COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 542

LINEAR COMPLEMENTARITY AND THE DEGREE OF MAPPINGS

by

Roger Howe

April 1980

LINEAR COMPLEMENTARITY AND THE DEGREE OF MAPPINGS*

by

Roger Howe

1. Introduction

Let M be an $n \times n$ real matrix and q an n-vector. The problem

Find n-vectors x and y such that

(1a) $$x - My = q$$

(1b) $$x \geq 0 , \quad y \geq 0$$

(1c) Either $x_i = 0$ or $y_i = 0$ for $1 \leq i \leq n$

is called the linear complementarity problem. As is explained in [C-D], several significant mathematical programming problems can be formulated as linear complementarity problems. For that reason, linear complementarity has been the subject of a considerable literature (see [K], [M], [L2], [G] and the papers cited there). For the most part in this literature the problem is treated from the algorithmic point of view, with the specification of procedures for solving the problem under various assumptions on the matrix M as a major goal. An exception to this rule is the paper [E-S] of Eaves and Scarf. In that paper, a general class of algorithms is discussed from a geometric point of view, and the linear complementarity problem is given a geometric interpretation,

*This work was partially supported by NSF Grant MCS-79-05018.

making it amenable, for certain M, to the general methods of the paper. The purpose of the present paper is to pursue the investigation of linear complementarity from a geometric point of view. Specifically, it will be shown that the topological theory of mapping degree (as exposed in, say, [G-P]) has direct bearing on the problem. No new algorithms will be proposed, but it is hoped the considerations here will lend insight into how and why existing algorithms work, and what can be expected of them.

In Section 2, the geometrization of the linear complementarity problem is reviewed. This relates the problem to a mapping of the (n-1)-sphere S^{n-1} to itself. In Section 3 we review the notion of the degree of a mapping from S^{n-1} to itself, and recite some of the basic properties of degree, in particular, its relation to the more familiar notion of index. In Section 4 we apply the degree theory to the maps coming from the linear complementarity problem. This allows us to recapture quickly many of the known results on the problem, including results of Eaves [E], of Murty [M], and of Kojima-Saigal [K-S]. In particular, degree theory immediately explains the widely noted fact that, under appropriate non-degeneracy assumptions, the parity of the number of solutions of (1) for fixed M and a variable q is constant. Also, the degree is relevant to the study of the class Q (see [C], [L2]) of matrices M such that (1) always has a solution for any vector q, since if the map associated to M has non-zero index, then automatically M is in class Q. The Section 5 studies the behavior of the degree as a function of M. It is shown that the degree can grow exponentially in n, the dimension of the problem. By contrast, most algorithms work with matrices whose associated degree is 1. In Section 6 implications of this approach for algorithms are discussed, and some natural questions left unanswered here are given.

The author would like to acknowledge the talks, given by L. van der Heyden and R. W. Cottle at the School of Organization and Management at Yale University, on linear complementarity. These were the catalyst for this paper. Conversations with Herb Scarf were also beneficial.

2. Geometrization of the Linear Complementarity Problem

The basic construction of this section may be found in [E-S]. By an orthant in $\mathbb{R}^n$, we mean the closure of a convex cone where all coordinates have a fixed sign. Thus $\mathbb{R}^{n+}$, the positive orthant, is the cone in which all coordinates are non-negative. Clearly there are 2^n orthants all together. Given a matrix M, we will construct a piecewise linear map

(2) $$P_M : \mathbb{R}^n \to \mathbb{R}^n .$$

The map P_M will be linear on each orthant. Let e^i be the i^{th} standard basis vector, and let m^i be the vector which defines the i^{th} column of M. Define a matrix M_Q by the recipe:

(3a) If x_i, the i^{th} coordinate of $x \in Q$, is ≥ 0, then the i^{th} column of M_Q is e^i,

(3b) If $x_i \leq 0$ on Q, then the i^{th} column of M_Q is m^i.

Then define P_M by:

(4) If $x \in Q$, then $P_M(x) = M_Q x$.

It is easy to check that if x belongs to more than one orthant, that is, if some coordinates of x are zero, then $M_Q x$ does not depend on which Q we might choose, so that $P_M(x)$ is well-defined. This also implies that the M_Q fit together on their common faces so that P_M is continuous. It is clearly linear on each orthant, hence piecewise linear overall. An alternative definition of P_M is as follows. For $x \in \mathbb{R}^n$, define $|x|$, the "absolute value" of x by

$$(5) \qquad |x| = (|x_1|, |x_2|, \ldots, |x_n|)$$

where $|x_i|$ denotes the usual absolute value of the i^{th} coordinate of x. It is obvious that $|x|$ is continuous and linear on each orthant. (Also, $|x| = P_{(-I)}$, where I is the identity matrix.) It is not hard to see that

$$(6) \qquad P_M(x) = \tfrac{1}{2}(x + |x|) - \tfrac{1}{2}M(|x| - x) .$$

Some pictures of P_M in the 2-dimensional case are in Figure 1.

We next observe that solving the linear complementarity problem (1) is equivalent to inverting P_M. For let x and y be vectors solving (1). Put $z = x-y$. Then $|z| = x+y$, and by comparison with (6) we find that (1) simply says $P_M(z) = q$. This establishes one direction of the first lemma. The other direction is just as easy.

<u>Lemma 2.1</u> (Eaves-Scarf). The linear complementarity problem (1) is equivalent to the problem:

Given the n-vector q, find z such that

$$(7) \qquad P_M(z) = q .$$

Since P_M is linear on orthants, it is in particular positive homogeneous of degree 1. That is,

$$P_M(\lambda x) = \lambda P_M(x)$$

for positive numbers λ. Let S^{n-1} be the unit sphere in $\mathbb{R}^n$. That is, S^{n-1} is the set of vectors x such that the Euclidean norm

$$||x|| = (\sum_{i=1}^{n} x_i^2)^{1/2}$$

equals 1. If $P_M(x) \neq 0$ whenever $x \neq 0$, we may define a map

$$\Pi_M : S^{n-1} \rightarrow S^{n-1}$$

by the formula

$$(8) \qquad \Pi_M(x) = ||P_M(x)||^{-1} P_M(x) , \quad x \in S^{n-1}$$

We will refer to the intersection of S^{n-1} with an orthant in $\mathbb{R}^n$ as an orthant in S^{n-1}. Clearly Π_M is smooth (in fact, analytic) on each orthant of S^{n-1}.

When will it be true that $P_M(x) \neq 0$ whenever $x \neq 0$? Certainly the condition

(ND) Each matrix M_Q is non-singular

will guarantee that $P_M(x)$ is zero only when x is. However, since we are only concerned with the behavior of M_Q on the orthant Q, it would in fact be enough to assume the weaker condition that

(WND) The kernel of the matrix M_Q intersects Q only in the point 0.

A matrix is non-singular if and only if its determinant does not vanish. Since M_Q has columns which are either standard basis vectors or columns of M, it is easy to see that the determinant $\det M_Q$ is the determinant of a "principal minor" submatrix. More specifically, if the coordinates $i_1, i_2, \ldots, i_k$ are the ones which are negative on Q, then

$$(9) \qquad \det M_Q = \det \begin{vmatrix} m_{i_1 i_1} & \cdots & m_{i_1 i_k} \\ \vdots & & \\ m_{i_k i_1} & \cdots & m_{i_k i_k} \end{vmatrix}$$

Therefore, we may reformulate condition (ND) as

(ND)' Each principal minor of M is non-singular.

We summarize the basic facts about Π_M in a lemma.

<u>Lemma 2.2</u>. The map Π_M on S^{n-1} can be associated to any matrix M satisfying condition (WND), which is implied by condition (ND)'. The map Π_M then depends continuously on M. Solving problem (1) is equivalent to inverting the map Π_M in the sense that a pair x,y of n-vectors satisfy (1) if and only if

$$(10) \qquad \Pi_M(z\|z\|^{-1}) = q\|q\|^{-1}$$

where $z = x-y$.

For aesthetic reasons, in a geometry-oriented article such as this one, it is desirable to have as coordinate-free formulation as is possible. In the present situation, that will not be very coordinate-free, since the orthant structure on $\mathbb{R}^n$ is compatible with relatively few coordinate systems. However, the system of orthants is preserved by:

i) Permutations of the coordinates (effected by permutation matrices).

ii) Dilations of the coordinate axes (effected by diagonal matrices).

Dilations may be further decomposed into:

ii-a) Dilations preserving all orthants (effected by positive diagonal matrices).

ii-b) Dilations preserving the set $\{\pm e_i\}_{i=1}^n$ of standard basis vectors and their negatives (effected by diagonal matrices with diagonal entries equal to ±1).

These transformations generate a group, which we might call the <u>orthant group</u>. Every element E of the orthant group can be written uniquely as a product

$$E = SD = S|D|(\operatorname{sgn} D) \tag{11}$$

where S is a permutation matrix, D is a diagonal matrix, $|D|$ is the matrix whose entries are the absolute values of the entries of D, and $\operatorname{sgn} D = |D|^{-1}D$.

We want to describe a collection of maps that are essentially the same as the maps P_M, as described by (7), but which exhibit the full symmetry of the orthant group. Let $\mathcal{P}$ be the set of continuous maps T from $\mathbb{R}^n$ to $\mathbb{R}^n$ such that the restriction of T to any orthant is (the restriction of) a linear map. Clearly $\mathcal{P}$ is a vector space under addition. Also if A is an $n \times n$ matrix, then the composition AT is in $\mathcal{P}$ if T is. By AT we mean

$$(AT)(x) = A(Tx) .$$

Although we can follow $T \varepsilon P$ by a matrix A, we cannot precede it by an arbitrary linear transformation and stay in the class P, because in general, the "breaks" in TA, where it is not linear, will occur not on the faces of the orthants, but interior to them. However, if E is in the orthant group then TE is again in P. Hence P is a vector space allowing multiplication on the left by all matrices and multiplications on the right by the orthant group.

We can parameterize elements of P by pairs of matrices L, M. For, given L,M, we can define

$$(12) \qquad T_{L,M}(x) = \left(\frac{1}{2}\right)L(x+|x|) + \frac{1}{2}M(x-|x|) .$$

Comparing with formula (6) we see that the P_M of that formula is equal to $T_{I,M}$ as defined in (12), where I is the identity matrix.

Lemma 2.3. Every element of P is represented uniquely in the form (12). Precisely, if $T \varepsilon P$, then $T = T_{L,M}$ if and only if

$$(13) \qquad T\big|_{(\mathbb{R}^n)^+} = L\big|_{(\mathbb{R}^n)^+} \qquad T\big|_{-(\mathbb{R}^n)^+} = M\big|_{-(\mathbb{R}^n)^+} .$$

If A is any $n \times n$ matrix, and S is a permutation then we have

$$(14) \qquad AT_{L,M} = T_{AL,AM} \qquad T_{L,M}S = T_{LS,MS} .$$

Proof. The formulas (14) are straightforward verifications. Also, it is clear by inspection of (12) that if $T = T_{L,M}$, then formulas (13) hold, which implies that T determines L and M. On the other hand, if L and M are the linear maps which agree with T on the positive and negative orthants respectively, then we may form $T_{L,M}$. The difference $T - T_{L,M}$ will be zero on the positive and negative orthants, and so will send the standard basis vectors e_i and their negatives to zero. But any orthant is spanned by a set consisting of certain of the e_i's and the negatives of the rest. Therefore $T - T_{L,M} = 0$, or $T = T_{L,M}$, and the lemma is proved.

We may define non-degeneracy similarly to above. By abuse of notation, given $T \in \mathcal{P}$, and an orthant Q, we will let $T_{|Q}$ stand for the linear transformation whose restriction to Q agrees with the restriction of T to Q. We will say T is <u>non-degenerate</u> if it satisfies

(ND)" For every orthant Q, $T_{|Q}$ is a non-singular matrix.

Denote the set of non-degenerate elements of $\mathcal{P}$ by ND$\mathcal{P}$. We will say T is <u>weakly non-degenerate</u> if it satisfies

(WDN)' For every Q, the intersection of Q with the kernel of $T_{|Q}$ is the single point 0.

Denote this set by WND$\mathcal{P}$. Clearly WND$\mathcal{P}$ is open in $\mathcal{P}$.

For $T \in$ WND$\mathcal{P}$, we can define a map Π_T on S^{n-1} in direct analogy with formula (8), viz.,

(15) $$\Pi_T(x) = \|T(x)\|^{-1}T(x) , \quad x \in S^{n-1} .$$

If $T = T_{L,M}$, then we will also write

$$\Pi_{T_{L,M}} = \Pi_{L,M} .$$

If $T = T_{L,M} \in \mathcal{P}$, and L is invertible, then by formula (14) we may write

(16) $$T_{L,M} = LT_{I,L^{-1}M} = LP_{L^{-1}M}$$

where P is as in (6). If $T \in ND\mathcal{P}$, then L is invertible by assumption. Thus we may conclude:

<u>Lemma 2.4</u>. Every $T \in ND\mathcal{P}$ may be written uniquely in the form

(17) $$T = AP_M$$

where A is a non-singular $n \times n$ matrix and P_M is as in (6), and M satisfies (ND)'.

From this lemma we see that, at least in the non-degenerate case, inverting the maps $T \in ND\mathcal{P}$, or their associated maps Π_T as in (15), is an essentially trivial generalization of problem (1). Hence we will feel free to discuss arbitrary maps in $\mathcal{P}$, not simply the maps P_M .

To close this section, we will discuss how some well-known transformations of M in problem (1) fit into our formulation of (1) in terms of $\mathcal{P}$. The transformations are

(18a) Conjugating M by a permutation matrix

$$S : M \to M' = SMS^{-1}$$

(18b) Pivots of M : if M is partitioned

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A is $k \times k$ and D is $(n-k) \times (n-k)$, then the <u>pivot</u> of M around A is

$$M' = \begin{bmatrix} A^{-1} & A^{-1}B \\ -CA^{-1} & D-CA^{-1}B \end{bmatrix}$$

The transformations (18a) are easy to understand. In fact, we may read off directly from formulas (14) that

$$\text{(19)} \qquad ST_{L,M}S^{-1} = T_{SLS^{-1},SMS^{-1}}$$

whence in particular

$$\text{(20)} \qquad P_{SMS^{-1}} = SP_MS^{-1} .$$

That is, the problem (1) for SMS^{-1} is just the conjugate by S of problem (1) for M . The two problems are therefore essentially equivalent. The particularly simple form of the transformation laws (19) and (20) result from the fact that S preserves the positive and negative orthants, in terms of which the coordinates $T = T_{L,M}$ were defined.

We may also conjugate elements of $\mathcal{P}$ by dilations. For positive dilations, formulas similar to (19) and (20) result, but for non-positive dilations, the transformations are more complicated because the positive orthant is not preserved. For example, let E_k be the diagonal matrix whose first k diagonal entries are -1, the rest being +1. Then we can compute that

(21) $$P_M E_k = \begin{bmatrix} A & O \\ C & I \end{bmatrix} P_{M'}$$

where M' is as in (18). Thus principal pivoting arises by multiplying P_M on the right by E_k, or some similar diagonal matrix with ±1 entries, then expressing the result as the product of $P_{M'}$ followed by a linear map.

3. Degree Theory

Consider the n-1-sphere S^{n-1}, and let

$$\phi : S^{n-1} \to S^{n-1}$$

by any continuous map. Algebraic topology allows us to attach to ϕ an integer, the degree of ϕ, written $\deg \phi$. The basic intuition about degree is that it measures the number of times ϕ wraps the sphere around itself. Thus it is a generalization to n-dimensions of the notion of winding number. See Figure 2.

The basic idea of the paper is to apply the ideas of degree theory to derive facts about the maps P_M, for M satisfying condition (WND) of Section 2 or, slightly more generally the maps T in WND$\mathcal{P}$. However degree theory in its most usual formulation does not apply directly to the maps P_M and T, but rather to the associated maps Π_T and Π_M defined on S^{n-1} by formulas (8) and (15). Thus at some point the standard results about Π_T must get transcribed into results about T. Because of the very simple relation between T and Π_T, the transcription is not difficult, but it must be done. Since we are treating degree theory as a black box in this paper, it seems simplest to just put the transcription in the black box too, and formulate the results we will use so that they apply directly to T . Thus what we present below is not only a summary, but also a slightly modified account of degree theory. For a more detailed and standard treatment of degree theory, see [G-P] or [Lf].

Let $\mu : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous map. We will say that μ is positive homogeneous of degree 1, or homogeneous for short, if

$$\text{(22)} \qquad \mu(tx) = t\mu(x) \ , \quad x \in \mathbb{R}^n \ , \quad t \in \mathbb{R} \ , \quad t > 0 \ .$$

Clearly, if μ is homogeneous, then $\mu(0) = 0$. We will say μ satisfies (WND) if $x \neq 0$ implies $\mu(x) \neq 0$. That is, if μ satisfies (WND) then μ defines a map from $\mathbb{R}^n - \{0\}$ to itself. The sphere S^{n-1} sits inside $\mathbb{R}^n - \{0\}$, and there is an obvious projection

$$\text{(23)} \qquad p : \mathbb{R}^n - \{0\} \to S^{n-1} \ ; \quad p(x) = \|x\|^{-1}x \ , \quad x \in \mathbb{R}^n - \{0\}$$

from $\mathbb{R}^n - \{0\}$ onto S^{n-1} . Clearly $p(x) = x$ for x in S^{n-1} . The projection p was implicitly used in defining the maps Π_T . We see that, in fact, given any homogeneous μ satisfying (WND), we can define

$$\Pi_\mu : S^{n-1} \to S^{n-1}$$

by

$$\text{(24)} \qquad \Pi_\mu(x) = p(\mu(x)) \ , \quad x \in S^{n-1} \ .$$

The usual degree theory would attach an integer $\deg \Pi_\mu$ to Π_μ . We will write

$$\text{(25)} \qquad \deg \mu = \deg \Pi_\mu \ .$$

In the rest of this section, μ will always be a homogeneous map satisfying (WND).

Probably the most basic property of degree is that it is a homotopy invariant. Let μ_0 and μ_1 be two homogeneous maps of $\mathbb{R}^n$ to itself satisfying (WND). Recall that the μ's are called <u>homotopic</u> if one can be continuously deformed into the other. Formally, μ_0 and μ_1 are homotopic if there is a continuous map

$$\Phi : \mathbb{R}^n \times [0,1] \to \mathbb{R}^n$$

such that for each t, $0 \leq t \leq 1$, the map $x \to \Phi(x,t)$ of $\mathbb{R}^n$ is homogeneous and satisfies (WND), and $\Phi(x,0) = \mu_0(x)$ and $\Phi(x,1) = \mu_1(x)$ for all $x \in \mathbb{R}^n$.

<u>Remark</u>. This is not the notion of homotopy between maps of $\mathbb{R}^n$, but rather the appropriate notion to guarantee the two maps Π_{μ_0} and Π_{μ_1} to be homotopic. The requirement that each $\Phi(\ ,t)$ satisfy (WND) is essential here.

(DEG1) We have $\deg \mu_0 = \deg \mu_1$ if and only if μ_0 and μ_1 are homotopic.

We will call a homogeneous map μ of $\mathbb{R}^n$ a <u>constant map</u> if μ has the form

$$\mu(x) = t(x)x_0, \quad t(x) > 0 \quad \text{for} \quad x \in \mathbb{R}^n - \{0\}$$

for some scalar-valued function $t\cdot$, and some fixed point $x_0 \neq 0$. The terminology derives from the fact that the corresponding map Π_μ of S^{n-1} will take only the value $p(x_0)$. Since S^{n-1} is arcwise connected, any two constant maps are homotopic. More generally, if μ is not surjective, i.e., if not every point of $\mathbb{R}^n$ is in the image of μ, then μ is homotopic to a constant map. This is best seen in terms of the squashed map Π_μ, which will certainly be surjective if and only if μ is. If Π_μ omits some point, say the north pole, we may shrink Π_μ to the constant map to the south pole, by pulling it southward along lines of longitude.

(DEG2) A constant map has degree zero. Hence every map of non-zero degree is surjective.

For maps that are smooth or piecewise smooth, such as the elements T of $\mathcal{P}$, there is a beautiful way to express precisely the intuition that a map of degree d covers S^{n-1} exactly d times. This involves the notion of <u>local index</u>. If μ is piecewise smooth, we call a point x in $\mathbb{R}^n$ a <u>regular point</u> of μ if i) μ is differentiable at x, and ii) the Jacobian matrix of μ is non-singular. Otherwise, we call x a singular point. Clearly whether x is regular or singular depends only on $p(x) \in S^{n-1}$, not on where x happens to lie on the ray through $p(x)$. A point $y \in \mathbb{R}^n$ is called a <u>singular value</u> of μ if it is the image under μ of some singular point. The result known as Sard's Theorem tells us that the set of singular values of μ is closed and of measure zero. (For our maps T this will be completely obvious.)

If x is a regular point of μ, we define the <u>index of μ at</u> x to be the sign of the determinant of the Jacobian matrix, $J_\mu(x)$, of μ at x. That is

$$(26) \qquad \operatorname{ind}_\mu(x) = \begin{cases} +1 & \text{if } \det J_\mu(x) > 0 \\ -1 & \text{if } \det J_\mu(x) < 0 \end{cases} .$$

For $y \varepsilon \mathbb{R}^n$, set

$$\mu^{-1}(y) = \{x \varepsilon \mathbb{R}^n : \mu(x) = y\} .$$

Observe that $\mu^{-1}(y)$ has the same cardinality as $(\Pi_\mu)^{-1}(p(y))$.

(DEG3) For any regular value y of μ , the cardinality of $\mu^{-1}(y)$ is finite, and we have the formula

$$\deg \mu = \sum_{x \varepsilon \mu^{-1}(y)} \mathrm{ind}_\mu(x) .$$

In particular, the cardinality of $\mu^{-1}(y)$ has the same parity as $\deg \mu$.

The reader may wish to verify the formula of (DEG3) for the maps of Figure 2.

The behavior of degree under composition is another very basic property. Suppose μ_1 is another homogeneous map of $\mathbb{R}^n$ satisfying (WND). It is quickly verified that the composed map $\mu_1 \circ \mu$ is again homogeneous and satisfies (WND). Also $\Pi_{\mu_1 \circ \mu} = \Pi_{\mu_1} \circ \Pi_\mu$. One has

(DEG4) $\deg(\mu_1 \circ \mu) = (\deg \mu_1)(\deg \mu)$.

Finally, especially for use in Section 5, we consider the degree of direct sums. Let μ be as usual and let ν be a homogeneous map of $\mathbb{R}^m$ satisfying (WND). We can define

$$\mu + \nu : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$$

by the obvious formula:

(27) $(\mu + \nu)(x, x') = (\mu(x), \nu(x'))$, $x \in \mathbb{R}^n$, $x' \in \mathbb{R}^m$.

It is obvious that $\mu + \nu$ is homogeneous and satisfies (WND).

(DEG5) $\deg(\mu + \nu) = (\deg \mu)(\deg \nu)$.

If μ and ν are piecewise smooth, it is not hard to deduce (DEG5) from the index formula (DEG3).

4. Applications of Degree Theory to Linear Complementarity

In this section, we apply degree theory to the maps T in WND$\mathcal{P}$, and in particular to the maps P_M of formula (6). Several results are virtually immediate. From (DEG1) and (DEG2), we can make the following conclusions:

Theorem 4.1. For $T \in$ WND$\mathcal{P}$, the integer $\deg T$ is well defined, and is constant on connected components of WND$\mathcal{P}$. If $\deg T \neq 0$, then T is surjective. In particular, for an $n \times n$ matrix M , if $\deg P_M \neq 0$, then M is a $\mathcal{Q}$ matrix.

Proof. The only thing that perhaps should be remarked is that an arc T_t , $0 \leq t \leq 1$, between T_0 and T_1 in WND$\mathcal{P}$ defines a homotopy Φ between T_0 and T_1 , by simply taking $\Phi(x,t) = T_t(x)$.

Two elements in the same component of (WND$\mathcal{P}$) will be called linearly homotopic.

Next we note that T is certainly piecewise smooth, so that (DEG3) applies. The second theorem gives some initial conclusions from (DEG3).

Theorem 4.2. For $T \in$ WND$\mathcal{P}$, and for y a regular value of T , the number of solutions of $T(x) = y$ has the same parity as $\deg T$. In particular, for a matrix M satisfying (WND) if q is a regular value of P_M, the number of solutions of (1) has the same parity as $\deg P_M$. Thus if (1) has an odd number of solutions for some open set of q's, then $\deg P_M$ is odd, and M is a Q-matrix. If M is not a Q-matrix, then (1) has an even number of solutions for all q's which are regular values of P_M.

Of course, it is quite possible for $\deg P_M$ to be even, in which case M would be a Q-matrix, but (1) would always have an even number of solutions. We will see examples in Section 5.

Another very simple consequence of (DEG3) and (DEG4), when combined with formula (16), reduces the computation of $\deg T$, at least for $T \in NDP$, to the computation of $\deg P_M$.

<u>Proposition 4.3</u>. a) For an invertible $n \times n$ matrix A, we have

$$(28) \qquad \deg A = \text{sign}(\det A) = \begin{cases} +1 & \text{if } \det A > 0 \\ -1 & \text{if } \det A < 0 \end{cases} .$$

b) For $T \in WNDP$, and E in the orthant group, we have

$$(29) \qquad \deg(ATE) = (\deg A)(\deg T)(\deg E) = \pm \deg T .$$

<u>Proof</u>. Formula (29) is immediate from formula (28) and (DEG4). Formula (28) is immediate from the formula of (DEG3), since $A^{-1}(y)$ for invertible A always consists of just one point.

A class of matrices that has figured prominently in the literature on linear complementarity is the class of <u>strictly semimonotone matrices</u> denoted L_* by Eaves [E1] and E by Lemke [L1] and Cottle [C]. This class is describable in various ways. In fact Cottle [C] gives several equivalent characterizations of strict semimonotony. For our purposes, the relevant property of semimonotone matrices is that if M is semimonotone, then problem (1) has a unique solution if q is non-negative. (This unique solution will clearly be q itself.) Again using (DEG3) we find

Proposition 4.4. If $T \in (WNDP)$, and $T^{-1}(y)$ consists of a single point for y in an open set, or for y a regular value of T, then $\deg T = \pm 1$. In particular, if M is strictly semimonotone, then $\deg P_M = 1$, and M is a $\mathcal{Q}$-matrix.

Proof. The first statement is immediate from the formula of (DEG3) (and the fact that the critical values have no interior). The sign of $\deg P_M$ for M semimonotone comes from noting that the Jacobian of P_M in the positive orthant is the identity.

Remark. The same argument shows Eaves' class L_1 gives degree 1 maps as do Garcia's $E^*(d)$, $d > 0$.

So far we have required only very crude properties of the maps T. We now look more closely at them. First, consider the regular points of T. It is clear that the interior points of orthant Q are regular points if and only if $T_{|Q}$ is non-singular. Typically, this will exhaust the set of regular points; however, it will occasionally happen that $T_{|Q_1} = T_{|Q_2}$ for two different orthants, and then some points on the supporting hyperplanes may be regular also. In any case, we see from the definition of index that for x in the interior of Q we have

$$\text{(30)} \qquad \operatorname{ind}_T(x) = \operatorname{sign} \det T_{|Q} .$$

Consider the special case when $T = P_M$ for some $n \times n$ matrix M. Let $\min_Q(M)$ denote the principal minor of M on the right hand side of formula (9). Then formula (9) tells us that for x in the interior of Q, we have

$$\text{ind}_{P_M}(x) = \text{sign}\det(\min_Q(M)) . \tag{31}$$

This formula (31) reveals the significance of the signs of the determinants of the minors for the understanding of problem (1). For $T \in NDP$, we will call the assignment of ±1 to the orthants according to formula (30) the <u>sign pattern</u> of T. It is clear that the sign pattern of T will strongly influence the properties of T. It might be hoped that the sign pattern of T would determine the degree of T. However, this is not always so, as is already seen in the 2-dimensional case. In Figure 3, which lists the sign patterns and degrees of the maps of Figure 1, we see that maps a) and c) have the same sign pattern but different degrees.

Nevertheless, some sign patterns do determine the degree of any T having them, or very nearly do so. The prime example of this in the literature (see [M], [El], [S-Th-W]) is that if the sign pattern of T consists only of +1's (or if $T = P_M$, then all principal minors of M have positive determinant, so M is a so-called P-matrix) then $\deg T = 1$; and moreover T is one-to-one (so if $T = P_M$, then (1) has always a unique solution). Another result of this sort is the result of Kojima and Saigal [K-S] who consider the case when all minors of M have negative determinant. We will see how these results can be understood and extended using degree theory.

First we consider the case of P_M with all minors of M having positive determinant. Whenever all minors of M are non-zero, we know from Section 2 that $P_M \in ND\mathcal{P}$. Of course $ND\mathcal{P}$ is a proper open subset of $WND\mathcal{P}$, so the connected components of $ND\mathcal{P}$ are contained inside connected components of $WND\mathcal{P}$. We have already agreed to call two elements in the same component of $WND\mathcal{P}$ linearly homotopic. We will call two elements in the same component of $ND\mathcal{P}$ <u>strongly linearly homotopic</u>. Clearly two strongly linearly homotopic elements will have the same sign pattern, but as the examples of Figure 3 show, the converse is not true. Similarly, those examples (e.g., a) and b), or c), d) and e)) show that elements with differing sign patterns may be linearly homotopic. In the other direction, we have the following result.

<u>Theorem 4.5</u>. Let M be an $n \times n$ matrix satisfying (ND)', so that $P_M \in ND\mathcal{P}$. Let E be a diagonal matrix with diagonal entries of ± 1. Then P_M is strongly linearly homotopic to P_E if and only if M has the same sign pattern as E. In particular P_M is strongly linearly homotopic to I, the identity map, if and only if all principal minors of M have positive determinant. In this case the degree of P_M is 1. If M is strongly linearly homotopic to an $E \neq I$, then $\deg P_M = 0$.

<u>Remark</u>. If the examples of Figure 3, example b) has the sign pattern of $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, example d) the pattern of $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$, and e) of $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$. Thus Theorem 4.5 predicts the degree of these maps. Slightly more abstractly, we see that of the 8 possible 2-dimensional sign patterns with + in the positive orthant (as will always be the case with P_M), 4 or half of them are accounted for by diagonal matrices. However with increasing dimension, the number of sign patterns of diagonal matrices

becomes insignificant compared with the total number of conceivable sign patterns: 2^n out of $2^{(2^n-1)}$. In this regard we remark that it is not clear that all possible assignments of ± 1 to the orthants are actually sign patterns of maps $T \in ND\mathcal{P}$.

Before proving Theorem 4.5, we note that it implies the uniqueness result of Murty *et al*.

Corollary 4.6 (Murty, *et al*.). If M has principal minors of positive determinant, then there is only one solution of (1) for all regular values q of P_M.

Proof. By Theorem 4.5, $\deg P_M = 1$. Plug this into the index formula of (DEG3), and observe there can be no cancellation on the right hand side because all local indices are +1.

Remark. Actually, it is known that for M with positive principal minors, system (1) has a unique solution for any q. This slightly more delicate result will follow from the local analysis of P_M to be given below.

Proof of Theorem 4.5. Define for $0 \le t \le 1$ the matrix

$$M_t = tE + (1-t)M .$$

Clearly if M_t satisfies (ND)' for all t, then P_{M_t} defines a path in $ND\mathcal{P}$ from P_M to P_E. Since E is diagonal, we have for all orthants Q

$$\det \min_Q(M_t) = (\det \min_Q E)(\det \min_Q(t + (1-t)E^{-1}(M))$$

$$= \pm\left(\sum_{k=0}^{\ell} t^{\ell-k}(1-t)^k \alpha_t\right)$$

where ℓ is the rank of $\min_Q M_t$, and α_k is the sum of the determinants of the principal minors of rank k of $\min_Q E^{-1}M$. But since E and M have the same sign pattern, and E is diagonal, the matrix $E^{-1}M$ has all principal minors positive. Hence $\det \min_Q(M_t)$ is a sum of terms all of the same sign, and thus never vanishes. Thus proves the main assertion of the theorem. As for the degrees, it is clear that $\deg P_I = \deg I = 1$. If E has some negative entries, though, we see that P_E cannot be surjective--no vectors with negative coordinates in any place where E has a -1 can be hit by P_E . Hence by Theorem 4.1, we have $\deg P_E = 0$.

We can increase the applicability of Theorem 4.5 by analyzing the "local structure" the maps P_M , and we will do this shortly. First, however, we note some other instances of strong linear homotopy. It is known and easy to show that if D is a positive diagonal matrix, then M is a $\mathcal{Q}$-matrix if and only if DM , or MD , is. Here is an analogue of that result.

Proposition 4.7. If $T = T_{L,M} \in ND\mathcal{P}$, and if D is a positive diagonal matrix, then the maps

$$T_{DL,M} ,\quad T_{L,DM} ,\quad T_{LD,M} ,\quad T_{L,MD}$$

are all strongly linearly homotopic to T .

Proof. Since the group of positive diagonal matrices is connected, it suffices to show for example, that $T_{DL,M}$ is in $(ND\mathcal{P})$ for all D . But in a given quadrant Q , the matrix of $(T_{DL,M})_Q$ is obtained from T_Q by multiplying certain columns (the ones corresponding to coordinates which are positive on Q) by the corresponding entries of D . Thus

$(T_{DL,M})_Q$ is certainly invertible if T_Q is, and the proposition follows.

Proposition 4.8. Suppose M satisfying (ND)' has the block triangular form

$$M = \begin{bmatrix} M' & N \\ 0 & M'' \end{bmatrix}$$

where M' is $k \times k$ and M'' is $(n-k) \times (n-k)$. Then P_M is strongly linearly homotopic to $P_{M'} \oplus P_{M''}$. Thus

(32) $$\deg P_M = (\deg P_{M'})(\deg P_{M''}) .$$

Proof. Formula (32) follows directly from the main conclusion and fact (DEG5). To prove the strong homotopy set

$$M_t = \begin{bmatrix} M' & tN \\ 0 & M'' \end{bmatrix} , \quad 0 \le t \le 1 .$$

It is easy to see that all principal minors of M_t will have the form

$$\begin{bmatrix} A & tC \\ 0 & B \end{bmatrix}$$

where A is a principal minor of M' and B is a principal minor of M'' . Thus all principal minors have determinants independent of t , and P_{M_t} is a path in ND^P connecting P_M with $P_{M_1} \oplus P_{M_2}$.

We now turn to the "local analysis" of maps in ND^P. By a semiorthant we mean a set obtained by specifying the sign (i.e., ≥ 0 or ≤ 0) of some of the coordinates. The semiorthant determined by specifying the sign of only one coordinate is a halfspace; and specifying signs for all the coordinates yields the orthants, which are the minimal semiorthants.

Let V_1 be the subspace of $\mathbb{R}^n$ determined by requiring certain coordinates to be zero, and the rest to be arbitrary. Let V_2 be the orthogonal subspace, so that $\mathbb{R}^n = V_1 \oplus V_2$. If Q_1 is an orthant in V_1, then $S = Q_1 \oplus V_2$ is a semiorthant, and all semiorthants can be put in this form in a unique way. We call the decomposition of S into $Q_1 + V_2$ the <u>standard decomposition</u> of S. We call V_2 the <u>spine</u> of the semiorthant S.

Given $T \in NDP$, we want to study the restriction of T to the semiorthant S. Let $S = Q_1 + V_2$ be the standard decomposition of S. It will simplify matters if we assume that T restricted to Q_1 is the identity map. We can arrange this by multiplying T on the left by a linear transformation, if necessary. Also, by permuting the coordinates by means of the orthant group, we may as well assume that V_1 is defined by setting the last $n-k$ coordinates equal to zero, and V_2 by setting the first k coordinates equal to zero. The following arguments are carried out under these assumptions.

Consider an orthant $Q \subseteq S$, and write $Q = Q_1 \oplus Q_2$ where Q_2 is an orthant of V_2. We may write T_Q in the form

$$T_Q = \begin{bmatrix} I_k & N_Q \\ 0 & T_{Q_2} \end{bmatrix} \tag{33}$$

where I_k is the $k \times k$ identity matrix, and T_{Q_2} is an $(n-k) \times (n-k)$ matrix. Since T is in NDP, the matrix T_Q is non-singular. Hence the matrix T_{Q_2} is non-singular also. It is clear that the maps T_{Q_2} fit together to define a continuous map T_2 from V_2 to itself. But then we see that in fact $T_2 \varepsilon \mathcal{P}_2$, the analogue of $\mathcal{P}$ for V_2; and since the T_{Q_2} are non-singular, we even have $T_2 \varepsilon ND\mathcal{P}_2$. We call T_2 the <u>local map around</u> Q_1 determined by T, and we call $\deg T_2$ the <u>local degree</u> of T at Q_1.

Thus, elements of $ND\mathcal{P}$ yield hereditarily elements in $ND\mathcal{P}$ of the spines of semiorthants. Note that elements of $WND\mathcal{P}$ do not have this property. For if in the discussion above T were only assumed to be in $WND\mathcal{P}$, the map T_2 might map some non-zero vectors to zero and thus fail to satisfy (WND). We remark also that if $T = P_M$ for some matrix M, and if Q_1 is some face of the positive orthant, then the local map around Q_1 is P_{M_2}, where M_2 is the obvious principal minor of M.

Another way of looking at the semiorthant $S = Q_1 + V_2$ is as the union of all orthants containing Q_1. Thus if x is any point in the relative interior of Q_1, the smallest union of orthants containing a full neighborhood of x in $\mathbb{R}^n$ is precisely the semiorthant S. A clearly interesting question about the behavior of T near x is whether $T(S)$ fills up a neighborhood of $T(x)$ or whether T "folds S over" somehow and x is left exposed on the edge of $T(S)$.

<u>Lemma 4.8</u>. Let S be a semiorthant and let $S = Q_1 + V_2$ be the standard decomposition of S. Let x be a point in the relative interior of Q_1, and let y be a point in V_2. Take $T \varepsilon ND\mathcal{P}$, and let $T_2 \varepsilon ND\mathcal{P}_2$ be the local map around Q_1 determined by T. Then for all sufficiently small scalars $s > 0$, the cardinalities of the sets $T^{-1}(x+sy) \cap S$ and $T_2^{-1}(y)$ are the same. In general $T^{-1}(x+sy) \cap S$ has fewer elements than $T_2^{-1}(y)$. Thus $T(S)$ covers a neighborhood of $T(x)$ if and only if T_2 is surjective from V_2 to V_2. Hence if $\deg T_2 \neq 0$, the image of S under T will cover a neighborhood of x.

Proof. Clearly the statements of this lemma are not affected by the normalizations used in deriving the existence of T_2 . Hence we may assume $T(x) = x$, and that for an orthant $Q \subseteq S$, the matrix T_Q has the form (33). Suppose $u \in Q$ and $T(u) = x+sy$. Write $u = u_1 + su_2$ with $u_1 \in Q_1$ and $u_2 \in Q_2$. Then we see that $T_{Q_2}(u_2) = y$, and $u_1 + sNu_2 = x$. Hence $u_2 = T_{Q_2}^{-1}(sy)$ is uniquely determined, and so is $u_1 = x - Nu_2 = x - sNT_{Q_2}^{-1}(y)$. Conversely, suppose $y = T_{Q_2}(u_2)$ for some $u_2 \in Q_2$. Since x is in the relative interior of Q_1 , the vector $u_1 = x - sNu_2$ will be in Q_1 for all sufficiently small s , so we can reverse the process and find $u = u_1 + su_2$ such that $T(u) = x+sy$. This establishes a bijection between the two sets in the first statement of the lemma. The rest of the lemma follows easily.

Our local analysis allows us to prove a substantial refinement on the uniqueness result, Corollary 4.6. Before stating it we make an observation about the sign pattern of the local maps. First, note that for a matrix A and a map $T \in NDP$, the sign pattern of AT is sign det A times the sign pattern. That is, if $\det A > 0$, then AT and T have the same sign pattern, while if $\det A < 0$, the sign pattern of AT is just the reverse of the sign pattern of T , with -1's in place of +1's and vice versa. We will call this the reverse sign pattern to the original. Since multiplying T by a linear map is geometrically not a radical thing to do to T , we would expect that T's with mutually reverse sign patterns would be similar in many ways.

The orientation of the spine V_2 of the semi-orthant S is not determined, so the sign pattern of T_2 , the local map around Q_1 , is determined only up to reversal. However, from (33) we may immediately assert:

(34) With appropriate choice of sign, the sign pattern of T_2, the local map around Q_1 determined by T, agrees with the sign pattern of T restricted to S (when one makes the obvious identification between orthants in S and orthants in V_2).

Theorem 4.9. Let $T \in NDP$ be given, and let $S \subseteq \mathbb{R}^n$ be a semiorthant. Suppose the sign pattern of T gives the same sign to all the orthants in S. Then T is injective on S. That is, no two points of S are mapped to the same point by T.

Proof. As with Lemma 4.8, the truth of this theorem is unchanged if T is composed with an invertible matrix, so we may as well assume T is normalized as in formula (33). Let $S = Q_1 + V_2$ be the standard decomposition of S, and let V_1 be the span of Q_1. Choose $y \in \mathbb{R}^n$ and write $y = y_1 + y_2$ with $y_i \in V_i$. The argument of Lemma 4.8 shows there is at most one element of $T^{-1}(y) \cap S$ for each element of $T_2^{-1}(y_2)$. Hence to prove the theorem it is enough to show that T_2 is one-to-one. But since, with appropriate normalization, T_2 has all +1's in its sign pattern, this is just the strong version of Corollary 4.6. We may finish its proof as follows. Suppose for 2 distinct points u and v in V_2 we have $T_2(u) = T_2(v)$. Let S_1' and S_2' be the smallest semiorthants in V_2 containing u and v in their respective interiors. Then the sign patterns of the local maps determined by T_2 around u and v have sign patterns all of one sign. By Theorem 4.5, and the remarks on sign patterns just above, these local maps have degree

± 1, and so are surjective. Hence T_2 maps S_1' and S_2' both onto a neighborhood of $T_2(u) = T_2(v)$. But then one sees easily that T_2 maps any neighborhood of u onto a neighborhood of $T_2(u)$ and similarly for v . Hence any point sufficiently near $T_2(u)$ has a T_2-inverse image near u and near v . Choosing such a point to be a regular value of T_2 contradicts Corollary 4.6. So T_2 must be one-to-one, and the theorem is proved.

This local version of the uniqueness theorem allows us to estimate the degrees of maps with quite varied sign patterns. We will pursue this theme at some length in Section 5. To finish this section we give one example which connects with the existing literature.

Theorem 4.10. Let $T \in NDP$ be given. Suppose the sign pattern of T assigns the same sign to all the orthants in a halfspace S . Then $|\deg T| \leq 1$. More precisely either $\deg T = 0$, or $|\deg T| = 1$, and the sign of $\deg T$ is the sign of the orthants in S .

Proof. To show this, it suffices to prove, by the index formula of (DEG3) that there is a point $y \in \mathbb{R}^n$ which is covered by T either not at all, or only once and with index the same as the orthants in S . To prove this, it is clearly enough, by Theorem 4.9, to show there is some point not in $T(-S)$ where $-S$ is the opposite halfspace to S . This is very easy, but it seems a sufficiently significant fact to state separately, so we put it into the next lemma, which will complete the proof of the theorem.

Lemma 4.11. For any $T \in P$ and any halfspace S , the image $T(S)$ is not all of $\mathbb{R}^n$.

Proof. Let $S = Q_1 + V_2$ be the standard decomposition of S. Since S is a halfspace, Q_1 is a single ray, and V_2 is a hyperplane. Choose $x \in Q_1 - \{0\}$, and consider the point $-T(x)$. Let Q be any orthant in S. Then either T_Q is singular or non-singular. If T_Q is non-singular, then $T_Q(Q)$ is a closed pointed cone containing $T(x)$, and so there is an open neighborhood of $-T(x)$ disjoint from $T_Q(Q)$. By taking an intersection, we can find a neighborhood U of $-T(x)$ disjoint from all $T_Q(Q)$ for Q such that T_Q is non-singular. But if T_Q is singular, the cone $T_Q(Q)$ is contained in a linear subspace of $\mathbb{R}^n$ of dimension less than n. So there must be points of U not contained in these $T_Q(Q)$ either, and these points are not in $T(S)$.

Corollary 4.12 (Kojima and Saigal). Given $T \in WNP$, if the sign pattern of T assigns all orthants the same sign except for one orthant which receives the opposite sign, then $|\deg T| \leq 1$. In particular, if M is a matrix all of whose principal minors have negative determinant, then $|\deg F_M| \leq 1$. Moreover, let Q be the exceptional orthant. If $\deg T = 0$, then the image of T is contained in $T(Q)$, and $T^{-1}(q)$ consists of 2 points, one in Q, one not, for $q \in T(Q)$. If $|\deg T| = 1$, then $T^{-1}(q)$ is a single point for $q \in T(Q)$, and contains 3 points if $q \in T(Q)$.

Proof. The first statement follows directly from Theorem 4.10. The statement about numbers of inverse images follows for regular values of T by plugging into the index formula of (DEG3) and noting that the only possible cancellation comes from Q. For non-regular values we may argue as in Theorem 4.9. For example, suppose $\deg T = 0$, and for some point $q \in T(Q)$, there are 2 points, u and v in $T^{-1}(q)$, and not in Q . Let S_1 and S_2 be the minimal orthants containing open neighborhoods of u and v respectively. Just as in Theorem 4.9, we can reach a contradiction if we can show at least one of the local degrees of T on S_1 or S_2 is non-zero. But if $Q \not\in S_i$, then the sign pattern of T on S_i is all one sign, so the local degree is ±1 by Theorem 4.5. But if $Q \in S_i$ for $i = 1, 2$, then we have u and v both contained in Q , contradicting our original assumption. The other cases proceed in the same way.

Remark. In the case $T = P_M$, with M having all principal minors negative, one can easily distinguish the degree 0 and the degree 1 cases. In fact, if $\deg P_M = 0$, then we say in Corollary 4.11 that $\operatorname{im} P_M \subseteq \mathbb{R}^{n+}$. In particular $M(-\mathbb{R}^{n+}) \subseteq \mathbb{R}^{n+}$, which says M is a negative matrix.

5. Degree Computations

One of the facts emerging from Section 4 is that most of the classes of matrices M which have been considered in the literature on linear complementarity give rise to maps P_M of degree 1. Thus it seems natural to wonder if in fact $|\deg T| \leq 1$ for all $T \in WND\mathcal{P}$, or at least whether $|\deg T|$ is bounded by some fairly slow growing (say polynomial) function of n (as in $\mathbb{R}^n$). In this section we will see that this is definitely not the case; contrariwise, the possible range of $|\deg T|$ grows exponentially with n . We will explicitly construct maps of very large degree. In the other direction, we will give some estimates on deg T .

It is obvious from (DEG3) that on $\mathbb{R}^n$ one has $|\deg T| \leq 2^n$, since no regular value of T can have more than 2^n inverse images, one per orthant. One can do a little better than that.

Proposition 5.1. Choose $T \in ND^P$. Let the sign pattern of T assign -1 to k orthants. Then

$$\text{(35)} \qquad \deg T < \min\left\{\frac{3}{2}k,\ \frac{1}{2}(2^n - k)\right\} .$$

In particular, all $T \in ND^P$ satisfy

$$\text{(36)} \qquad |\deg T| \leq 3\cdot 2^{n-3} - 1 \quad \text{for} \quad n \geq 3 .$$

Proof. If $k = 0$ or 1, then we know from Section 4 that $\deg T \leq 1$, so (35) clearly holds. So we may assume $k > 1$. If Q and Q' are any two orthants, we can find a halfspace H such that $Q \subseteq H$, and $Q' \subseteq -H$. Thus if $k > 1$ we can break up $\mathbb{R}^n$ into the halfspaces H and $-H$ respectively containing ℓ and ℓ' negative orthants for the sign pattern of T , with $\ell \geq 1 \leq \ell'$ and of course $\ell + \ell' = k$. If $\ell > 1$, we may further break up H into quarter spaces, each containing some negative orthants. Continuing in this fashion, we can break $\mathbb{R}^n$ up into semiorthants $S_1, S_2, \ldots, S_k$, such that the interiors of the S_j are disjoint from one another, and such that each S_j contains exactly one negative orthant. (Some S_j may consist of a single negative orthant.)

The index formula of (DEG3) says that if $\deg T = d > 0$, then each point in $\mathbb{R}^n$ (except for a set of measure zero) must be covered at least d times positively by T . Since a single orthant is mapped to a convex cone, it will cover less than half (measured say by taking the measure of the intersection of its image cone with the sphere S^{n-1}) of the points of $\mathbb{R}^n$. The estimate $\deg T < (1/2)(2^n - k)$ follows. Further, by combining Corollary 4.12, describing the local map defined by T on one of the S_j , and Lemma 4.8, we see that $T(S_j)$ can at most cover all points one time positively and a cone of points another time positively. Altogether, this allows for all points being covered less than $(3/2)k$ times positively. The combination of these two estimates yields (35). Clearly the estimate (35) yields the worst result when both of its bounds are equal, that is, when $(3/2)k = (1/2)(2^n - k)$, or $k = 2^{n-2}$. Plugging that value of k in (35), noting that $\deg T$ is an integer, and recalling that if $\deg T = d$, then $\deg AT = -d$ if A is a matrix of negative determinant, we obtain (36).

Remarks. a) For $n = 2$, estimate (35) gives $|\deg T| \leq 1$, which is clearly sharp. For $n = 3$, estimate (36) gives $|\deg T| \leq 2$, and we will see shortly that this is sharp also. For $n = 4$, estimate (36) gives $|\deg T| \leq 5$, but a more refined analysis shows in fact $|\deg T| \leq 3$. Presumably estimate (36) gets progressively worse with increasing n . The reason for this seems to be that, for $\deg T$ to be large, there need to be substantial numbers of negative orthants around to allow the folding of the positive orthants necessary for wrapping the sphere around itself many times. On the other hand, if there are many negative orthants, they begin interfering with themselves and begin to prevent the folding, or they cause too much folding, resulting in large amounts of cancellation in the index formula.

b) One can also conclude from the argument of Proposition 5.1 that when the sign pattern of T has k (-1)'s in it, there are at most $3k$ inverse images of any point, extending the Kojima-Saigal result.

We will now give some explicit examples of matrices M with $|\deg P_M| > 1$. These examples generalize one of Garcia [G].

Example 5.2. The matrix

$$M_{k,n} = \begin{bmatrix} -1 & 2 & 2 & 2 & \cdots & & & 2 \\ 2 & -1 & & & & & & \\ & & \ddots & & & & & \\ & & -1 & 2 & 2 & & & \\ 2 & & 2 & -1 & 2 & & & \\ & & 2 & 2 & 1 & & & \\ & & & & & & 2 & \\ 2 & & & & & 2 & 1 & \end{bmatrix}$$

defined by

(37) $$m_{ii} = -1 \quad \text{for} \quad 1 \le i \le k \ , \quad m_{ii} = 1 \quad \text{for} \quad k < i \le n$$

$$m_{ij} = 2 \quad \text{for} \quad i \neq j$$

yields a map $P_{M_{k,n}} = T_{k,n}$ of degree

(38) $$\deg T_{k,n} = 1-k \ .$$

In particular, $\deg T_{n,n} = 1-n$.

Proof. It will suffice, by (DEG3), to show that any point q with all positive coordinates is covered once positively and k times negatively by $T_{k,n} = T$. Such a point q is clearly covered once positively by $T(\mathbb{R}^{n+})$. Let Q_j be the orthant where all coordinates are positive except the j^{th} . For $1 \le j \le k$; we see that $M_{k,n}(-q_j e_j)$ has its j^{th} coordinate equal to q_j and its other coordinates negative. Hence

$$q = M_{k,n}(-q_j e_j) + r = T(-q_j e_j + r)$$

where r is a non-negative vector with j^{th} coordinate zero. Thus q is covered (negatively) by $T(Q_j)$ if $j \leq k$. Since all columns of $M_{k,n}$ beyond the k^{th} are positive, $T(Q_j)$ will consist of vectors with non-positive j^{th} coordinate for $j > k$.

Let N be any principal minor of $M_{k,n}$ of rank at least 2. Since the off-diagonal entries of N are positive and larger in absolute value than the diagonal entries it is easy to see that any positive linear combination of the rows of N can have at most 1 coordinate negative. Therefore, if Q is the orthant corresponding to N, $T(Q)$ cannot cover $q > 0$. Hence altogether q is covered once positively by $\mathbb{R}^{n+}$, and k times negatively by the $T(Q_j)$ for $j \leq k$ for a total degree of $1-k$, as asserted.

Once we have produced maps of degree greater than 1, the flood gates are open, for we can make the degree grow rapidly by taking direct sums.

Proposition 5.3. On $\mathbb{R}^n$, there exist elements $T \in N\mathbb{P}$ such that $\deg T \geq 2^{(2/5)n-2}$.

Proof. It suffices to exhibit a T of this index. Write $n = 5m + \ell$ with $\ell = 0, 1, 2, 3,$ or 4. We will take $T = P_M$ where

$$M = \begin{bmatrix} M_{5,5} & 0 & \cdots & 0 & 0 \\ 0 & M_{5,5} & & & \\ \vdots & \vdots & & \vdots & \\ & & & 0 & \\ 0 & & 0 & M_{5,5} & \\ 0 & & & & \tilde{M}_\ell \end{bmatrix}$$

There are m 5×5 blocks $M_{5,5}$, as in Example 5.2, plus one $\ell \times \ell$ block $\tilde{M}_\ell$ at the end. The $\tilde{M}_\ell$ block is the identity if $\ell = 0$ or 1, and is $M_{\ell,\ell}$ of Example 5.2 if $\ell = 2,3,$ or 4. According to (DEG5) and formula (38) we have

$$\deg P_M = 4^m \max(1, \ell-1) = 2^{2m}\max(1, \ell-1) \geq 2^{(2n/5)-2}$$

as claimed.

By a continuity argument, one can see that if there is a $T \in ND^P$ of degree $d > 0$, there is another T of any positive degree less than d . Thus the possibilities for $\deg T$ grow exponentially with n . This would seem to make it unlikely that there is any simple method of computing $\deg T$ in all cases.

6. Remarks on Algorithms

From the geometric viewpoint adopted here, two classes of algorithms occur as candidates for solving problem (1): path-following algorithms and homotopy algorithms. In the former type, one would start with points x and y for which one knew $P_M(x) = y$. (Taking $x = y$ in $\mathbb{R}^{n+}$ is an obvious choice.) Then one would draw some path from y to q . (There are many possibilities, the straight line being the obvious first choice.) One then would attempt to "lift" this path via P_M^{-1} , and follow it along until one arrives at $P_M^{-1}(q)$. The path following algorithm with the obvious choices of starting point and path suggested above essentially amounts to the Lemke-type algorithms of [L], [E1], and [G]. It seems that the algorithm of van der Heyden [vH] is another path-following algorithm, but following a broken line from y to q with individual segments parallel to the coordinate axes.

The only problem that path-following algorithms could run into is the following. Let γ be the path from y to q, and let $\tilde{\gamma}$ be the lift of γ via P_M^{-1}, beginning at x. As we proceed along γ, we might come to the interface between two orthants which have opposed signs in the sign pattern of P_M. When that happens, proceeding forward along $\tilde{\gamma}$ corresponds via P_M to reversing direction and proceeding backwards along γ. Thus it might happen that after proceeding along $\tilde{\gamma}$ for a while we might return to $P_M^{-1}(y)$ instead of arriving at $P_M^{-1}(q)$. Thus for example in Figure 2 if we start at x and head for the inverse image of q, we will not reach it, but will return to z lying over y again. Of course we can persevere, and in the 1-dimensional case we eventually will reach a point over q. But in higher dimensions, one can be caught on a closed loop that will cycle and cycle and never find q.

One way to be assured that this problem will not occur is to start at a point y such that $P_M^{-1}(y)$ consists of a single point x. Thus we see the significance of the Lemke class L_1, for which all strictly positive q permit only one solution to (1) and the more refined Garcia classes [G] which for some positive q (hence all nearby q) there is only one solution to (1). For these classes path-following techniques will work. However, assuming a single inverse for a regular value of P_M implies $|\deg P_M| = 1$, so to guarantee success of the path-following technique one must restrict oneself to maps of degree ± 1. This explains our empirical observation in Section 4 that most of the classes of matrices considered in connection with (1) did give maps of degree 1.

The homotopy algorithms are not so transparently constructed, and much less prevalent in the literature. In fact the paper of Eaves and Scarf [E-S] proposes the only homotopy algorithm known to the author. The basic idea is to construct a homotopy from the given P_M to a well-understood map P_{M_0}, solve the problem (1) for M_0, then follow the homotopy to arrive at a solution of (1) for M. This could also be looked on as a path-following technique, but the path is a path of maps rather than of points. Unlike the regular path-following technique, the homotopy method poses some conceptual problems at the outset, namely what standard model M_0 to choose, and how to define the homotopy. About the only map that immediately strikes one as "well-understood" is the identity map P_I, or some reasonably mild perturbation of it. But choosing $I = M_0$ immediately limits one to maps of degree 1 since homotopy preserves degree. Furthermore, choosing M_0 still leaves one with a second non-obvious problem, namely how to perform the homotopy.

Despite these difficulties, the homotopy method is essentially of the same power as the path-following method. That is, for the class of maps for which a path-following technique will certainly work, namely those for which some regular value is assumed only once, one can specify a homotopy to a standard map. We describe how to do this.

Without essential loss of generality, let us assume that for the vector $q_0 = (1, 1, \ldots, 1)$ there is only the one solution $x = q_0$, $y = 0$ to (1). Instead of P_I, we will use as our basic mapping P_{M_0}, where

$$M_0 = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \\ 1 & \cdots & 1 \end{bmatrix}$$

The reader may convince himself that, indeed P_{M_0} is relatively easy to understand, and $P_{M_0}^{-1}(q)$ consists of a single point for almost all q. For a homotopy we will use the naive one P_{M_t} where

$$(39) \qquad M_t = (1-t)M_0 + tM \, , \quad 0 \le t \le 1 \, .$$

The salient observation then is

Lemma 6.1. Suppose $P_M \in WND^P$, and for $q = q_0$, the system (1) has only the solution $x = q_0$, $y = 0$. Then M_t, as in (39) is in WND^P for all t in $[0,1]$, so that P_{M_t} does constitute a homotopy in WND^P from P_M to P_{M_0}.

Proof. We must show that M_t satisfies the condition (WND) of Section 2 for all t. Since q_0 and M_0 are invariant under permutations of coordinates, it will be enough to check (WND) for quadrants Q_k, in which the first k coordinates are positive and the rest are negative. For Q_k, we have (using the notation of (WND))

$$(M_t)_{Q_k} = \left[\begin{array}{c|c|c|c} I_k & & & \\ \cline{1-1} & (M_t)_{k+1} & \cdots & (M_t)_n \\ 0 & & & \end{array} \right]$$

where I_k is the $k \times k$ identity matrix and $(M_t)_j$ is the j^{th} column of M_t. To verify (WND) for Q_k, we must check that the dependence relation

$$(40) \qquad \sum_{i=1}^{k} a_i e_i = \sum_{j=k+1}^{n} b_j (M_t)_j$$

has no solution with non-negative a's and b's not all zero. Suppose (40) does have a solution. Write $(M_t)_j = (1-t)q_0 + tM_j$. Then (40) becomes

$$(41) \qquad \sum_{i=1}^{k} a_i e_i = \sum_{j=k+1}^{n} b_j(1-t)q_0 + \sum_{j=k+1}^{n} b_j M_j ,$$

or

$$(42) \qquad \sum_{i=1}^{k} a_i e_i - \sum_{j=1}^{n} b_j M_j = (\sum_{j=k+1}^{n} b_j(1-t))q_0 .$$

But (42) just says that the vectors $x = (a_1, \ldots, a_k, 0, \ldots, 0)$ and $y = (0, 0, \ldots, 0, b_{k+1}, \ldots, b_n)$ are a non-trivial solution to (1) for $q = sq_0$, where $s = (\sum_{j=k+1}^{n} b_j(1-t)) > 0$. But this is impossible by our assumption on M . Hence (40) has no solution and M_t is in (WND) for all t in $[0,1]$. This finishes the lemma.

Thus we see that all algorithms proposed so far are more or less equally capable of solving problem (1), and that they can do so under the condition that some positive vector have a unique inverse image under P_M , a condition that entails that $\deg P_M = 1$. We will close the paper with some questions that suggested by this conclusion and the other results of the paper.

1. Is it possible to reasonably characterize the class $\mathcal{P}_1 = \{T \in \text{WND}\mathcal{P} : \deg T = 1\}$?

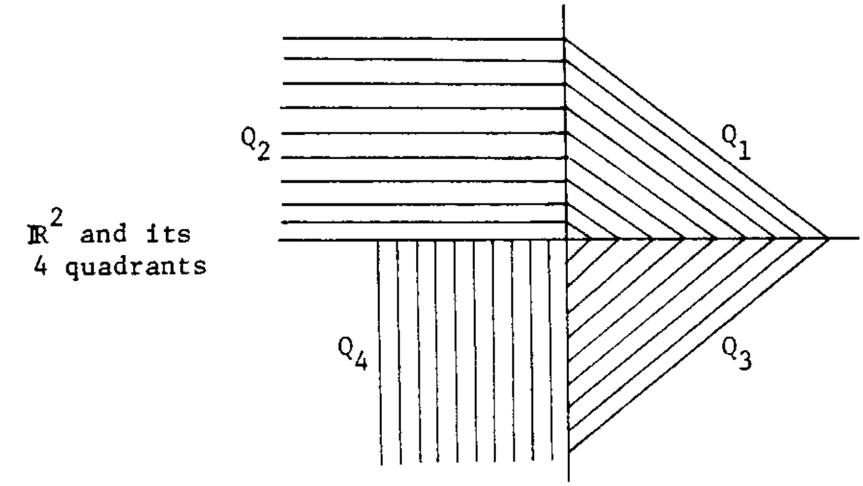
2. In particular, is $\mathcal{P}_1$ connected, so that every element of $\mathcal{P}_1$ can be deformed in WND$\mathcal{P}$ to the identity map?

3. Is it conceivable that for every member T of $\mathcal{P}_1$ there is some regular value of T with only one inverse image, so that the known algorithms essentially apply to all of $\mathcal{P}_1$?

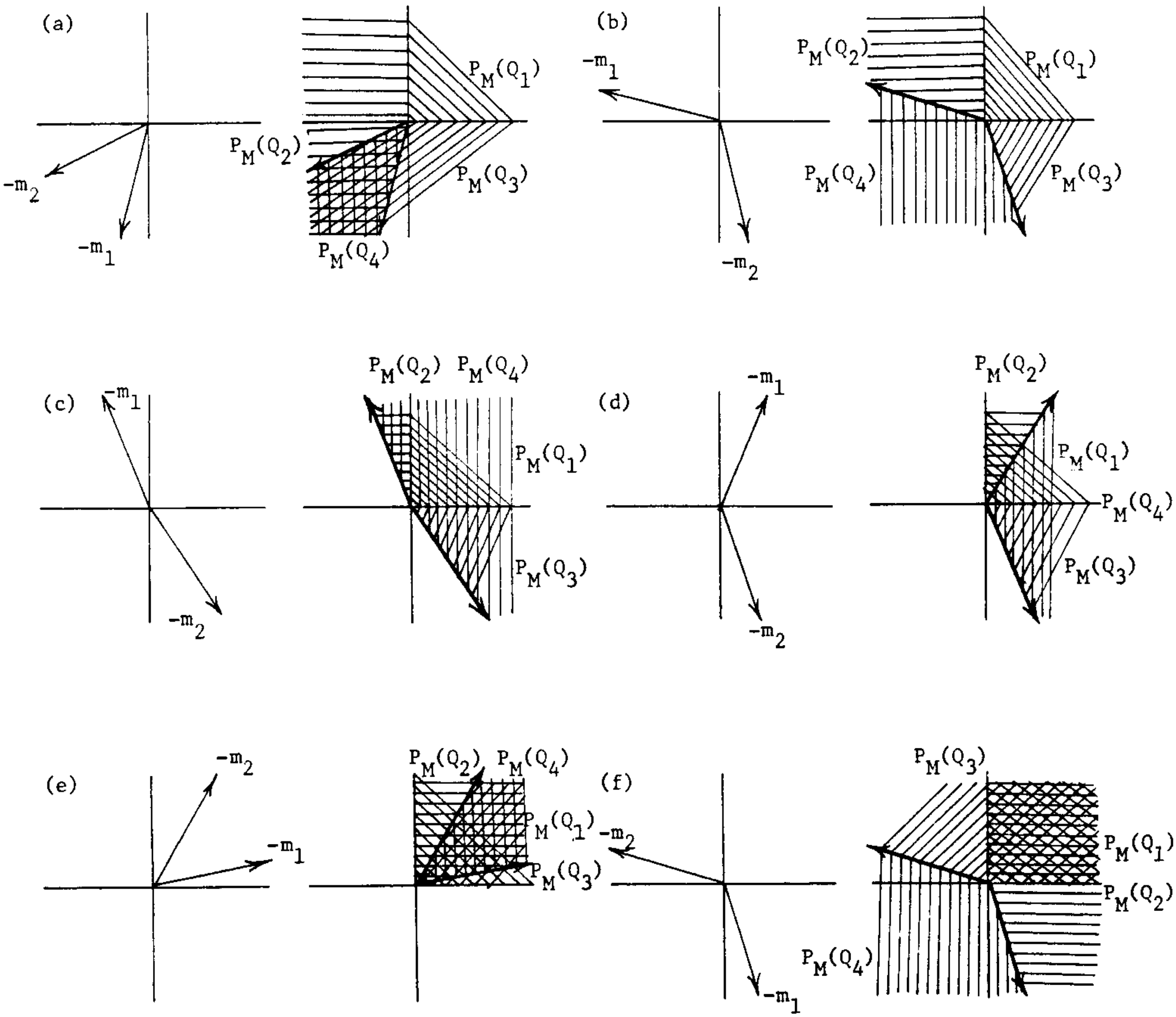
4. In a similar vein, is $\deg T \neq 0$ a necessary as well as a sufficient condition for $T \in \mathrm{WND}\mathcal{P}$ to be surjective? That is, if M is a $\mathcal{Q}$-matrix, and M satisfies (WND), is $\deg P_M \neq 0$?

5. Are there any naturally occurring problems of type (1) for which $|\deg P_M| > 1$? If so, what can be done about the higher degree case?

FIGURE 1



$\mathbb{R}^2$ and its
4 quadrants
Q_1
Q_2
Q_3
Q_4



(a)
$-m_2$
$-m_1$
$P_M(Q_1)$
$P_M(Q_2)$
$P_M(Q_3)$
$P_M(Q_4)$
(b)
$-m_1$
$-m_2$
$P_M(Q_2)$
$P_M(Q_1)$
$P_M(Q_4)$
$P_M(Q_3)$
(c)
$-m_1$
$-m_2$
$P_M(Q_2)$ $P_M(Q_4)$
$P_M(Q_1)$
$P_M(Q_3)$
(d)
$-m_1$
$-m_2$
$P_M(Q_2)$
$P_M(Q_1)$
$P_M(Q_4)$
$P_M(Q_3)$
(e)
$-m_2$
$-m_1$
$P_M(Q_2)$ $P_M(Q_4)$
$P_M(Q_1)$
$P_M(Q_3)$
(f)
$-m_2$
$-m_1$
$P_M(Q_3)$
$P_M(Q_1)$
$P_M(Q_2)$
$P_M(Q_4)$

FIGURE 2. Examples of maps of S^1, the circle to itself. Actually, to promote visualization, we have drawn curves in the plane. These may be imagined to be the images of maps from the circle. To obtain maps from the circle to itself, simply project radially. The arrows → indicate the direction of traversal.

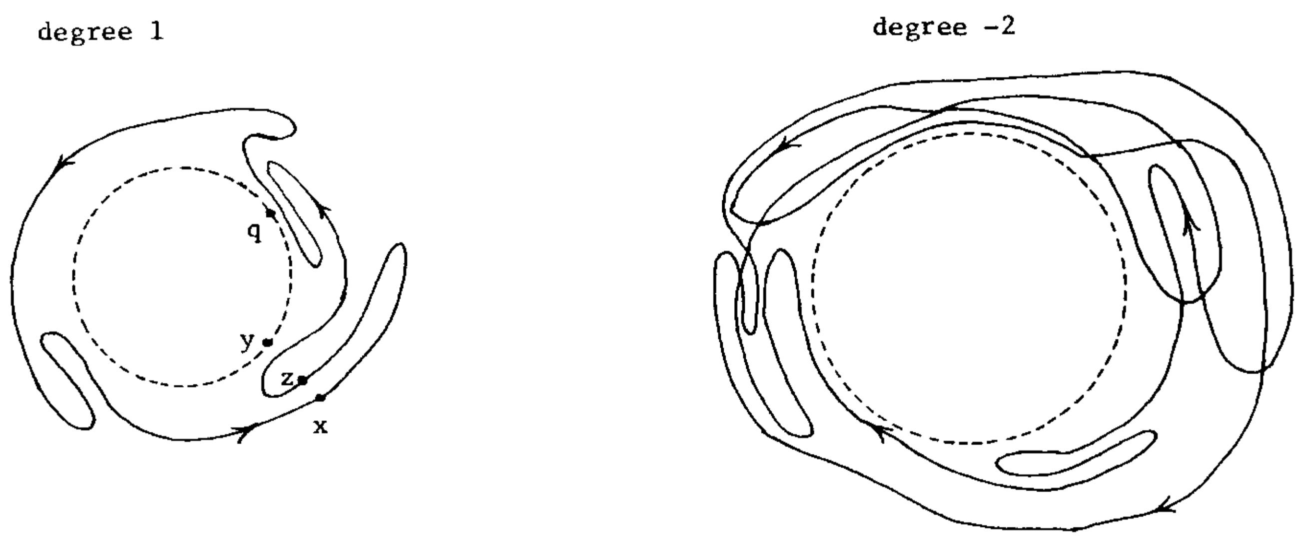


FIGURE 3. The sign patterns and degrees of the maps of Figure 1. The degree of P_M in each orthant is indicated by a + or – sign in that orthant.

a) degree 1

+	+
–	+

b) degree 1

+	+
+	+

c) degree 0

+	+
–	+

d) degree 0

–	+
–	+

e) degree 0

–	+
+	–

f) degree –1

–	+
–	–

REFERENCES

[C] R. Cottle, Completely-Q Matrices, Stanford University Department of Operations Research Technical Report 79-12, Sept. 1979.

[C-D] R. Cottle and G. Dantzig, Complementary Pivot Theory of Mathematical Programming, Lin. Alg. and its Applications 1 (1968), pp. 103-125.

[E] B. C. Eaves, The Linear Complementarity Problem, Management Sci. 17 (1971), pp. 612-634.

[E-S] B. C. Eaves and H. Scarf, The Solution of Systems of Piecewise Linear Equations, Math. Op. Res. 1, No. 1 (1976), pp. 1-27.

[G] C. Garcia, Some Classes of Matrices in Linear Complementarity Theory, Math. Programming 5, No. 3 (1973), pp. 299-310.

[G-P] V. Guillemin and A. Pollack, <u>Differential Topology</u>, Prentice Hall (1974), Engelwood Cliffs, N.J.

[vH] L. van der Heyden, A Variable Dimension Algorithm for the Linear Complementarity Problem, J. F. Kennedy School Discussion Paper Series, Number 67D, June 1979.

[K] S. Karamardian, The Complementarity Problem, Math. Programming 2 (1972), pp. 107-129.

[M] K. G. Murty, On the Number of Solutions to Complementarity Problems and the Spanning Properties of Complementary Cones, Lin. Alg. and Applic. 5 (1972), pp. 65-108.

[K-S] M. Kojima and R. Saigal, On the Number of Solutions to a Class of Linear Complementarity Problems, Math. Programming 17 (1979), pp. 136-139.

[Lf] S. Lefschetz, <u>Introduction to Topology</u>, Princeton University Press, Princeton, N.J., 1949.

[L1] C. Lemke, Bimatrix Equilibrium Points and Mathematical Programming, Management Sci. 11, No. 7 (1965), pp. 681-689.

[L2] C. Lemke, Recent Results on Complementarity Problems, in <u>Nonlinear Programming</u>, Eds. J. Rosen, O. Mangasarian, and K. Ritter, Academic Press, New York, 1970.

[S-Th-W] H. Samuelson, R. M. Thrall, and O. Wesler, A Partition Theorem for Euclidean n-space, Proc. Amer. Math. Soc. 9 (1958), pp. 805-807.