + \tau + \delta < \gamma.$$

Then there is some k such that $y_k \in \sigma$, $y_{k+1} \notin \sigma$. Clearly we can choose ϵ small enough so that $\|x_{k+1} - y_{k+1}\| < \epsilon$ implies that $x_k \in \sigma$, $x_{k+1} \notin \sigma$. ∎

Next we show that termination will occur after a finite number of repetitions of Step 6.

Theorem 3.13. ϵ, τ, and $\delta > 0$ can be chosen so that the path following algorithm will compute an approximate solution to x^* after a finite number of steps.

Proof. Let B^i be the facet of σ containing x^*. Let $\gamma > 0$ be the radius of the domain of attraction for x^* using Newton's method on

$$G_k(x) = \begin{pmatrix} F(x) \\ b_i(x) \end{pmatrix} .$$

Let $\bar{\gamma} < \gamma$ be the radius of a ball around x' such that $C \cap B(x^*,\bar{\gamma})$ is connected, and if $\hat{x} = C \cap \partial B(x^*,\bar{\gamma})$ then $d(C \setminus B(\hat{x}',\bar{\gamma}), B^i) = d(x,B^i)$, that is, the minimum distance from the curve outside of the $\bar{\gamma}$-neighborhood of x^* is the facet B^i is attained at the intersection of the curve C and the boundary of the ball. Such a $\bar{\gamma}$ exists by the continuity compactness of C and B^i. Let $\theta = d(x,B^i)$.

We want to show that the first point in the sequence $\{x^k\}$ which is not in σ, call it $x^{\ell+1}$, is such that $x^{\ell+1} \in B(x^*,\gamma)$, and hence Newton's method in Step 6 will converge to x^*.

So $x^\ell \in \sigma$, and $x^{\ell+1} \notin \sigma$. Now $x^\ell \in B(y^\ell,\delta)$ and $x^{\ell+1} \in B(y^{\ell+1},\delta)$ but there are three cases to consider depending on whether y^ℓ and $y^{\ell+1}$ are within or outside of σ. We will choose $\epsilon, \tau, \delta > 0$ so that $\bar{\gamma} + \epsilon + \tau + \delta \leq \gamma$ and $\epsilon < \gamma$.

Case 1. $y^{\ell+1} \in \sigma$. Then $d(y^{\ell+1},x^{\ell+1}) < \epsilon$ means that $d(y^{\ell+1},x^*) \leq \bar{\gamma}$. So, we conclude

$$d(x^{\ell+1},x^*) \leq \epsilon + \bar{\gamma} < \gamma .$$

Case 2. $y^{\ell+1} \notin \sigma$, $y^{\ell} \in \sigma$.

By part of the proof of Theorem 3.12 (pg. 12) for any $t \in [0,\tau]$,

$$d(x^{\ell} + tu(x^{\ell}),\ h(x^{\ell},t)) \leq \delta\ .$$

Now since $h(x^{\ell},0) = y^{\ell} \in \sigma$ and $h(x^{\ell},\tau) = y^{\ell+1} \notin \sigma$, an intermediate value argument along the curve C would show that there is some $s \in [0,\tau]$ such that $h(x^{\ell},s) = x^*$. (See Figure 3.14.)

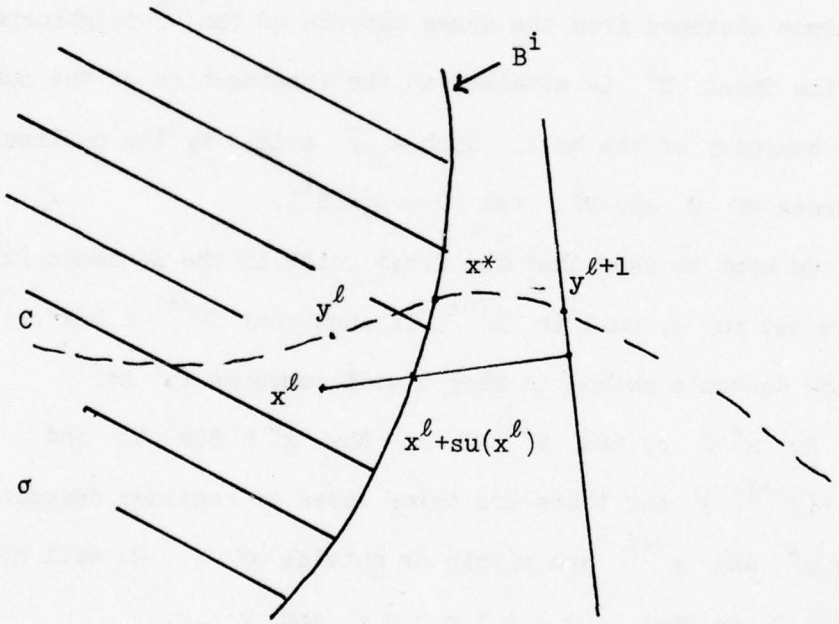


FIGURE 3.14

Let $x^s = x^{\ell} + su(x^{\ell})$. Then $d(x^s,x^*) \leq \delta$ and $d(x^{\ell+1},x^s) \leq \tau$. By the triangle inequality

$$d(x^{\ell+1},x^*) \leq \delta + \tau < \gamma\ .$$

Case 3. $y^{\ell} \notin \sigma$. This implies that

$$x^{\ell} \in B(x^*, \bar{\gamma})$$

so

$$d(x^{\ell+1}, x^*) \leq \bar{\gamma} + d(x^{\ell}, x^{\ell+1})$$

$$\leq \bar{\gamma} + \tau + \delta < \gamma.$$

Hence, in all cases $d(x^{\ell+1}, x^*) < \gamma$ and hence $x^{\ell+1}$ is in the domain of attraction of x^* for the iterative process in Step 6. ∎

From the large number of times we used worlds like "if ϵ is chosen small enough" one may get the impression that the tolerances must be chosen so small that the algorithm would be grossly inefficient. In fact, from our computational experience, quite large stepsizes can be chosen and the algorithm still converges.

CHAPTER IV

PATH-FOLLOWING METHODS

In the first two sections of this chapter we prove some theorems concerning the existence of paths which are closely related to the paths of Kellogg, Li, and Yorke [1976] and S. Smale [1976]. In the remainder of the chapter, two new path methods are discussed which are closely related to the paths defined by certain fixed point algorithms.

II.1. Kellogg, Li, and Yorke's Continuation Method.

Let σ be a bounded, convex cell in $\mathbb{R}^n$ and $f : \sigma \to \mathbb{R}^n$ a C^2 function which satisfies

Assumption 1.1. $f(x)$ points into σ for all points x in the boundary of σ.

Define the solution set $E = \{x \in \sigma \mid f(x) = 0\}$. Let $h : \sigma \setminus E \to \partial\sigma$ be denoted as

$$h(x) = x - \mu(x) \cdot f(x) ,$$

where

$$\mu(x) = \{\mu \geq 0 \mid x - \mu f(x) \in \partial\sigma\} \quad . \tag{1.1}$$

Note that $\mu(x)$ is well defined because $f(x) \neq 0$ for $x \notin E$, σ is convex, and the boundary condition (Assumption 1.1) implies that for $x \in \partial\sigma$, $\mu(x) = 0$. Hence, h is the identity on the boundary of σ. The function h is often called a retraction of the set $\sigma \setminus E$

to the boundary of σ. This method is concerned with the path $h^{-1}(x^0)$ where x^0 is some point on the boundary of σ. This path method was motivated by Hirsh's [1963] proof that there is no continuous retraction of a set to its boundary.

Intuitively, $h^{-1}(x^0)$ is the set of points x for which $-f(x)$ points at x^0. Figure 1.2 depicts this behavior.

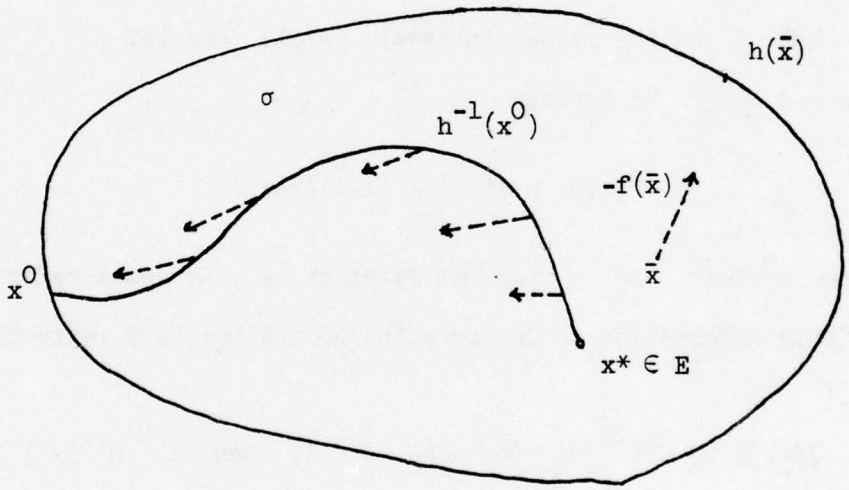


FIGURE 1.2

The main result of Kellogg, Li, and Yorke [1976] can be stated as

Theorem 1.3. For almost every $x^0 \in \partial\sigma$ the set $h^{-1}(x^0)$ consists of a number of components, each one of which is a diffeomorphic image of a circle or interval. Furthermore, for any $\epsilon > 0$, the connected

component $\gamma \subset h^{-1}(x^0)$ leading from x^0 is such that

$$\gamma \cap B(E,\epsilon) \neq \varphi .$$

In other words, γ gets arbitrarily close to some zero of f. The proof of Kellogg, Li, and Yorke used only results from advanced calculus, and, hence. it was quite long. We shall prove a very similar result using the theory of Chapter II. The new result will be just as useful for computational purposes.

Let I be the closed interval $[-1,1]$ and let $F:M = \sigma \times I \to \mathbb{R}^n$ be defined

$$F(x,\theta) = \theta f(x) - (1-\theta)(x-x^0) \tag{1.2}$$

for some $x^0 \in \tau^0$ and τ is some facet of σ. We shall refer to the use of this deformation to define paths as the homotopy retraction method.

If $(x,\theta) \in F^{-1}(0)$ for some $\theta < 1$, then $x \in h^{-1}(x^0)$ because

$$\theta f(x) - (1-\theta)(x-x^0) = 0$$

implies

$$x - \frac{\theta}{1-\theta} f(x) = x^0 ,$$

so

$$\mu(x) = \frac{\theta}{1-\theta} \qquad \text{and} \qquad h(x) = x^0 .$$

A nice feature of our deformation is that $F(x,\theta)$ is defined for $x \in E$ while $h(x)$ is not. Clearly $(x,1) \in F^{-1}(0)$ implies that

$f(x) = 0$ or $x \in E$. Next, we have the result which is essentially equivalent to Theorem 1.3.

Theorem 1.4. If Assumption 1.1 holds and 0 is a good value of $F: M \to \mathbb{R}^n$, then there is a unique point $(x^0, \theta^0) \in \partial M \cap F^{-1}(0)$ and $\theta^0 = 0$. The component $\gamma \subset F^{-1}(0)$ containing $(x^0, 0)$ has an opposite boundary point $(x^*, 1)$. Also, $x^* \in E$ and $|E|$ is odd.

Proof. Since f is C^2, and 0 is a good value, Theorem I.3.15 gives us that $F^{-1}(0)$ is a 1-manifold with boundary, neat in M.

The reason that $I = [-1, 1]$ was chosen for the domain of θ rather than $I' = [0, 1]$ is that $[x^0, 0]$ would be in two facets of $\sigma \times I'$ which would make it impossible for 0 to be a good value.

Consider solutions to

$$\begin{aligned} F(x, \theta) &= \theta f(x) - (1-\theta)(x - x^0) = 0 \\ \theta &< 1 \\ (x, \theta) &\in \partial M \end{aligned} \tag{1.3}$$

It is clear that $F^{-1}(0) \cap (\sigma \times \{0\}) = (x^0, 0)$. Parametrize γ by $\varphi = [0, T] \to M$, such that $\varphi(0) = (x^0, 0)$, $\varphi(T) \in \partial M$, and $\varphi(t) \in M^0$, $t \in (0, T)$. If we can write $\varphi(t) \equiv (x(t), \theta(t))$, then $\dot{\theta}(0) > 0$ would imply $\theta(\bar{t}) > 0$, $\bar{t} \in (0, T]$ because if not, then $\theta(0) = 0$, $\theta(\bar{t}) = 0$ implies that γ is not neat in M, a contradiction.

The orientation of $\varphi(t)$ is determined by guaranteeing that $\dot{x}(0)$ point into σ. But $(\dot{x}(0), \dot{\theta}(0)) = \alpha \cdot (v, 1)$ where α is chosen so that αv points into σ, and

$$F'(x,\theta)\Big|_{(x^0,\theta)} \begin{pmatrix} v \\ 1 \end{pmatrix} = 0$$

or

$$[\theta \cdot f'(x) - (1-\theta)I|f(x) + (x-x^0)]\Big|_{(x^0,0)} \begin{pmatrix} v \\ 1 \end{pmatrix} = 0 ,$$

$$-Iv = -f(x^0) ,$$

$$v = f(x^0) .$$

Since $f(x^0)$ points into σ, $\alpha > 0$, and, hence, $\dot{\theta}(0) > 0$, and $\theta \geq 0$ for $(x,\theta) \in \gamma$.

Suppose (x',θ') satisfies (1.2) for $\theta \in (0,1)$. Then $x' \in \partial\sigma$, and $f(x') \neq 0$, and

$$x' - \frac{\theta'}{1-\theta'} f(x') = x^0 \tag{1.4}$$

$f(x') \neq 0$ implies $x' \neq x^0$. Now $x' \in \partial\sigma$, $f(x')$ points into σ, and $x^0 \in \partial\sigma$ means that (1.4) is a contradiction of the convexity of σ. Hence, there is a unique solution to (1.3) when $\theta \geq 0$.

We have shown that $\partial\gamma \setminus \{(x^0,0)\} \subset \{(x,\theta)|(x,\theta) \in \partial M,\ \theta > 0,\ x \notin \partial\sigma\}$ which means that $\partial\gamma \setminus \{(x^0,0)\} = (x^*,1)$ and $f(x^*) = 0$.

Suppose there is another point $x' \in E$, then $(x',1) \in F^{-1}(0) \cap \partial M$. Suppose that $\gamma' \subset F^{-1}(0)$ is the connected component containing $(x',1)$. If we can show that for any $(x,\theta) \in \gamma'$, $\theta > 0$, then, by the arguments above. $\partial\gamma' \setminus \{(x',1)\} = \{(x^2,1)\}$ and $f(x^2) = 0$. Suppose there is some $(x,\theta) \in \gamma'$ with $\theta \leq 0$, then, by continuity, there is some

$(\bar{x}, \bar{\theta}) \in \gamma'$ with $\bar{\theta} = 0$, which implies $\bar{x} = x^0$. But $\gamma' \neq \gamma$ and $F^{-1}(0)$ is a neat submanifold, a contradiction.

Thus we have shown that each solution $x' \neq x^*$ is connected by some component (Fig. 1.4) of $F^{-1}(0)$ to one other solution. Since 0 is a good value and M is compact, there are a finite number of elements in E and the theorem is proved. ∎

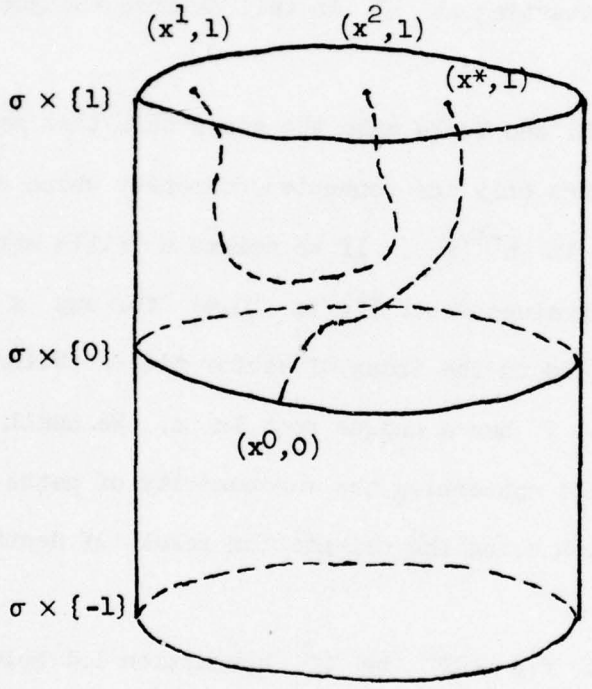


FIGURE 1.4

Kellogg, Li, and Yorke [1976, p. 478] explore some conditions which imply that x^0 will be guaranteed to be a good value for k. One of these conditions was the following

Assumption 1.5. (Eigenvalue condition) Suppose that the matrix $f'(x)$ has no eigenvalues which lie on $(0,\infty)$ for any $x \in \sigma \setminus E$.

The result which follows immediately is

Theorem 1.6. Let $f:\sigma \to \mathbb{R}^n$ and suppose assumptions 1.1 and 1.5 are true. Then each x^0 on a C^1 part of $\partial\sigma$ is a regular value of h, and the curve starting at x^0 is well defined and goes to the set E.

Kellogg, Li and Yorke make the conjecture that assumption 1.5 implies that E has only one connected component which can be joined to $\partial\sigma$ by a path in $h^{-1}(x^0)$. If we assume a little more, i.e., that $f'(x)$ has no eigenvalues which lie in $[0,\infty)$ for any x in σ, then using Hopf's theorem on the index of vector fields [Milnor, 1965], one can show that f has a unique root in σ. We shall prove this result and a result concerning the monotonicity of paths with this stronger assumption using the orientation result of Section III.2.

Theorem 1.7. Let $f:\sigma \to \mathbb{R}^n$ be C^2, assumption 1.1 hold, and for every $x \in \sigma$, $f'(x)$ has no eigenvalue which lies on $[0,\infty)$. Then $f(x) = 0$, $x \in \sigma$ has a unique solution, and for any $x^0 \in \partial\sigma$ defining F, there is one component in $F^{-1}(0)$. Furthermore, if $F^{-1}(0) = \{(x,\theta) \mid (x,\theta) = (x(t),\theta(t)),\ t \in [0,T]\}$, where $(x(0),\theta(0)) = (x^0,0)$ and $(x(T),\theta(T)) = (x^*,1)$, then $\dot{\theta}(t) > 0$ for any $t \in [0,T]$.

Proof. We will prove the last statement first. From the proof of Theorem 1.4 we have that $\dot{\theta}(0) > 0$ and that $\theta(t) > 0$ for any $t > 0$. Thus, we only need to show that $\dot{\theta}(t) > 0$ for $t \in [0,T]$.

Recall that

$$F'(x,\theta) = [\theta f'(x) - (1-\theta)I | f(x) + (x-x^0)]$$

$$= [E(x) | e(x)].$$

Any eigenvalue of $E(x)$ has the form $\theta\lambda - (1-\theta) < 0$ where $\lambda < 0$ is an eigenvalue of $f'(x)$. Hence, $E(x)$ is never singular for any $x \in \sigma$. This means that $\dot{\theta}(t) \neq 0$ for any $t \in (0,T]$ by Theorem III.2.3. Hence, by the continuity of $\dot{\theta}(t)$, $\dot{\theta}(t) > 0$ for any $t \in [0,T]$.

Suppose there was another component $\bar{\gamma} \subset F^{-1}(0)$, then by studying the proof of Theorem 1.4, we see that $\bar{\gamma}$ would have boundary points at $(x^1,1)$, $(x^2,1)$ where x^1 and x^2 are in E. Setting up a parametrization of $\bar{\gamma}$, $(\bar{x}(t),\bar{\theta}(t)) = [0,T] \to M$ it is clear that for some $t \in (0,T)$, $d\bar{\theta}(t)/dt = 0$ which implies that $\det E(x(t)) = 0$, a contradiction. Hence, $F^{-1}(0)$ has a unique connected component. Similar considerations would show that x^* is the only solution of $f(x) = 0$, $x \in \sigma$. ∎

Thus we have shown that Assumption 1.5 is so strong that the homotopy parameter θ is guaranteed to increase monotonically from zero to one. The literature of continuation methods for solving equations commonly makes assumptions necessary to insure that the homotopy parameter increases monotonically along the path (Avila [1974]).

However, it is well known that monotonic behavior by any variable can rarely be depended upon in equation solving problems (Eaves and Scarf, [1976]). The path following algorithm of Section III.3 does not assume monotonicity in any variable. The orientation results of Section III.2 are what allow us to follow paths which are not monotone in any variable.

IV.2. The Global Newton Method

In this section, we discuss a path method that satisfies the differential equation version of the Newton recursion $x^{k+1} - x^k = f'(x^k)^{-1} f(x^k)$, $k = 0,1,\ldots$, except for a factor of ± 1 which is determined by the determinant of $f'(x)$. To be precise, if $x(t)$ was a parametrization of the path, then $x(t)$ is specified by the differential equation

$$f'(x) \frac{dx}{dt} = -\lambda(x) f(x), \qquad x \in D \tag{2.0}$$

where λ is an arbitrary scalar function of x such that $\operatorname{sign} \lambda(x) = \pm \operatorname{sign} \det Df(x)$.

This method was first described in a paper by Smale [1976]. The global Newton method is a differentiable analogue of Scarf's fixed point algorithm [1965] and Eaves' "vector-labelling" algorithm [1969]. Varian [1977] described a method which allows the path to move in and out of D. In the next section we will present a generalization of Varian's method and prove that the path produced by this algorithm is

identical to the path produced by Eaves' algorithm in the limit as the mesh of the triangulation goes to zero.

Once again, we are to find $x \in D \subset \mathbb{R}^n$ such that $f(x) = 0$, where D is a bounded cell with one facet ∂D. Suppose that f is C^2 and the following boundary condition holds:

Assumption 2.1. For $x \in \partial D$, $\det f'(x) \neq 0$ and there is a choice, (a) $\operatorname{sign} \lambda(x) = \operatorname{sign} \det f'(x)$, all $x \in \partial D$, or (b) $\operatorname{sign} \lambda(x) = -\operatorname{sign} \det f'(x)$, all $x \in \partial D$, which makes $-\lambda(x) f'(x)^{-1} f(x)$ point into D at each $x \in \partial D$.

Again, let $E = \{x \in D \mid f(x) = 0\}$. Then we can define the C^2 map $g = D \setminus E \to S^{n-1}$ as

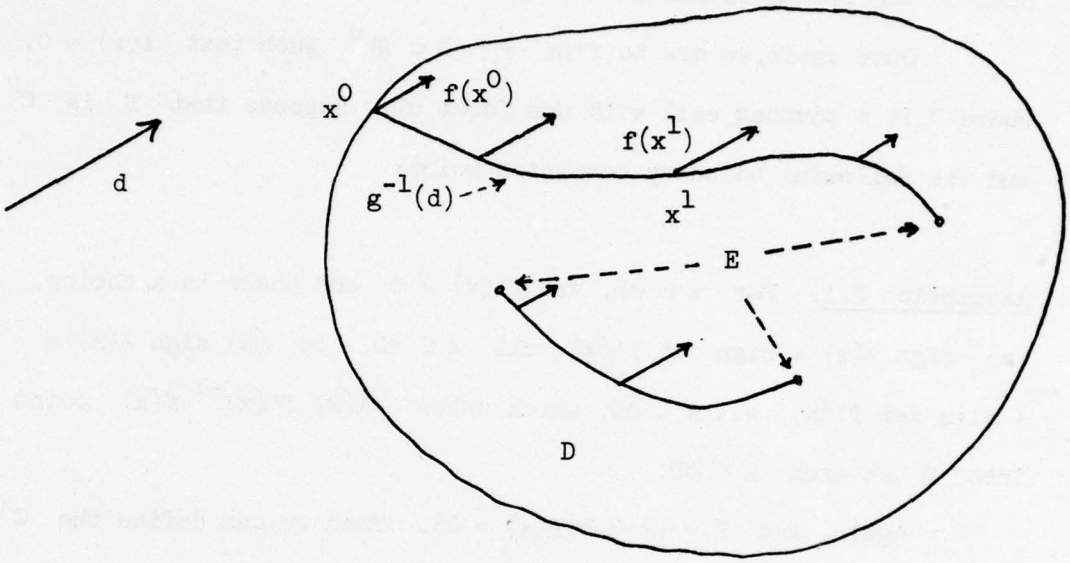
$$g(x) = f(x)/\|f(x)\| \quad . \tag{2.1}$$

So g maps every point in D which is not a zero of f to a point on the unit sphere in $\mathbb{R}^n$. Smale's main result was

Theorem 2.2. Let $f: D \to \mathbb{R}^n$ be C^2 and satisfy Assumption 2.1. Then for almost every $x^0 \in \partial D$, there exists a C' curve $\varphi = [t_0, t_1) \to D$ with $\|\dot{\varphi}(t)\| = 1$, and t_1 maximal, $t_1 \leq \infty$. Also if $d = g(x^0)$, $g^{-1}(d) = \{x \mid x = \varphi(t),\ t \in [t_0, t_1)\}$ and φ converges to E as $t \to t_1$.

He also proves a companion theorem with the added assumption that 0 is a regular value of f with the stronger result that φ converges to x^* as $t \to t_1$ and $x^* \in E$. We shall reformulate the problem and prove a result similar to the companion theorem.

Example 2.3.



$g^{-1}(d)$ is the set of points x for which $f(x)$ points in the direction d. Note that $g^{-1}(d)$ does not include its "boundary points" which are elements of E.

Let $F: D \times \mathbb{R}^1_+ \to \mathbb{R}^n$ be defined

$$F(x,\theta) = f(x) - \theta d \,, \tag{2.2}$$

where $d = f(x^0)/\|f(x^0)\|$ for some point $x^0 \in \partial D$. Then $(x,\theta) \in F^{-1}(0)$ for $\theta > 0$ implies $f(x)/\theta = d$ and since d has norm 1, $\theta = \|f(x)\|$. Thus $x \in g^{-1}(d)$. Clearly $x \in g^{-1}(d)$ implies $(x,\theta) \in F^{-1}(0)$ for $\theta = \|f(x)\|$. In this case, we have solved the problem when we find $(x,0) \in F^{-1}(0)$.

Since D is compact, we can define $K = \max_{x \in D} \|f(x)\|$, and $K < \infty$.

Pick an $L \in (K,\infty)$. Then $M = D \times [0,L]$ is the domain of interest for F.

Theorem 2.3. If $0 \in \mathbb{R}^n$ is a good value for F and Assumption 2.1 holds, then the path $\gamma \subset F^{-1}(\theta)$ beginning at (x^0,θ^0) has another endpoint $(x^*,0)$ such that $f(x^*) = 0$.

Proof. Consider solutions of the system

$$\begin{aligned} F(x,\theta) &= f(x) - \theta d = 0 \qquad \left(d = \frac{f(x^0)}{\|f(x^0)\|}\right) \\ (x,\theta) &\in \partial M \\ \theta &> 0 \end{aligned} \tag{2.3}$$

Clearly (x^0,θ^0) solve (2.3) because $x^0 \in \partial M$ and Assumption 2.1 implies that $f(x^0) \neq 0$ and so $\theta^0 = \|f(x^0)\| > 0$. No point $(x,\theta) \in F^{-1}(0)$ can have $\theta = L$ because otherwise $\|f(x)\| = L > K$ contradicting the definition of K.

Suppose there is an $(\bar{x},\bar{\theta}) \in \gamma$ such that $\bar{x} \in \partial D$, $\bar{\theta} \in (0,L)$. Let $(x(t),\theta(t))$ be a parametrization of γ by arclength with $(x(0),\theta(0)) = (x^0,\theta^0)$. Then $(\dot{x}(t),\dot{\theta}(t))$ is the tangent of γ at $(x(t), \theta(t))$ in the direction away from (x^0,θ^0). To compute $(\dot{x}(0),\dot{\theta}(0))$ we choose $v \in \mathbb{R}^n$ such that

(a) v points into D

(b) $f'(x)v = \lambda(x(0))d$

(c) $\lambda(x) = \operatorname{sgn} \det Df(x^0)$, WLOG

Then

$$(\dot{x}(0), \dot{\theta}(0)) = \frac{(v,\ \lambda(x^0))}{\|(v, \lambda(x^0))\|}$$

(WLOG means that if $(v, \lambda(x(0)))$ does not point into M for this definition, we can define $\lambda(x) = -\text{sgn det } f'(x)$.) Assumption 2.1 implies that the solution v of (b) is unique. As we move along $(x(t), \theta(t))$ let $\bar{t}$ such that $(x(\bar{t}), \theta(\bar{t})) = (\bar{x}, \bar{\theta})$ be the first $t > 0$ such that $(x(t), \theta(t)) \in \partial D \times (0, L)$. Then $\dot{x}(\bar{t})$ does not point into D and there is a $v \in \mathbb{R}^n$ such that

$$\frac{(\bar{v},\ \lambda(\bar{x}))}{\|(\bar{v}, \lambda(\bar{x}))\|} = (\dot{x}(\bar{t}), \dot{\theta}(\bar{t}))$$

and

$$f'(\bar{x})\bar{v} = \lambda(\bar{x})d = \lambda(x)\ \frac{f(\bar{x})}{\|f(\bar{x})\|}$$

because $\bar{x} \in g^{-1}(d)$. But

$$\bar{v} = f'(\bar{x})^{-1}\ \frac{\lambda(\bar{x})}{\|f(\bar{x})\|}\ f(\bar{x})$$

does not point into D and

$$v = f'(x^0)^{-1}\ \frac{\lambda(x^0)}{\|f(x^0)\|}\ f(x^0)$$

does point into D, a contradiction of Assumption 2.1. Thus there is no point in γ other than (x^0, θ^0) which satisfies (2.3) so by Corollary II.3.19 the other boundary point (x^*, θ^*) of γ must

$\theta^* = 0$, and, therefore, $f(x^*) = 0$, $x \in D$. ∎

The proof indicates the connection between the curve $g^{-1}(d)$ and the differential equation (2.0). Clearly the path following algorithm could be used if D was defined as $\{x|b_i(x) \leq 0,\ i \in \underline{m}\}$ for some smooth functions b_i, $i \in \underline{m}$ and F was defined as in (2.2).

IV.3. The Strong Path Method

In this section we define a deformation F for which one can follow $F^{-1}(0)$ to find a solution of $f(x) = 0$, $x \in D$, with quite weak boundary conditions on D. We shall then show that Varian's method [1977] is in some sense a special case of our method, and we shall use a result of Friedenfelds [1976] to show that the path we define is the same as the limiting path of Eaves' [1971] vector labelling algorithm as the mesh size goes to zero.

Let $D \subset \mathbb{R}^n$ be a cell defined as in Chapter I, $D = \{x|b_i(x) \leq 0,\ i \in \underline{m}\}$. Clearly D has a nonempty interior because of the constraint qualification on the b_i's. In this section we make the following

Assumption 3.1. There is some interior point $c \in D$ such that for any $x \in \partial D$, there is no $\alpha > 0$ such that $\alpha f(x) = x-c$.

If $D = B^n$ and $c = 0$, then the assumption states that there is no $x \in S^{n-1}$ such that $f(x)$ points radially outwards. If the condition fails for f we could test it for $-f$ and solve for $-f(x) = 0$ so that a weaker form of 3.1 could be given.

Assumption 3.2. There is some $c \in D^0$ such that there are no $x, y \in \partial D$ for which there are scalars $\beta < 0 < \alpha$ such that $\beta f(x) = x-c$, $\alpha f(y) = y-c$.

Next we define the deformation $F: M \to \mathbb{R}^n$ where $(M, \mathcal{M})$ is a subdivided $(n+1)$-complex.

$$F(x,\theta,\rho) = \theta f(x) + \rho(c-x) - (1-\theta)d \,, \tag{3.1}$$

where $x \in D$, $0 \leq \theta \leq 1$, $\rho \geq 0$. For the moment we define

$$\mathcal{M} \equiv \{\sigma_0, \sigma_1, \cdots, \sigma_m\},$$

in which

$$\sigma_0 \equiv \{(x,\theta,\rho) \,|\, x \in D,\ \theta \in I,\ \rho = 0\} \,,$$

$$\sigma_i \equiv \{(x,\theta,\rho) \,|\, x \in \tau_i,\ \theta \in I,\ \rho \geq 0\} \,, \qquad i \in \underline{m},$$

$$\tau_i \equiv \{x \,|\, b_i(x) = 0,\ b_j \leq 0,\ j \neq i\}, \qquad i \in \underline{m},$$

and

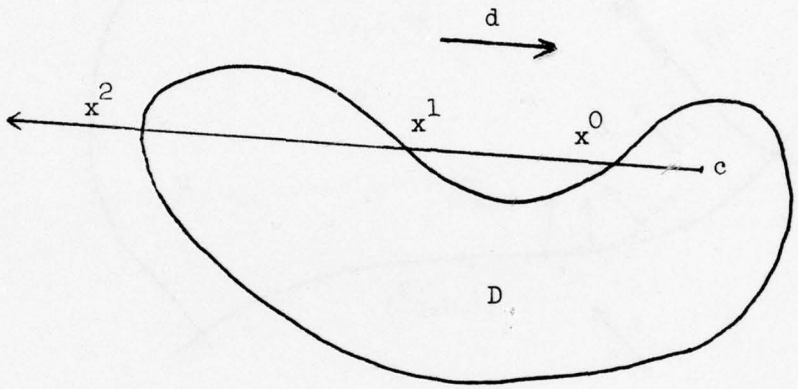
$$I = [0,1].$$

When x is in D^0, then $(x, \theta, \rho) \in M$ implies that $\rho = 0$ and F is essentially the deformation defined for the Global Newton method (2.2); $F\Big|^{-1}_{\sigma_0}(0) \ni (x,\theta,0)$ for $\theta < 1$ implies that $f(x) = \theta/(1-\theta)d$ or $f(x)$ points in the same direction as d.

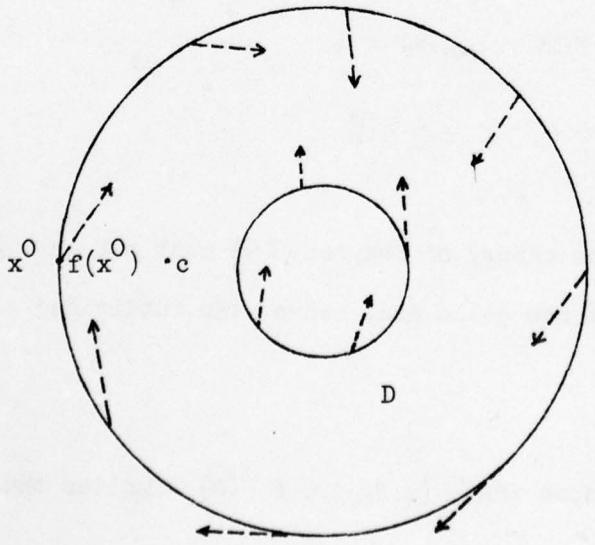
When $x \in \partial D$, then $(x,\theta,\rho) \in M$ is still in an $n+1$ cell because ρ is allowed to increase from zero. Note that any facet of both σ_i and σ_0 for some $i \in \underline{m}$ is defined by the fact that ρ is identically zero on that facet.

Assumption 3.3. c and $d \in \mathbb{R}^n$ are chosen so that the ray $r = \{y | y = c - \alpha d,\ \alpha \geq 0\}$ intersects ∂D in only one point $\{x^0\}$, and x^0 is in only one facet τ_k of D.

Example 3.4. a) fails to satisfy Assumption 3.3.

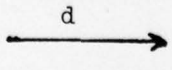


b) satisfies Assumptions 3.3 and 3.1.

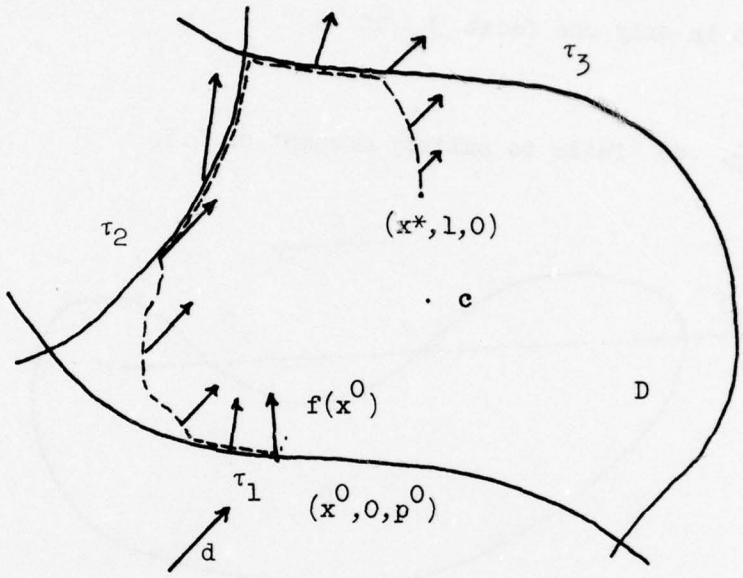


$D = \overline{B(0,2) \setminus B(0,1)}$

The vectors on the boundary of D represent the value of f(x) at those points.



Example 3.5. The path $F^{-1}(0)$ projected onto the set D might look like this.



Define $P_x(S)$ for some set $S \subset M$ to be $\{x \mid (x,\theta,\rho) \in S\}$. Points in $P_x(F^{-1}(0)) \cap \partial D$ have the property that d is in the cone spanned by $f(x)$ and $(c-x)$, i.e., if $\theta < 1$,

$$\frac{\theta}{1-\theta} f(x) + \rho(c-x) = d$$

$$\frac{\theta}{1-\theta} \geq 0, \qquad \rho \geq 0 .$$

In order to apply the theory of Chapter II we must put an upper bound on ρ to make each of the cells and, hence, the subdivided complex M compact sets.

Lemma 3.6. There is a Q such that $(x,\theta,\rho) \in F^{-1}(0)$ implies that $\rho < Q$.

Proof. Since $\rho \equiv 0$ for points $(x,\theta,\rho) \in F^{-1}(0)$ where $x \in D^0$. Assume there is a sequence $\{x^k, \theta^k, \rho^k\} \subset F^{-1}(0)$ such that $x^i \in \partial D$ for all k and for any $Q > 0$, there is a k such that $\rho^k > Q$. Now

$$\frac{\theta^k}{\rho^k} f(x^k) + (c-x^k) - \frac{(1-\theta^k)}{\rho^k} d = 0 .$$

Since $(x^k,\theta^k) \in \partial D \times I$, a compact set, and f is continuous,

$$\lim_{k\to\infty} \frac{\theta^k f(x^k)}{\rho^k} = 0 = \lim_{k\to\infty} \frac{(1-\theta^k)}{\rho^k} .$$

This implies that $\lim_{k\to\infty} c - x^k = 0$, which is impossible because c is an interior point for D. ∎

Thus we can let $J \equiv [0,Q]$ and redefine $\sigma_i \equiv \tau_i \times I \times J$, $i \in \underline{m}$, and let $(M,\mathcal{M})$ be redefined accordingly.

Theorem 3.7. Given Assumptions 3.1 and 3.3 and that 0 is a good value for F with respect to $(M,\mathcal{M})$, the path $\gamma \subset F^{-1}(0)$ with $(x^0,0,\rho^0)$ $(x \in \partial D,\ \rho^0 = \|d\|/\|c-x^0\|)$ as one boundary point has an opposite boundary point $(x^*,1,0)$ and $x^* \in E = \{x \in D | f(x) = 0\}$. Furthermore $|E|$ is odd.

Proof. Clearly, $(x^0,0,\rho^0) \in \partial M \cap F^{-1}(0)$ by the definition of x^0 in Assumption 3.3. ρ^0 is positive, so $(x^0,0,\rho^0)$ is in only one facet of the cell σ_k (remember, $x^0 \in \tau_k$). Since 0 is a good value, there

is a unique connected component $\gamma \subset F^{-1}(0)$ which contains $(x^0,0,\rho^0)$.

Next we show that there is only one solution of

$$\begin{aligned} F(x,\theta,\rho) &= 0 \\ \theta &< 1 \\ (x,\theta,\rho) &\in \partial M \end{aligned} \tag{3.2}$$

By studying the defintion of $(M,\mathcal{\eta})$, one notes that $(x,\theta,\rho) \in \partial M$ means either $\theta = 0$ or 1, or $\rho = Q$. By the definition of Q, the latter type of boundary point is impossible.

If $\bar{\theta} = 0$, then $F(\bar{x},\bar{\theta},\bar{\rho}) = 0$ implies

$$\bar{\rho}(c-\bar{x}) = d \Rightarrow \bar{\rho} > 0 \Rightarrow \bar{x} \in \partial D ,$$

or

$$\bar{x} = c - \frac{1}{\bar{\rho}} d .$$

By Assumption 3.3, $\bar{x} = x^0$, and the solution to (3.2) is unique.

Thus, the other endpoint of $\gamma, (x^*,\theta^*,\rho^*)$, must have $\theta^* = 1$. Suppose $\rho^* > 0$, then $x^* \in \partial D$ and

$$f(x^*) = \rho^*(x^*-c) ,$$

a contradiction of Assumption 3.1. Hence, $\rho^* = 0$ and $f(x^*) = 0$.

The fact that $|E|$ is odd is a simple consequence of the following facts: $|F^{-1}(0) \cap \partial M|$ is even (Prop. I.3.18), (3.2) has a unique solution, and $(x,1,\rho) \in F^{-1}(0)$ implies $\rho = 0$ and $x \in E$. ∎

The assumption that 0 is a good value for F with respect to $(M, \mathcal{M})$ removes the possibility of boundary solutions, i.e., $\bar{x} \in \partial D$ such that $f(\bar{x}) = 0$. In this case $(\bar{x},1,0) \in F^{-1}(0) \cap \partial M$, $(\bar{x},1,0) \in \sigma_0$ and $(\bar{x},1,0) \in \sigma_i$ for some $i \in \underline{m}$, a contradiction that $F^{-1}(0)$ meets only the interior of facets of cells in $\mathcal{M}$. One could easily relax the regularity conditions on Theorem 3.7 so that x* could be a boundary point of D, but the oddness of $|E|$ could not be guaranteed.

IV.4. Varian's Modification of the Global Newton Method.

Varian [1977] proposed a modification of Smale's method which allowed the boundary conditions to be weakened considerably. We shall show how Varian's modification is related to the strong path method.

Suppose $f: D^n \to \mathbb{R}^n$ is a C^2 function, where $D^n = B(0,1)$ the closed unit disc in $\mathbb{R}^n$. The problem is, as usual to find x which solves $f(x) = 0$, $x \in D^n$. Suppose f satisfies the following

Assumption 4.1. At all $x \in \partial D^n$, there is no $\alpha > 0$ such that $f(x) = \alpha x$.

It is clear that this boundary condition implies Assumption 3.1 when $D = D^n$ and $c = 0 \in \mathbb{R}^n$.

Let D_2^n be a disc in $\mathbb{R}^n$ of radius 2. Let $s(x) = \|x\| - 1$ and define the following function on D_2^n:

$$h(x) = -s(x)\cdot\frac{x}{\|x\|} + (1-s(x))\, f\left(\frac{x}{\|x\|}\right), \qquad 1 < \|x\| \leq 2$$
$$= f(x), \qquad 0 \leq \|x\| \leq 1. \tag{4.1}$$

This function coincides with f on D^n and is a continuous extension on $D_2^n \setminus D^n$. In fact h is C^2 on $D_2^n \setminus D^n$ and on the interior of D^n, but not on ∂D^n.

We must define a subdivided complex $(M, \mathcal{M})$, where $\mathcal{M} = \{D^n \times \mathbb{R}^1, \overline{D_2^n \setminus D^n} \times \mathbb{R}^1_+\}$, so that h is continuous on M and C^2 on pieces of $\mathcal{M}$. The deformation is identical to that of Smale [1976].

$F: M \to \mathbb{R}^n$ is denoted by $F(x,\theta) = h(x) - \theta d$, and $d = h(x^0)/\|h(x^0)\|$ for some $x^0 \in \partial D_2^n$.

For $x \in \partial D_2^n$, $h(x) = -x/2$ and $h'(x) = -I/2$; hence it is easy to see that h satisfies Assumption 2.1. Since D_2^n is compact, we can define $K = \max_{x \in D_2^n} \|h(x)\|$, and pick $L > K$. Then

$\mathcal{M} = \{D^n \times [0,L], \overline{D_2^n \setminus D^n} \times [0,L]\}$ is a compact subdivided complex, and we can prove, almost immediately, the

Proposition 4.2. If 0 is a good value for F with respect to $(M, \mathcal{M})$ and Assumption 4.1 holds, then $\gamma \subset F^{-1}(0)$ beginning at $(x^0, \theta^0) \in \partial D_n^2 \times (0,L)$ has another boundary point $(x^*,0)$ such that $f(x^*) = 0$.

Proof. Since Assumption 2.1 is satisfied for h, it is immediate that (x^*, θ^*) has $\theta^* = 0$ by the proof of Theorem 2.3. All we need to show is that $x^* \in D^n$, because then $F(x^*,0) = h(x^*) = f(x^*) = 0$.

Suppose x^* is in $D_2^n \setminus D^n$. If $\|x^*\| < 2$, then $1 - s(x^*) > 0$, and

$$h(x^*) = - s(x^*) \frac{x^*}{\|x^*\|} + (1-s(x^*)) f(\frac{x^*}{\|x^*\|}) = 0$$

or

$$f\left(\frac{x^*}{\|x^*\|}\right) = \frac{s(x^*)}{1-s(x^*)} \frac{x^*}{\|x^*\|}$$

but this contradicts Assumption 4.1.

If $\|x^*\| = 2$, then $h(x^*) = -x^*/2 \neq 0$, a contradiction. Hence $x^* \in D^n$. ∎

The formulation above is essentially the same as Varian's, with the following differences. Varian considers $g: D_2^n | E \to S^{n-1}$ denoted by $g(x) = h(x)/\|h(x)\|$ and the path defined by $g^{-1}(d)$, where d is defined as above. The regularity assumption Varian makes is that d is a regular value for g restricted to $D_2^n \setminus D^n$, g restricted to D^n and g restricted to the boundary of D^n. This assumption implies that d is a good value if one assumes there are no zeros of f on ∂D^n. It is easy to see that $\{x | (x,\theta) \in F^{-1}(0)$ for some $\theta \geq 0\} \setminus E = g^{-1}(d)$. That is, the two formulations provide essentially identical paths.

Figure 4.3 provides some geometrical insight into the method.

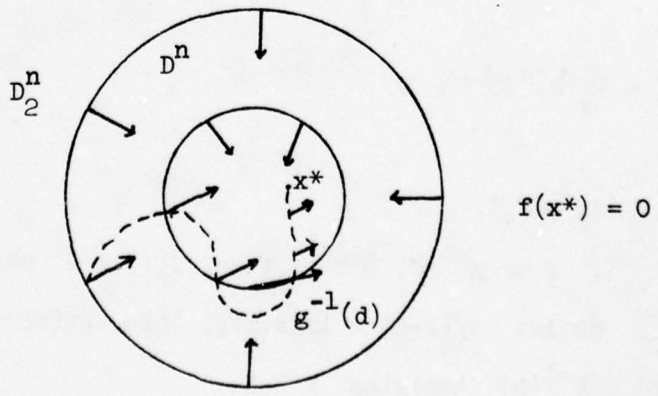


FIGURE 4.3. The vectors along the path $g^{-1}(d)$ are all parallel to each other.

The relationship of Varian's method to the strong path method is that if the portion of the path $g^{-1}(d)$ contained in $D_2^n \setminus D^n$ were projected onto the boundary of D^n, then the resulting path would be identical to the path followed by the strong path method for D^n in the x-variables. The following proposition makes this statement more precise.

Remember the deformation for the path method (3.1)

$$F(x,\theta,\rho) = \theta f(x) + \rho(c-x) - (1-\theta)d ,$$

defined on the subdivided (n+1)-complex $(M,\mathcal{H})$ where $\mathcal{H} \equiv \{D^n \times [0,1] \times 0,\ \partial D^n \times [0,1] \times [0,Q]\}$, where $Q > 0$ is chosen so that $\rho < Q$ for any $(x,\theta,\rho) \in F^{-1}(0)$. The interior point c is chosen as the origin and $d \equiv -x^0/2$ for some initial point $x^0 \in \partial D^n$.

Proposition 4.4. The following sets are identical:

$$S_1 \equiv (g^{-1}(d) \cap D^n) \cup \{x \mid x = y/\|y\|,\ y \in (g^{-1}(d) \cap D_2^n \setminus D^n)\}$$

and

$$S_2 \equiv P_x(F^{-1}(0)) \setminus E.$$

Proof. $(S_1 \subset S_2)$.

If $x \in (g^{-1}(d) \cap D^n)$, then $f(x) \neq 0$ so we can rule out $\theta = 1$ in S_2. So let $\theta/(1-\theta) = 1/\|f(x)\|$, then $\theta f(x) = (1-\theta)d$ or $(x,\theta,0) \in F^{-1}(0)$ implying $x \in S_2$.

If $y \in (g^{-1}(d) \cap D_2^n \setminus D^n)$, then $h(y) \neq 0$ so we have

$$\frac{1}{\|h(y)\|}\,(s(y)(-\frac{y}{\|y\|}) + (1-s(y))\, f(\frac{y}{\|y\|}) = d\ ,$$

and if $x = y/\|y\|$, $\theta/(1-\theta) = (1-s(y))/\|h(y)\|$, and $\rho = (s(y)/\|h(y)\|)\cdot(1-\theta) \geq 0$, we have

$$\theta f(x) + \rho(-x) - (1-\theta)d = 0$$

or $F(x,\theta,\rho) = 0$. Of course, $x \in \partial D^n$ implies that $(x,\theta,\rho) \in M$, and we have that $S_1 \subset S_2$.

($S_2 \subset S_1$)

We shall prove only the case when $x \in \partial D^n$. We must show that there is some $\alpha \in [0,1]$ such that $y_\alpha = (1+\alpha)x \in g^{-1}(d)$. Clearly, for such y, $s(y) = \alpha$ and $h(y_\alpha) = \alpha(-x) + (1-\alpha)\, f(x)$.

$y_\alpha \in g^{-1}(d)$ if and only if $h(y_\alpha)$ points in the same direction as d. We have $\theta f(x) + \rho(-x) = (1-\theta)d$ for some $\theta,\ \rho \geq 0$. We have $\theta + \rho > 0$, so

$$\frac{\theta}{\theta+\rho}\, f(x) + \frac{\rho}{\theta+\rho}\,(-x) = \frac{1-\theta}{\theta+\rho}\ .$$

Let $\alpha = \rho/(\theta+\rho)$, then $1-\alpha = \theta/(\theta+\rho)$ and

$$\alpha(-x) + (1-\alpha)\, f(x) = \frac{1-\theta}{\theta+\rho}\, d\ ,$$

so if $y = (1+\alpha)x$, $y \in g^{-1}(d)$. ∎

IV.5. The Strong Path Method in Relation to a Fixed Point Algorithm.

The path prescribed by the strong path method, is very nearly, the path which would be followed by some fixed point algorithms for solving the same problem. This is not surprising because the strong path method is related to Smale's [1976] "Global Newton Method" which was a differential version of Scarf's [1973] fixed point algorithm stated in terms of solving systems of equations.

The algorithm we shall discuss is a vector labelling algorithm for solving Kakutani fixed points developed by Eaves [1971], and discussed by Friedenfelds [1976].

For a complete description of this algorithm one should see one of the above.

Suppose f is a map from C to C^*, the set of all convex subsets of C, which has a closed graph. Then Kakutani's theorem [1941] states that there is an $x \in C$ such that $x \in f(x)$. Such an x is called a fixed point of f. Assume that C is contained in the hyperplane $H = \{x \in \mathbb{R}^n | \sum_{i=1}^{n} x_i = 1\}$, and that a point $c \in C^0$ is available. Let N be an arbitrary positive integer, and let Π be the set of points in H defined as

$$\Pi \equiv \{x \in \mathbb{R}^n | x = \frac{1}{N} y, \text{ where } \sum_{i=1}^{n} y_i = N, \text{ and } y_i \text{ is an integer, } i \in \underline{n}\}$$

Assume that H is triangulated by Kuhn's [1968] method with Π as the set of

We will define a piecewise linear approximation to f below. First, extend f to $H \supset C$ by defining for each $x \in H$,

$$\bar{f}(x) \equiv \begin{cases} f(x) & \text{if } x \in \text{int } C \\ \text{conv}[c,\ f(x)] & \text{if } x \in \partial C \\ c & \text{if } x \notin C. \end{cases}$$

The extended $\bar{f}$ is a closed map taking points of H into convex subsets of C. Furthermore, the fixed points of the extended f' are precisely those of the original f on the domain of C (Friedenfelds [1976], p. 16).

For each point $x \in \Pi$, define the function $g(x) = y-x$, where $y \in f'(x)$; and let d be an arbitrary vector in $\mathbb{R}^n$ such that $\sum_{i=1}^{n} d_i = 0$. Let $\bar{g}: H \to \{x | x \in \mathbb{R}^n,\ \sum_{i=1}^{n} x_i = 0\} \equiv H^0$ be the continuous function which is linear on each $\sigma \in \mathcal{M}$ and agrees with g at the vertices of σ.

The fixed point algorithm consists of following the path $G^{-1}(0)$ where $G: H \to H^0$ is defined as

$$G_N(x, \theta) = \bar{g}(x) + \theta d\ , \quad \text{and}$$

$(\bar{M}, \overline{\mathcal{M}})_N$ is defined by $\overline{\mathcal{M}}_N = \{\sigma \times \mathbb{R}^1_+ | \sigma \in \mathcal{M}_N\}$. Define, the direction ray, r, as $r \equiv \{(x,\gamma) \in \mathbb{R}^{n+1} | x = c + \gamma d,\ \gamma \in \mathbb{R}^1_+\}$. Assume the points of r are such that no point of r is a convex combination of less than $(n-1)$ points of Π. For any $\sigma \in \overline{\mathcal{M}}$, we say that σ touches r if $r \cap \sigma \neq \emptyset$. Then it can be shown that if $(x, \theta) \in \sigma \subset H \setminus C$ and $(x, \theta) \in G^{-1}(0)$ then σ touches r. Friedenfelds ([1976], p. 17) shows that if one begins in a cell σ_0 that touches $r \setminus C$, and moves

along $G^{-1}(0)$ towards C, after passing through a finite number of cells, one arrives at a point $(\bar{x},0) \in G^{-1}(0)$ imply that $\bar{g}(\bar{x}) = 0$. $\bar{x}$ is called an approximate fixed point of f. Eaves [1971] shows that, if $\bar{x}^N$ is the approximate fixed point for the algorithm when N is the denominator in the Kuhn triangulation of H, then each cluster point of $\{\bar{x}^N\}$ is a fixed point of f.

Define the collection of cells

$$C^N = \{\sigma \in (M, \mathcal{M})_N | \sigma \cap G_N^{-1}(0) \neq \varphi\} .$$

The set of <u>almost-complete points</u> A_d is defined as

$$\begin{aligned} A_d &\equiv \{x \in H | 0 \in \text{conv}[\bar{f}(x)-x,\ d]\} \\ &= \{x \in H | x \in \text{conv}[\bar{f}(x),\ x+d]\} \end{aligned}$$

In terms of the original function, f, the set, A_d, within C is given by

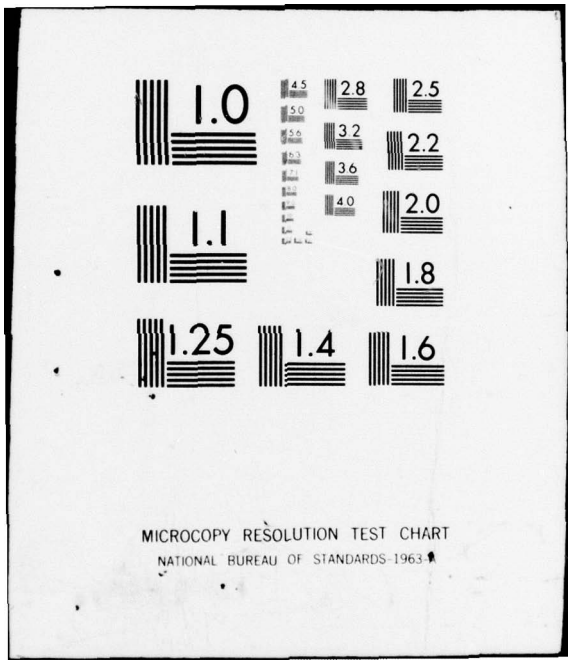
$$A_d \cap C = \left\{ x \in C \middle| x \in \begin{cases} \text{conv}[f(x), x+d] & \text{if } x \in \text{int } C \\ \text{conv}[f(x), x+d, c] & \text{if } x \in \partial C \end{cases} \right\} \qquad (5.1)$$

It is easy to check that every sequence of cells $\{\sigma^N\}$ with $\sigma^N \subset C^N$ converges to A_d as $N \to \infty$. It is easy to check that the points of A_d outside of C are precisely those on the direction ray r. For any $N = L$ the fixed point algorithm can be started at some $\sigma_0 = \text{conv}[p^1, \ldots, p^n] \subset H \setminus C$. Let r_0 be an arbitrary point in $\sigma_0 \cap r$. Then for each $N \geq L$, the algorithm can be started so that

AD-A053 200 STANFORD UNIV CALIF SYSTEMS OPTIMIZATION LAB F/G 12/1
ON THE SOLUTION OF NONLINEAR EQUATIONS BY PATH METHODS.(U)
OCT 77 T R ELKEN N00014-75-C-0267
UNCLASSIFIED SOL-77-25 NL

2 OF 2
AD
A053200

END
DATE
FILMED
6-78
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963

r_0 is in the convex hull of the original set. We call r_0 the initial almost-complete point. Application of the fixed point algorithm for any $N \geq L$ then gives us a succession of points of H, $p_1,\ldots,p_n$, $p_{n+1},\ldots,p_\ell$, where $r_0 \in \text{conv}[p_1,\ldots,p_n]$; $\text{conv}[p_{i+1},\ldots,p_{i+n}] \in C^N$ for $i = 0, 1,\ldots,\ell-n$; and $\text{conv}[p_{\ell-n+1},\ldots,p_\ell]$ contains an approximate fixed point. Let $p_N(r_0) \equiv \{p_1,\ldots,p_\ell\}$ and let $\bar{p}_N(r_0)$ be the continuous path formed by joining p_i to p_{i+1}, $i = 1,\ldots, \ell-1$. We call $\bar{p}_N(r_0)$ the approximate almost-complete path corresponding to N. One of Friedenfelds [1976] main results concerned the limiting properties of $\bar{p}_N(r_0)$. Let $A'_d \equiv (A_d \cap \text{conv}[c,r_0]) \cup (A_d \cap C)$--i.e., A'_d is A_d with the points of ray r further from C than r_0 excluded. Let $P_0(r_0)$ denote the connected component (i.e., the largest connected subset) of A'_d which contains r_0.

Theorem 5.1. (Friedenfelds [1976])

(i) $P_0(r_0)$ is a closed set,

(ii) $\bar{p}_N(r_0) \to P_0(r_0)$ as $N \to \infty$
(i.e., given $\epsilon > 0$, $\bar{p}_N(r_0) \cap B(P_0(r_0),\epsilon)$ for all N sufficiently large).

The conclusion that $P_0(r_0)$ is a path cannot be drawn with the assumption that f is a closed point to set map. However, if f is C^2 then we can apply the results of Section II.3 to show that $P_0(r_0)$ is a path. In fact if $P_0(x_0) = P_0(r_0) \cap C$ where $x_0 = r \cap \partial C$, then $P_0(x_0)$ is the same path as that produced by the strong path

method, projected onto H. To make this statement precise requires some preliminaries.

First, since the strong path method is for solving systems of equations, rather than finding fixed points, we will be solving for a zero of

$$g(x) \equiv f(x) - x , \qquad x \in C .$$

Second, for simplicity, we will define the subdivided complex on which

$$F(x,\theta,p) = \theta g(x) + \rho(c-x) - (1-\theta)(-d) \tag{5.2}$$

is defined as $(M,\mathcal{M})$ where

$$\mathcal{M} = \{\sigma_1,\sigma_2\} = \{c \times [0,1] \times 0,\ \partial C \times [0,1] \times [0,Q]\}$$

$$M = \sigma_1 \cup \sigma_2 ,$$

where $Q > 0$ is chosen as in Lemma 3.6 so that $\rho < Q$ for any $(x,\theta,p) \in F^{-1}(0)$. This subdivision ignores the fact that C must be defined by a set of inequalities for any practical applications. We use $-d$ in the definition of F (5.2) because the almost complete path is the set of points for which $g(x)$ points in the direction opposite to d (i.e., $0 \in \text{conv}[g(x),d]$) for points in the interior of C.

From Theorem 3.8 we have that if 0 is a good value for F, and $\gamma \in F^{-1}(0)$ is the path containing $(x^0,0,\rho^0)$, then the other boundary point of γ, $(x^*,1,0)$, is such that $g(x^*) = 0$.

Proposition 5.2. Given Assumptions 3.1 and 3.2, we have $A_d \cap C = P_x(F^{-1}(0))$.

Proof. Suppose that $x \in P_x(F^{-1}(0))$, then if $x \in \partial C$, there is some pair θ, $\rho \geq 0$ such that

$$\theta g(x) + \rho(c-x) = -(1-\theta)d \ .$$

$\vartheta = 1$ implies that $\rho = 0$ and $g(x) = 0$, otherwise x violates Assumption 3.1. But $g(x) = 0 \Rightarrow x \in A_d$. If $\theta < 1$, then let $\alpha = \rho/(1-\theta) \geq 0$, $\beta = \rho/(1-\theta) \geq 0$ and we have

$$\alpha(f(x) - x) + \beta(c-x) = - d,$$

$$\alpha f(x) + \beta c = (-d-x) + (\alpha + \beta + 1)x \ .$$

If $\mu = \alpha + \beta + 1$, then

$$\frac{\alpha}{\mu} f(x) + \frac{\rho}{\mu} c + \frac{1}{\mu} (d+x) = x \ ,$$

which means that $x \in \text{conv}[f(x), x+d, c]$. Thus $x \in A_d \cap \partial C$ by the definition (5.1). Clearly, the argument can be reversed to show that $x \in A_d \cap \partial C$ implies that $x \in P_x(F^{-1}(0))$ because any choice of $\lambda_1, \lambda_2, \lambda_3$ such that $\sum \lambda_i = 1$, $\lambda_i \geq 0$, $i \in \underline{3}$ can be written, α/μ, β/μ, $1/\mu$.

If $x \in P_x(F^{-1}(0)) \cap \text{int } C$, then ρ must be zero, and there is some $\theta \in [0,1]$ such that

$$\theta g(x) + (1-\theta)d = 0$$

If $\theta = 1$, $g(x) = 0$ and $x \in A_d \cap C$. If $\theta < 1$, then let $\alpha = \theta/(1-\theta)$ and we have

$$\alpha(f(x)-x) + d + x = x$$

$$\iff \alpha f(x) + d + x = (1 + \alpha)x$$

$\iff$ there exist $\lambda_1, \lambda_2 \geq 0$, $\lambda_1 + \lambda_2 = 1$ such that $\lambda_1 f(x) + \lambda_2(d+x)=x$

$\iff$ $x \in \text{conv}[f(x), d+x]$.

Thus, by (5.1), $x \in A_d \cap C$. This proves the proposition. ∎

The following relates the theorem to the paths which the algorithms follow.

Corollary 5.3. $P_0(x^0) = P_x(r)$.

We have shown that path methods essentially follow the same course that a fixed point algorithm follows. One might ask what the point is in considering path methods when the fixed point methods are more robust and do not require differentiability or continuity assumptions. It is the author's contention that many equation solving problems deal with differentiable functions which are sufficiently well behaved so that path methods can process them faster than fixed point methods. Our computational results in Chapter V, Part 2 help to support this contention. Of course, fixed point methods are necessary when the functions are non-differentiable or map points into sets.

IV.6. A Class of Path Methods

A number of fixed point algorithms can be written in a rather general form. One need merely specify a triangulation of $\mathbb{R}^n \times (0,D]$, an artificial map $r:\mathbb{R}^n \to \mathbb{R}^n$ for labeling $\mathbb{R}^n \times \{D\}$, and a method for labeling the vertices in $\mathbb{R}^n \times (0,D)$ using r and $\ell(x) = f(x)-x$. In particular, the algorithm of Eaves and Saigal [1972] chooses r as a one-to-one linear map with a unique point x_0 such that $r(x_0) = 0$, and a decreasing map $\alpha = (0,1] \to (0,\infty)$ such that $\alpha(1) = 0$, $\alpha(t) \to +\infty$ as $t \to 0$. Then the labeling L of a vertex $(x,t) \in \mathbb{R}^n \times (0,D]$ is defined by

$$L(x,t) = f(x) - x \quad \text{if} \quad \alpha(t) > \sum_{i=1}^{n} |x_i|$$

$$= r(x) \qquad \text{if not .}$$

A geometric interpretation of the algorithm is to let $F(x,t)$ be the continuous extension of $L(x,t)$ which is linear on each piece of the subdivision of $\mathbb{R}^n \times (0,D]$. Then, beginning at (x_0,D), follow $F^{-1}(0)$ until one reaches a point (x^*,t^*) where t^* is close to zero or until $\|x^*\| \geq K$ for some large constant K. The former termination means that x^* is an approximate fixed point, the latter means that the algorithm appears to have failed.

One condition for success is contained in

Theorem 6.1. Let C be an open bounded set containing x_0 such that $\ell(x) + \rho r(x) \neq 0$ for all $\rho \geq 0$ and $x \in \partial C$. Then there is a $D > 0$ such that the algorithm will compute a fixed point of f.

Proof. See Saigal [1976a], for example.

Now it is clear that an analogous class of path methods can be defined with only minor changes. Let

$$r(x) = Ax - a$$

for some nonsingular matrix A, and $f:\mathbb{R}^n \to \mathbb{R}^n$ is the function we are trying to find a zero of. Define the deformation

$$F(x,\theta) = \theta r(x) + (1-\theta)f(x) , \qquad x \in \sigma,\ \theta \in [0,1] , \tag{6.1}$$

where σ is a bounded n-cell in $\mathbb{R}^n$. Let $\mathcal{H} = \{\sigma \times [0,1]\}$. Then if 0 is a good value for F one could follow the path $\gamma \subset F^{-1}(0)$ containing $(x_0,1)$, where $x_0 = A^{-1}a \in \sigma^0$. The algorithm would terminate at (x^*,θ^*) if either $\theta = 0$ or $x^* \in \partial\sigma$. Since $r(x)$ has a unique zero, it is impossible for γ to hit the facet $\sigma \times \{1\}$.

It is trivial to prove the following

Theorem 6.2. Suppose that there is no $x \in \partial\sigma$ such that $f(x) + \rho r(x) = 0$ for some $\rho \geq 0$. Then if $0 \in \mathbb{R}^n$ is a good value for F w.r.t. $\sigma \times [0,1]$, the path $\gamma \subset F^{-1}(0)$ containing $(x_0,1)$ has an opposite boundary point $(x^*,0)$.

Proof. By the discussion preceding the theorem, we need only show that there is no $(\bar{x},\bar{\theta}) \in \gamma$ such that $\bar{x} \in \partial\sigma$. If there was, then

$$(1-\bar{\theta})\, f(\bar{x}) + \bar{\theta} r(\bar{x}) = 0$$

for some $\bar{\theta} < 1$. Hence, if $\bar{\rho} = \bar{\theta}/(1-\bar{\theta}) \geq 0$, then

$$f(\bar{x}) + \bar{\rho} r(\bar{x}) = 0 , \qquad \bar{x} \in \partial\sigma ,$$

a contradiction. ∎

As a special case of Theorem 6.2 we can achieve a path method which is convergent with assumptions just as weak as those necessary for the strong path method of Section 3.

Corollary 6.3. Let $r(x) = -x + c$ for some $c \in \sigma^0$. Suppose that Assumption 3.1 holds and that 0 is a good value for F as defined in (6.1), then $\gamma \subset F^{-1}(0)$ which contains (0,1) leads to a zero of f.

The advantage of this class of path methods is that the starting point x^0 is not a boundary point of σ. So if a good guess to a zero of f is known, that guess can be the starting point.

A useful area of research would be to apply the results of Saigal [1976b] to the path methods presented in this section. The results contained therein suggest that if x^0 is close to x^*, a root of f, then the path $\gamma \subset F^{-1}(0)$ will be shortest if $A = f'(x^*)$. Other results suggest that if $A^{-1}f'(x)$ has positive real eigenvalues for points x near x^*, then the path will be well defined.

REFERENCES

Davidenko, D. [1953a], "On a New Method of Numerically Integrating a System of Nonlinear Equations," Dokl. Akad. Nauk. SSR, 88, pp. 601-604 (in Russian).

Davidenko, D. [1953b], "On the Approximate Solution of a System of Nonlinear Equations," Ukrain. Mat. Z., 5, pp. 1963-206 (in Russian).

Eaves, B.C. [1971], "Computing Kakutani Fixed Points," SIAM J. Appl. Math., 21, pp. 236-44.

Eaves, B.C. [1976], "A Short Course in Solving Equations with P.L. Homotopies," SIAM-AMS Proceedings, 9, pp. 73-143.

Eaves, B.C. and R. Saigal [1972], "Homotopies for Computation of Fixed Points on Unbounded Regions," Math. Prog., 3, pp. 225-37.

Eaves, B.C. and H. Scarf [1976], "The Solution of Systems of Piecewise Linear Equations," Math. of Operations Research, 1, pp. 1-27.

Elken, T. [1977], "The Computation of Economic Equilibria by Path Methods," SOL Technical Report 77-26, Department of Operations Research, Stanford University, Stanford, California.

Friedenfelds, J. [1976], "Fixed Point Algorithms and Almost-Complementary Sets," Fixed Points: Algorithms and Applications, S. Karamardian, Editor, Academic Press.

Freudenstein and Roth [1963], "Numerical Solution of Systems of Nonlinear Equations," J.A.C.M., 10, pp. 550-56.

Hirsch, M. [1963], "A Proof of the Nonretractibility of a Cell onto its Boundary," Proc. Amer. Math. Soc. 14, pp. 364-65.

Hirsch, M. [1976], Differential Topology, New York: Springer Verlag.

Kakutani, S. [1941], "A Generalization of Brower's Fixed Point Theorem," Duke Math. J., 8, pp. 457-59.

Kellogg, R.B., T.Y. Li, and J. Yorke [1976], "A Constructive Proof of the Brouwer Fixed-Point Theorem," SIAM J. Numer. Anal., 13, pp. 473-83.

Kuhn, H.W. [1968], "Simplicial Approximation of Fixed Points," Proc. Nat. Acad. Sci. U.S.A., 61, pp. 1238-42.

Lahaye, E. [1934], "Une méthode de resolution d'une categorie d'equations transcendantes," C. R. Acad. Sci. Paris, 198, 1840-42.

Lahaye, E. [1948], "Solution of Systems of Transcendental Equations," Acad. Roy. Belg. Bull. Cl. Sci., 5, pp. 805-822.

Lemke, C.E. [1965], "Bimatrix Equilibrium Points and Mathematical Programming," Management Sciences, 11, pp. 681-89.

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

Meyer, G. [1968], "On Solving Nonlinear Equations with a One-Parameter Embedding," SIAM J. Numer. Anal., 5, pp. 739-52

Milnor, J. [1965], Topology from the Differentiable Viewpoint, Charlottesville: The University Press of Virginia.

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

Saigal, R. [1976a], "Fixed Point Computing Methods," Ency. of Comp. Sci. and Technology, New York: Marcel Dekker, Inc.; to be published.

Saigal, R. [1976b], "On Paths Generated by Fixed Point Algorithms," Math. of Operations Res., 1, pp. 359-80.

Scarf, H. [1967a], "The Approximation of Fixed Points of a Continuous Mapping," SIAM J. Applied Math., 15, pp. 1328-43.

Scarf, H. [1967b],"On the Computation of Equilibrium Prices," in Ten Essays in Honor of Irving Fischer, ed. Fellner, et al. pp. 207-230, New York: John Wiley and Sons.

Scarf, H. [1973]. The Computation of Economic Equilibria, New Haven: Yale University Press.

Smale, S. [1976], "A Convergent Process of Price Adjustment and Global Newton Methods," Journal of Math. Econ., 3, pp. 107-120.

Tucker, A.W. [1945], "Some Topological Properties of the Disc and Sphere," Proc. First Canadian Math. Congress, pp. 285-309.

Varian, H. [1977], "A Remark on Boundary Restrictions in the Global Newton Method," Unpublished communication.

Wilson, R. [1976], "The Bilinear Complementarity Problem and Competitive Equilibria of Linear Economic Models," Technical Report SOL 76-2, Department of Operations Research, Stanford University. To appear in Econometrica.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (14) SOL-77-25	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) (6) ON THE SOLUTION OF NONLINEAR EQUATIONS BY PATH METHODS.	5. TYPE OF REPORT & PERIOD COVERED (9) Technical Report.
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s) (10) Thomas R. Elken	8. CONTRACT OR GRANT NUMBER(s) (15) N00014-75-C-0267, N00014-75-C-0865
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Operations Research, SOL Stanford University Stanford, CA 94305	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-047-064 NR-047-143
11. CONTROLLING OFFICE NAME AND ADDRESS (11) Operations Research Program Office of Naval Research Arlington, VA 22217	12. REPORT DATE Oct [redacted] 77
	13. NUMBER OF PAGES 99 (12) 106p.
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

FIXED POINT ALGORITHMS
NONLINEAR EQUATIONS
HOMOTOPY
DIFFERENTIAL TOPOLOGY
CONTINUATION METHODS
ORIENTATION
GLOBAL NEWTON'S METHOD

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

The problem considered is that of finding a solution to a system of nonlinear equations subject to some auxiliary constraints. The methods studied here are called path methods, also referred to as "continuation" or "global Newton" methods, for solving equations. A general theory is developed which unifies the results from several papers and allows new methods to be analyzed easily. The new methods are shown to converge under more general boundary and monotonicity conditions than those assumed for the existing methods. A rigorous proof of convergence is given for an algorithm which implements a general path method.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)