SIAM J. MATRIX ANAL. APPL.
Vol. 10, No. 2, pp. 259-277, April 1989

011

CONTINUOUS HOMOTOPIES FOR THE LINEAR COMPLEMENTARITY PROBLEM*

LAYNE T. WATSON†, J. PATRICK BIXLER†, AND AUBREY B. POORE‡

Abstract. There are various formulations of the linear complementarity problem as a Kakutani fixed point problem, a constrained optimization, or a nonlinear system of equations. These formulations have remained a curiosity since not many people seriously thought that a linear combinatorial problem should be converted to a nonlinear problem. Recent advances in homotopy theory and new mathematical software capabilities such as HOMPACK indicate that continuous nonlinear formulations of linear and combinatorial problems may not be farfetched. Several different types of continuous homotopies for the linear complementarity problem are presented and analyzed here, with some numerical results. The homotopies with the best theoretical properties (global convergence and no singularities along the zero curve) turn out to also be the best in practice.

Key words. homotopy algorithm, globally convergent, linear complementarity problem, fixed point, expanded Lagrangian, nonlinear equations

AMS(MOS) subject classifications. 65H10, 65L10, 65L60

1. Introduction. Given a real $n \times n$ matrix M and a real n-vector q, the linear complementarity problem (LCP), denoted by (q, M), is to find n-vectors w and z such that

$$w - Mz = q,$$
$$w \geqq 0, \quad z \geqq 0, \quad w^t z = 0.$$

The constraint $w^t z = 0$ is called the complementarity condition since for any i, $1 \leqq i \leqq n$, $z_i = 0$ if $w_i \neq 0$, and vice versa. A solution where some $z_i = w_i = 0$ is called degenerate. The linear complementarity problem arises in such diverse areas as economic modeling [15], [16], [59]; bimatrix games [29], [32]; mathematical programming [10], [19], [34]; mechanics [17]; lubrication [28]; and numerical analysis [9].

There are numerous algorithms for solving special classes of linear complementarity problems. Those based on pivoting or simplex-type processes include Lemke's complementary pivot algorithm [29]; Cottle and Dantzig's principal pivot method [6]; Bard-type algorithms [4], [45], [60]; and the n-cycle algorithm [62], [64]. There are also linear iterative techniques, similar to those for solving linear systems of equations, such as SOR [2], [3], [8], [35], [50], [51], [61] and various related fixed point iteration schemes. A very different algorithm is the simplicial homotopy algorithm of Merrill [37], applied to a Kakutani fixed point formulation (solution is a fixed point of a point-to-set mapping) of the linear complementarity problem.

* Received by the editors December 23, 1987; accepted for publication (in revised form) October 21, 1988.

† Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061. The work of L. T. Watson was supported in part by Air Force Office of Scientific Research grant 85-0250.

‡ Department of Mathematics, Colorado State University, Fort Collins, Colorado 80523. The work of this author was supported in part by National Science Foundation grant DMS-8704679.

A more recent development was the formulation of the linear complementarity problem as a differentiable nonlinear system of equations [33], and the solution of this system of equations by a globally convergent homotopy method [66]. This approach has remained a curiosity because few people took seriously the formulation of a linear combinatorial problem (like the LCP) as a highly nonlinear problem. Recent advances in homotopy theory and mathematical software for nonlinear systems of equations [68]–[69], and new nonlinear formulations of linear, discrete, and combinatorial problems ([33], [53], [54], [66], [67]) suggest that nonlinear formulations of the linear complementarity problem should be investigated further.

The present paper proposes and analyzes several nonlinear homotopies for the linear complementarity problem. The existence theorems implied by the globally convergent homotopy theorems are as general as any derived by other methods. Section 2 defines some terminology, §§3–9 describe and analyze different homotopy maps, §10 describes some numerical experiments, and §11 summarizes.

2. Preliminaries. In this section we gather some terms and fundamental results about globally convergent homotopy methods. For additional background refer to [65], [68].

Let E^n denote n-dimensional, real Euclidean space and let $E^{n\times n}$ be the set of real $n \times n$ matrices. The ith component of a vector $v \in E^n$ is denoted by v_i, and for a matrix $A \in E^{n\times n}$, $A_{i\cdot}$ denotes the ith row and $A_{\cdot j}$ denotes the jth column. For subsets $\emptyset \neq I, J \subset \{1, \cdots, n\}$, A_{IJ} denotes the submatrix of A with rows indexed by I and columns indexed by J. Let $e \in E^n$ be the vector such that $e_i = 1$ for all i. For $v \in E^n$, $v+$ denotes the vector with components $(v+)_i = \max\{0, v_i\}$, and $v-$ denotes the vector with components $(v-)_i = \max\{0, -v_i\}$. The *support* of v, denoted by $S(v)$, is simply $\{i \mid v_i \neq 0\}$. We use the following notation when comparing a vector $a \in E^n$ to 0:

$$\begin{aligned} a &\geqq 0 \quad \text{if } a_i \geqq 0 \text{ for all } i,\\ a &\geq 0 \quad \text{if } a \geqq 0 \text{ and } a \neq 0,\\ a &> 0 \quad \text{if } a_i > 0 \text{ for all } i. \end{aligned}$$

Let $M \in E^{n\times n}$ be a real $n \times n$ matrix and let q be a real n-vector. M is *nonnegative* if each element of M is nonnegative, *copositive* if $x^tMx \geqq 0$ for all $x \geqq 0$, and *strictly copositive* if $x^tMx > 0$ for all $x \geq 0$. M is called *nondegenerate* if all of its principal minors are nonzero, and a *P-matrix* if all of its principal minors are positive. The vector q is *nondegenerate* with respect to M if q is not a linear combination of any $n-1$ columns of $(I, -M)$. Finally, M is *strictly semimonotone* if for each vector $x \geq 0$ there exists an index k such that $x_k(Mx)_k > 0$.

When $w \geqq 0$ and $z \geqq 0$ satisfy $w - Mz = q$, (w, z) is called a *feasible solution*. If $w^tz = 0$ also, (w, z) is called a *complementary feasible solution*.

A C^2 (twice continuously differentiable) function $F : E^n \to E^m$ is said to be *transversal to zero* if the $m \times n$ Jacobian matrix $DF(x)$ has rank m on $F^{-1}(0)$. The theoretical justification for modern probability-one homotopy methods rests on a result from differential geometry, known as a parameterized Sard's theorem [65]:

LEMMA 2.1. *Let $\rho : E^m \times [0,1) \times E^n \to E^n$ be a C^2 map which is transversal to zero, and define*

$$\rho_a(\lambda, z) = \rho(a, \lambda, z).$$

Then for almost all $a \in E^m$, the map ρ_a is also transversal to zero.

The significance of Lemma 2.1 is partially given by:

LEMMA 2.2. *In addition to the hypotheses of Lemma 2.1, suppose that for each $a \in E^m$ the system $\rho_a(0, z) = 0$ has a unique solution $z^{(0)}$. Then for almost all $a \in E^m$ there is a smooth zero curve $\gamma \subset [0,1) \times E^n$ of $\rho_a(\lambda, z)$, emanating from $(0, z^{(0)})$, along which the Jacobian matrix $D\rho_a(\lambda, z)$ has rank n. γ does not intersect itself or any other zero curves of ρ_a, does not bifurcate, has finite arc length in any compact subset of $[0,1) \times E^n$, and either goes to infinity or reaches the hyperplane $\lambda = 1$.*

LEMMA 2.3. *Under the hypotheses of Lemma 2.2, if the zero curve γ is bounded, then it has an accumulation point $(1, \bar{z})$. Furthermore, if rank $D\rho_a(1, \bar{z}) = n$, then γ has finite arc length.*

Conceptually, the algorithm for solving the nonlinear system of equations $F(z) = 0$ is simple. Using the lemmas above, just follow the zero curve γ, starting from some point $(0, z^{(0)})$ and ending at a point $(1, \bar{z})$, where $\bar{z}$ is a zero of $F(z)$. Computationally this may be nontrivial, but at least the idea is clear. A typical simple choice for the homotopy map is

$$\rho_a(\lambda, z) = \lambda F(z) + (1 - \lambda)(z - a).$$

Although this homotopy map has the same form as a standard continuation or embedding mapping, there are two important differences. First, in standard continuation the embedding parameter λ increases monotonically from 0 to 1 as the trivial problem $(z - a) = 0$ is continuously deformed to the given problem $F(z) = 0$. With the present homotopy method, turning points on γ cause no special difficulties and so λ can increase and decrease as the curve is being tracked. Secondly, the fact that the Jacobian matrix $D\rho_a$ has full rank along γ and the way in which the zero curve is tracked guarantee that there are never any "singular points" which afflict standard continuation methods.

3. The 1979 homotopy. To provide a backdrop for the homotopies presented in the next few sections, we briefly review the homotopy map of [66]. Mangasarian [33] has shown that the linear complementarity problem (q, M) can be reformulated as a zero finding problem

$$H(z) = 0,$$

where $H(z)$ can be made as smooth as desired. Taking $\theta(t) = t^3$ in Mangasarian's Theorem 1 [33], we define $H(z)$ by

$$H_i(z) = -|M_{i\cdot}z + q_i - z_i|^3 + (M_{i\cdot}z + q_i)^3 + {z_i}^3$$

and

$$\rho_a(\lambda, z) = \lambda H(z) + (1 - \lambda)(z - a).$$

By noting the signs of each term in H, it is clear that $z \geqq 0$, $Mz + q \geqq 0$, and $(Mz + q)^t z = 0$ if and only if $H(z) = 0$. That is to say, z solves the LCP if and only if $H(z) = 0$. The following result from [66] gives conditions on the matrix M to insure that a zero curve of the homotopy map ρ_a can be tracked to obtain a zero of H.

THEOREM 3.1. *Let $M \in E^{n\times n}$ be either positive definite, a P-matrix, nondegenerate strictly copositive, or nondegenerate strictly semimonotone, and let $q \in E^n$ be nondegenerate with respect to M. Then there exists $\delta > 0$ such that for almost all $a \geqq 0$ with $\|a\|_\infty < \delta$ there is a zero curve γ of $\rho_a(\lambda, z)$, along which $D\rho_a(\lambda, z)$ has full rank, having finite arc length and connecting $(0, a)$ to $(1, \bar{z})$, where $\bar{z}$ is a zero of $H(z)$.*

Although it was not stated in [66], the proof of Theorem 3.1 there showed that if the nondegeneracy assumptions are removed, then the conclusion still holds, except that $(1, \bar{z})$ is only an accumulation point of the zero curve γ (of possibly unbounded variation). The map ρ_a above is the standard homotopy map. In the context of this paper we can view it as relaxing all of the solution requirements of the LCP while the zero curve is being tracked. Initially z is set to some arbitrary point a having nothing to do with the solution to (q, M). As λ gets closer to 1 we can say that, in some sense, z gets closer to such a solution. However, for any $\lambda < 1$, z and $w = Mz + q$ do not necessarily form a feasible solution or a complementary solution to (q, M). These conditions are imposed only at the end, when $\lambda = 1$, and then all at once. In the next few sections we present several homotopies, based on Mangasarian's function, that attempt to maintain at least feasibility or complementarity for a modified LCP right from the start. The hope is that the homotopy process is then more efficient.

4. Relaxation of M. In this map, all of the continuation is applied to the matrix. We maintain a complementary feasible solution for some other matrix which is a convex combination of M and the identity. When $\lambda = 0$ the matrix is the identity, and when $\lambda = 1$ the matrix is M. We can view this map as relaxing only the matrix M as the zero curve is being tracked.

Define $\Lambda : [0, 1) \times E^n \to E^n$ by

$$\Lambda_i(\lambda, z) = -\left|[(1-\lambda)I + \lambda M]_{i\cdot} z + q_i - z_i\right|^3 + \left([(1-\lambda)I + \lambda M]_{i\cdot} z + q_i\right)^3 + z_i^3$$

for $i = 1, \cdots, n$.

Observe that since this is simply Mangasarian's map with a modified matrix for M, feasibility and complementarity are preserved wherever Λ is zero.

LEMMA 4.1. *Let P be any of the following properties:*

(a) *positive definite,*

(b) *P-matrix,*

(c) *nondegenerate strictly copositive,*

(d) *nondegenerate strictly semimonotone,*

and let $0 \leqq \lambda \leqq 1$. If a matrix $M \in E^{n\times n}$ has property P, then $(1-\lambda)I + \lambda M$ also has property P except possibly for finitely many values of λ.

Proof. (a) It follows from the definition of positive definite that

$$x^t[(1-\lambda)I + \lambda M]x = (1-\lambda)(x^t x) + \lambda(x^t M x) > 0$$

for all $x \neq 0$ whenever M is positive definite.

(b) It can be shown [13] that M is a P-matrix if and only if for all $x \neq 0$ there is an index k such that $x_k(Mx)_k > 0$. Let M be a P-matrix and let $x \neq 0$. Then

$$\begin{aligned} x_k\big([(1-\lambda)I + \lambda M]x\big)_k &= (1-\lambda)x_k(Ix)_k + \lambda x_k(Mx)_k \\ &= (1-\lambda)x_k^2 + \lambda x_k(Mx)_k \\ &> 0 \end{aligned}$$

for some index k.

Let M be nondegenerate and let $K \subset \{1, \cdots, n\}$. Because the determinant is multilinear we have

$$\det\big((1-\lambda)I + \lambda M\big)_{KK} = \sum_{J \subset K} (1-\lambda)^{|K|-|J|}\lambda^{|J|} \det M_{JJ},$$

which is simply a polynomial in λ. Notice that, since $\{(1-\lambda)^{k-j}\lambda^j \mid 0 \leqq j \leqq k\}$ forms a linearly independent set of polynomials, and $\det M_{JJ} \neq 0$ for any subset $J \subset \{1, \cdots, n\}$, this polynomial is not identically zero. (By convention, $\det M_{\phi\phi} = 1$.) This polynomial has only a finite number of zeros and so $(1-\lambda)I + \lambda M$ is nondegenerate except for finitely many values of λ.

(c) It follows from the definition of strictly copositive that

$$x^t[(1-\lambda)I + \lambda M]x = (1-\lambda)(x^t x) + \lambda(x^t M x) > 0$$

for all $x \geq 0$ whenever M is strictly copositive.

(d) An argument similar to that for (b) holds if M is strictly semimonotone and $x \geq 0$. □

Lemma 4.1, Theorem 3.1, and the subsequent remark give us the following theorem.

THEOREM 4.1. *Let $M \in E^{n\times n}$ be positive definite or a P-matrix, and let $q \in E^n$. Then there exists a zero curve γ of Λ emanating from $(0, q-)$ and reaching a point $(1, \bar{z})$, where $\bar{z}$ solves the* LCP (q, M).

Note that Theorem 4.1 does not include strictly copositive or strictly semimonotone matrices, nor any reference to the rank of the Jacobian matrix along the zero curve γ. If M is nondegenerate strictly copositive or nondegenerate strictly semimonotone, there is a solution to the LCP $\big(q, [(1-\lambda)I + \lambda M]\big)$ for every $\lambda \in [0, 1]$ by Theorem 3.1. However, there may be multiple solutions, and when the number of solutions changes at some $\bar{\lambda}$ some of the zero curves of Λ either "stop" or "start" at $\bar{\lambda}$. Thus there is no guarantee that a *single* zero curve of Λ will reach all the way from $\lambda = 0$ to $\lambda = 1$. For example, take

$$M = \begin{pmatrix} 3 & 2 & 10 \\ 0 & 1 & 10 \\ 1 & 0 & 1 \end{pmatrix}, \qquad q = \begin{pmatrix} -5.3 \\ -4.0 \\ -0.9 \end{pmatrix}.$$

M is nondegenerate strictly semimonotone, but the zero curve emanating from $(0, q-)$ disappears at $\lambda = 0.8$. Also we cannot say that the Jacobian matrix $D\Lambda(\lambda, z)$ is nonsingular along the entire zero curve γ. The ith row of the Jacobian matrix of Λ is

$$\begin{aligned}\big(D\Lambda(\lambda,z)\big)_{i\cdot} = \big(&-3|A|(A)(-I+M)_{i\cdot}z + 3(B)^2(-I+M)_{i\cdot}z,\\ &-3|A|(A)(\lambda m_{i1}) + 3(B)^2(\lambda m_{i1}), \cdots,\\ &-3|A|(A)(-\lambda+\lambda m_{ii}) + 3(B)^2(1-\lambda+\lambda m_{ii}) + 3z_i^2, \cdots,\\ &-3|A|(A)(\lambda m_{in}) + 3(B)^2(\lambda m_{in})\big),\end{aligned}$$

$$\begin{aligned}\text{where } A &= [(1-\lambda)I + \lambda M]_{i\cdot}z + q_i - z_i,\\ B &= [(1-\lambda)I + \lambda M]_{i\cdot}z + q_i.\end{aligned}$$

Observe that if $|q| > 0$, then rank $D\Lambda(0, q-) = n$, and so the starting point $z = q-$ for the zero curve is nonsingular.

PROPOSITION 4.2. *Let $M \in E^{n\times n}$ be positive definite or a P-matrix, and let $q \in E^n$. Whenever M and q are such that $(\bar{w}, \bar{z})$, the solution to (q, M), has $S(\bar{z}) \neq S(q-)$, the Jacobian matrix of Λ has singularities along the zero curve γ of Λ. There is at least one singularity for each element in the disjoint union*

$$\big(S(\bar{z}) \cup S(q-)\big) \setminus \big(S(\bar{z}) \cap S(q-)\big).$$

Proof. Let $(\bar{w}, \bar{z})$ be the solution to (q, M) and let $i \in S(\bar{z}) \setminus S(q-)$. First note that, on the (unique) zero curve γ of Λ, both z and w are continuous functions of λ. Since $z = q-$ when $\lambda = 0$, there must be a point λ_0 such that, along the zero curve, $z_i = 0$ for $0 \leqq \lambda \leqq \lambda_0$ and $z_i > 0$ for $\lambda_0 < \lambda < \lambda_0 + \epsilon$ for some ϵ. Since complementarity is maintained along the zero curve, $w_i = 0$ for $\lambda_0 < \lambda < \lambda_0 + \epsilon$. By continuity, w_i must be 0 at $\lambda = \lambda_0$. This means that both z_i and $w_i = [(1-\lambda)I + \lambda M]_{i\cdot}z + q_i$ are zero at $\lambda = \lambda_0$, and hence the Jacobian matrix $D\Lambda(\lambda_0, z(\lambda_0))$ is singular.

Similarly, let $i \in S(q-) \setminus S(\bar{z})$. There must be a point λ_1 such that, along the zero curve γ, $z_i > 0$ for $0 \leqq \lambda < \lambda_1$ and $z_i = 0$ at $\lambda = \lambda_1$. Again by complementarity and continuity, w_i must be 0 at $\lambda = \lambda_1$ and the Jacobian matrix $D\Lambda(\lambda_1, z(\lambda_1))$ is singular. $\square$

5. Relaxation of q.

We can also relax the right-hand side of the LCP keeping the matrix M fixed. This map maintains feasibility and complementarity, but uses a convex combination of the vectors q and $\|q\|_\infty e$ for the right-hand side of the equation. When $\lambda = 0$, we have the trivial problem $(\|q\|_\infty e, M)$ where the right-hand side has all components positive and, when $\lambda = 1$, we have the given problem (q, M).

Define $\Theta : [0,1) \times E^n \to E^n$ by

$$\Theta_i(\lambda, z) = -\big|M_{i\cdot}z + \lambda q_i + (1-\lambda)\|q\|_\infty - z_i\big|^3 + \big(M_{i\cdot}z + \lambda q_i + (1-\lambda)\|q\|_\infty\big)^3 + z_i^3$$

for $i = 1, \cdots, n$.

Since this is once again Mangasarian's map with a slightly different vector for q, feasibility and complementarity on the zero set of Θ is guaranteed. By Theorem 3.1

and the remark following it, we know that the LCP has a locally unique solution for any q whenever M is nondegenerate strictly semimonotone. Thus we easily have the following theorem about Θ.

THEOREM 5.1. *Let $M \in E^{n\times n}$ be positive definite or a P-matrix, and let $q \in E^n$. Then there exists a zero curve γ of Θ emanating from* $(0,0)$ *and reaching a point* $(1,\bar{z})$, *where $\bar{z}$ solves the* LCP (q,M).

Note that Theorem 5.1 does not include strictly copositive or strictly semimonotone matrices, nor any reference to the rank of the Jacobian matrix along the zero curve γ. If M is nondegenerate strictly copositive or nondegenerate strictly semimonotone, there is a solution to the LCP $(\lambda q + (1-\lambda)\|q\|_\infty e, M)$ for every $\lambda \in [0,1]$ by Theorem 3.1. However, there may be multiple solutions, and when the number of solutions changes at some $\bar{\lambda}$ some of the zero curves of Θ either "stop" or "start" at $\bar{\lambda}$. Thus there is no guarantee that a *single* zero curve of Θ will reach all the way from $\lambda = 0$ to $\lambda = 1$. For example, take

$$M = \begin{pmatrix} 1 & 5 & 10 \\ 5 & 1 & 10 \\ 1 & 1 & 1 \end{pmatrix}, \qquad q = \begin{pmatrix} -2 \\ -2 \\ -1 \end{pmatrix}.$$

M is nondegenerate strictly semimonotone, but the zero curve emanating from $(0,0)$ disappears at $\lambda = 4/5$.

Furthermore, we cannot say that the Jacobian matrix is nonsingular along the entire zero curve. The ith row of the Jacobian matrix of Θ is

$$\begin{aligned}\big(D\Theta(\lambda,z)\big)_{i\cdot} = \big(&-3|A|(A)(q_i - \|q\|_\infty) + 3(B)^2(q_i - \|q\|_\infty),\\ &-3|A|(A)(m_{i1}) + 3(B)^2 m_{i1}, \cdots,\\ &-3|A|(A)(m_{ii}-1) + 3(B)^2 m_{ii} + 3z_i^2, \cdots,\\ &-3|A|(A)(m_{in}) + 3(B)^2 m_{in}\big),\end{aligned}$$

$$\begin{aligned}\text{where } A &= M_{i\cdot}z + \lambda q_i + (1-\lambda)\|q\|_\infty - z_i,\\ B &= M_{i\cdot}z + \lambda q_i + (1-\lambda)\|q\|_\infty.\end{aligned}$$

Note that the first column and the diagonal element differ slightly in form from the rest of the entries. Also note that if z_i and $w_i = M_{i\cdot}z + \lambda q_i + (1-\lambda)\|q\|_\infty$ are both zero for some λ, then every entry in $(D\Theta)_{i\cdot}$ is 0. Hence, the Jacobian matrix is singular and we have the following proposition.

PROPOSITION 5.2. *Let $M \in E^{n\times n}$ be positive definite or a P-matrix, and let $q \in E^n$. Whenever M and q are such that $(\bar{w},\bar{z})$, the solution to (q,M), has $\bar{z} \neq 0$, the Jacobian matrix of Θ has singularities along the zero curve γ of Θ. There are at least as many singularities as there are nonzero components of $\bar{z}$.*

Proof. Let $(\bar{w},\bar{z})$ be the solution to (q,M) and let i be such that $\bar{z}_i > 0$. First note that, on the (unique) zero curve γ of Θ, both z and w are continuous functions of λ. Since $z = 0$ when $\lambda = 0$, there must be a point λ_0 such that, along the zero curve, $z_i = 0$ for $0 \leqq \lambda \leqq \lambda_0$ and $z_i > 0$ for $\lambda_0 < \lambda < \lambda_0 + \epsilon$ for some ϵ. Since complementarity is maintained along the zero curve, $w_i = 0$ for $\lambda_0 < \lambda < \lambda_0 + \epsilon$. By continuity, w_i

must be 0 at $\lambda = \lambda_0$. This means that both z_i and $w_i = M_{i\cdot}z + \lambda q_i + (1-\lambda)\|q\|_\infty$ are zero at $\lambda = \lambda_0$, and hence the Jacobian matrix $D\Theta(\lambda_0, z(\lambda_0))$ is singular. $\square$

Geometrically, the singularity corresponds to the point at which the vector $\lambda q + (1-\lambda)\|q\|_\infty$ passes through the boundary of one complementary cone [44], [48], [56], [62] and into another. If it happens that this vector stays in such a boundary for all λ in some interval $[\lambda_0, \lambda_1]$, then z_i and w_i are simultaneously 0, and the Jacobian matrix is singular, along that entire interval. Since there are a finite number (2^n) of complementary cones, however, we can always perturb the right-hand side by adding some $(\epsilon, \epsilon^2, \cdots, \epsilon^n)$, for example, so that there are only a finite number of singularities.

6. Relaxation of complementarity.

This section presents a map that uses the given matrix M and the given vector q, but does not maintain a complementary solution as we track the zero curve. Although nonnegativity of z is preserved along the curve, complementarity is enforced only at the very end of the curve, when $\lambda = 1$. Throughout this section, let $M \in E^{n\times n}$ and $q \in E^n$ be fixed.

Define $\Psi : E^n \times [0,1) \times E^n \to E^n$ by

$$\Psi_i(a,\lambda,z) = -\lambda|M_{i\cdot}z + q_i - z_i|^3 + \lambda(M_{i\cdot}z + q_i)^3 + z_i{}^3 - (1-\lambda)a_i{}^3$$

for $i = 1, \cdots, n$. For fixed $a \in E^n$ let $\Psi_a(\lambda, z) = \Psi(a,\lambda,z)$. The next few lemmas show that, for suitable matrices M, there is a zero curve of Ψ that can be tracked to obtain a solution to the LCP (q, M).

LEMMA 6.1. *If $a \geqq 0$, then $z \geqq 0$ on $\Psi_a^{-1}(0)$.*

Proof. Note that if both z_k and $M_{k\cdot}z + q_k$ are negative, then the entire sum comprising $(\Psi_a(\lambda,z))_k$ is negative. If, on the other hand, $z_k < 0$ and $M_{k\cdot} + q_k \geqq 0$, then $|M_{k\cdot} + q_k| < M_{k\cdot} + q_k - z_k$, and the sum is again negative. $\square$

LEMMA 6.2. *Let M be strictly semimonotone. Then there exists $r > 0$ such that $z \in E^n$, $z \geqq 0$, and $\|z\|_\infty = r$ implies that $z_k(Mz+q)_k > 0$ for some index k.*

Proof. First let

$$\Phi(z) = \max_{1\leqq i\leqq n} z_i(Mz)_i$$

and note that, because M is strictly semimonotone, $\Phi > 0$ for $z \geq 0$. Also note that since Φ is continuous and $\{z : z \geqq 0, \|z\|_\infty = 1\}$ is compact, Φ must assume its minimum on that set. Call that minimum $\bar{\Phi}$ and take $r > \|q\|_\infty/\bar{\Phi}$. Then for $z \geqq 0$ and $\|z\|_\infty = r$, there is some index k such that

$$\begin{aligned} z_k(Mz+q)_k &= \|z\|_\infty^2 \Phi(z/\|z\|_\infty) + z_k q_k \\ &\geqq \|z\|_\infty^2 \bar{\Phi} - \|z\|_\infty \|q\|_\infty \\ &= \|z\|_\infty(\|z\|_\infty \bar{\Phi} - \|q\|_\infty) \\ &> 0. \end{aligned}$$

$\square$

LEMMA 6.3. *Let M be strictly semimonotone. Then there exists $r > 0$ such that $\Psi_0(\lambda, z) \neq 0$ for $0 \leqq \lambda \leqq 1$ and $\|z\|_\infty = r$.*

Proof. By Lemma 6.1, it suffices to consider $z \geqq 0$. Let r and k be as in the conclusion of Lemma 6.2 above and simply notice that, since z_k and $(Mz+q)_k$ are both positive, $\Psi_0(\lambda, z)$ cannot be 0. $\square$

LEMMA 6.4. *Let M be strictly semimonotone. Then there exists $r > 0$ and $\delta > 0$ such that $0 \leqq \lambda \leqq 1$, $\|z\|_\infty = r$, and $\|a\|_\infty < \delta$ implies $\Psi_a(\lambda, z) \neq 0$.*

Proof. Let r be as in Lemma 6.3, and note that $\{(a, \lambda, z) \mid a = 0,\ 0 \leqq \lambda \leqq 1, \|z\|_\infty = r\}$ is disjoint from $\Psi^{-1}(0)$. Since the first of these sets is compact and the second is closed there is a positive distance $\delta > 0$ between them, measured in the max norm. This δ satisfies the conclusion of the Lemma. □

Notice that a positive definite matrix is also a P-matrix, a P-matrix is strictly semimonotone by the sign-reversal property of P-matrices [13], and a strictly copositive matrix is clearly strictly semimonotone. Hence, Lemmas 6.1–6.4 hold for any such matrix and we can state the following theorem.

THEOREM 6.5. *Let $M \in E^{n \times n}$ be positive definite, a P-matrix, strictly copositive, or strictly semimonotone, and let $q \in E^n$. Then there exists $\delta > 0$ such that for almost all $a > 0$, $\|a\|_\infty < \delta$ there is a zero curve γ of $\Psi_a(\lambda, z)$, along which the Jacobian matrix $D\Psi_a(\lambda, z)$ has full rank, emanating from $(0, a)$ and reaching a point $(1, \bar{z})$, where $\bar{z}$ solves the* LCP (q, M).

Proof. First observe that, for $a > 0$, Ψ is transversal to 0 (i.e., its Jacobian matrix has full rank on $\Psi^{-1}(0)$). To see this, note that $\partial\Psi_i/\partial a_j$ is zero if $i \neq j$, and nonzero if $i = j$. Thus, the n columns of $D\Psi$ corresponding to the partials of Ψ with respect to the a_i are linearly independent. Clearly, Ψ_a is C^2, and therefore by Lemma 2.1, for almost all $a > 0$, Ψ_a is also transversal to 0. Thus, by the implicit function theorem, Ψ_a has a zero curve γ, starting from $(0, a)$, along which the Jacobian matrix $D\Psi_a(\lambda, z)$ has full rank. All of this is true regardless of the conditions on the matrix M.

For M strictly semimonotone (positive definite, strictly copositive, or a P-matrix), Lemma 6.4 insures that there exists $\delta > 0$ such that the zero curve γ is bounded for $\|a\|_\infty < \delta$ and $0 \leqq \lambda \leqq 1$. Note that $(0, a)$ is the unique zero of Ψ_a at $\lambda = 0$, and by the implicit function theorem, γ cannot return to $(0, a)$. Since the curve can neither simply stop, nor return to $\lambda = 0$, nor go to infinity, it must reach a point $(1, \bar{z})$, where $\bar{z}$ solves the LCP (q, M). □

7. Expanded Lagrangian homotopy.

The expanded Lagrangian approach [54] may be described as an optimization/continuation approach and has in its simplest form two main steps.

Step 1. (Optimization phase).

At $r = r_0 > 0$ solve the unconstrained minimization problem

$$\min_{w,z} P(w, z, r),$$

where

$$P(w, z, r) = \frac{1}{2r}\|w - Mz - q\|_2^2 + \frac{1}{2r}\langle w, z\rangle^2 - r\sum_{i=1}^{n} \ln z_i - r\sum_{i=1}^{n} \ln w_i.$$

Step 2A. (Switch to expanded system).

A (local) solution of $\min P$ must satisfy

$$0 = \nabla_{(w,z)} P = \begin{pmatrix} I \\ -M^t \end{pmatrix} \frac{(w - Mz - q)}{r} + \begin{pmatrix} z \\ w \end{pmatrix} \frac{\langle w, z\rangle}{r} - r\left(\frac{1}{w_1}, \cdots, \frac{1}{w_n}, \frac{1}{z_1}, \cdots, \frac{1}{z_n}\right)^t.$$

Introduce the following variables:

$$\begin{aligned}
\beta &= \frac{w - Mz - q}{r}, \\
\theta &= \frac{\langle w, z\rangle}{r}, \\
\mu_i &= \frac{r}{w_i}, \qquad i = 1, \cdots, n, \\
\eta_i &= \frac{r}{z_i}, \qquad i = 1, \cdots, n,
\end{aligned}$$

which ultimately represent the Lagrange multipliers. This helps to remove the inevitable ill conditioning associated with penalty methods for small r and we thus obtain our equivalent but expanded system:

$$\begin{aligned}
\begin{pmatrix} I \\ -M^t \end{pmatrix} \beta + \begin{pmatrix} z \\ w \end{pmatrix} \theta - \begin{pmatrix} \mu \\ \eta \end{pmatrix} &= 0, \\
w - Mz - q - r\beta &= 0, \\
\langle w, z\rangle - r\theta &= 0, \\
\mu_i w_i - r &= 0, \qquad i = 1, \cdots, n, \\
\eta_i z_i - r &= 0, \qquad i = 1, \cdots, n.
\end{aligned}$$

(Remark. As a result of the optimization phase and the initial starting point with $r_0 > 0$, the solution $(w^{(0)}, z^{(0)})$ of $\min P(w, z, r_0)$ satisfies $z^{(0)} > 0$ and $w^{(0)} > 0$. As a consequence, $\mu^{(0)} > 0$ and $\eta^{(0)} > 0$ from the definitions of μ and η. They remain positive until $r = 0$, where we formally have

$$\begin{aligned}
\begin{pmatrix} I \\ -M^t \end{pmatrix} \beta + \begin{pmatrix} z \\ w \end{pmatrix} \theta - \begin{pmatrix} \mu \\ \eta \end{pmatrix} &= 0, \\
w - Mz - q &= 0, \\
\langle w, z\rangle &= 0, \\
\mu_i w_i &= 0, \qquad i = 1, \cdots, n, \\
\eta_i z_i &= 0, \qquad i = 1, \cdots, n, \\
w, z, \theta, \mu, \eta &\geqq 0,
\end{aligned}$$

which implies that we have solved the problem.)

In practice we do not solve the optimization problem $\min P$ to high accuracy since a highly accurate solution may have only a digit or two in common with the final answer. However, it is imperative that ∇P be reasonably small in magnitude, say, less than $r_0/10$. The expanded system is converted to a homotopy map by letting $r = r_0(1 - \lambda)$ and modifying the first equation to obtain:

$$\begin{aligned}
\begin{pmatrix} I \\ -M^t \end{pmatrix} \beta + \begin{pmatrix} z \\ w \end{pmatrix} \theta - \begin{pmatrix} \mu \\ \eta \end{pmatrix} - \frac{r}{r_0} \nabla P(w^{(0)}, z^{(0)}, r_0) &= 0, \\
w - Mz - q - r\beta &= 0, \\
\langle w, z\rangle - r\theta &= 0, \\
\mu_i w_i - r &= 0, \qquad i = 1, \cdots, n, \\
\eta_i z_i - r &= 0, \qquad i = 1, \cdots, n.
\end{aligned}$$

Write this system of $5n+1$ equations in the $5n+2$ variables λ, w, z, β, θ, μ, η as

$$\Upsilon(\lambda, w, z, \beta, \theta, \mu, \eta) = 0.$$

Step 2B. (Track the zero curve of Υ from $r = r_0$ to $r = 0$.)

Starting with arbitrary $r_0 > 0$, $w^{(0)} > 0$, and $z^{(0)} > 0$, the rest of the initial point $(0, w^{(0)}, z^{(0)}, \beta^{(0)}, \theta_0, \mu^{(0)}, \eta^{(0)})$ is given by

$$\begin{aligned}
\beta^{(0)} &= \frac{w^{(0)} - Mz^{(0)} - q}{r_0}, \\
\theta_0 &= \frac{\langle w^{(0)}, z^{(0)} \rangle}{r_0}, \\
\mu_i^{(0)} &= \frac{r_0}{w_i^{(0)}}, \qquad i = 1, \cdots, n, \\
\eta_i^{(0)} &= \frac{r_0}{z_i^{(0)}}, \qquad i = 1, \cdots, n.
\end{aligned}$$

This approach requires careful attention to implementation details. For example, the linear algebra and globalization techniques with dynamic scaling are critically important in the optimization phase. For degenerate problems the path can still be long. One possible resolution is the use of shifts and weights as developed in the method of multipliers [5], but holding $r = r_0$ fixed. (This approach is currently under investigation in the context of linear programming [53].) However, in keeping with the philosophy of the "pure" homotopy approach of the current work, we do not solve the optimization problem (Step 1), but instead use the above equations $\Upsilon(\lambda, w, z, \beta, \theta, \mu, \eta) = 0$ as a "pure" homotopy.

Logarithmic barrier potential functions are hardly new [5], and have been used recently by Kojima et al. [26], [27] and Mizuno et al. [38] to extend the ideas of Karmarkar to obtain polynomial-time algorithms for the LCP. The exact details of how the barrier parameter, step size selection, concomitant numerical linear algebra, and initial point computation are handled are crucial to the practical utility of such methods, and in practice are far more significant than theoretical polynomial complexity. It is reasonable that the pure expanded Lagrangian homotopy (without the optimization step) would behave significantly differently from other logarithmic barrier homotopies [26], [27], [38], which include a Phase 1 step equivalent to Step 1 here. These latter homotopies of Kojima et al. are certainly *not* globally convergent, since they require a nontrivial preliminary computation to get a special starting point at which to begin the homotopy.

8. Absolute Newton method. The method of this section is not a homotopy method, but is presented for the sake of comparison and as an example of what can be done with a Newton-type iterative scheme (see also [1] and [35]). Let $x = (w, z) \in E^{2n}$ and define $F : E^{2n} \to E^{2n}$ by

$$F(x) = \begin{pmatrix} w - Mz - q \\ w_1 z_1 \\ \vdots \\ w_n z_n \end{pmatrix}.$$

Then the LCP (q, M) is equivalent to $F(x) = 0$ for x nonnegative. $F(x) = 0$ is a polynomial system of equations of total degree 2^n, which in general has 2^n solutions over complex Euclidean space C^{2n}, counting multiplicities and solutions at infinity. Thus *all* solutions of the LCP (q, M) are among the zeros of $F(x)$, including degenerate solutions, which correspond to manifolds (in C^{2n}) of zeros of $F(x)$. The algebraic geometry theory of polynomial systems is rich and deep, and beyond the scope of this paper. Discussions of the pertinent aspects of algebraic and differential geometry for polynomial systems are in [39], [40], [41], and [68]. It suffices to note here that $F(x)$ is a polynomial system with a particularly simple structure.

The Jacobian matrix of F is

$$DF(x) = \begin{pmatrix} I & -M \\ \operatorname{diag}(z_1, \cdots, z_n) & \operatorname{diag}(w_1, \cdots, w_n) \end{pmatrix},$$

a $2n \times 2n$ matrix. The absolute Newton iteration is

$$x^{(k+1)} = \left| x^{(k)} - \left[DF(x^{(k)})\right]^{-1} F(x^{(k)}) \right|, \quad k = 0, 1, 2, \cdots$$

for an arbitrary starting point $x^{(0)} \in E^{2n}$. The absolute value signs mean to replace each component of the vector by its absolute value (precisely, $|x| = x+ + x-$). When this iteration is well defined is given by the following theorem:

THEOREM 8.1. *Let $M \in E^{n \times n}$ be nondegenerate and let $\bar{x} = (\bar{w}, \bar{z})$ be a zero of F. Then the Jacobian matrix $DF(\bar{x})$ is invertible if and only if $|\bar{w}| + |\bar{z}| > 0$.*

Proof. Suppose that $\bar{w}_k = \bar{z}_k = 0$. Then the $(n+k)$th row of $DF(\bar{x})$ is zero, so $DF(\bar{x})$ is not invertible.

Conversely, suppose that $|\bar{w}| + |\bar{z}| > 0$. Observe that $\bar{w}$ and $\bar{z}$ are complementary vectors, since $\bar{x} = (\bar{w}, \bar{z})$ is a zero of F. For each index k such that $\bar{z}_k \neq 0$ interchange the kth and $(n+k)$th columns of $DF(\bar{x})$. This produces a matrix of the form

$$\begin{pmatrix} A & * \\ 0 & \operatorname{diag}(\bar{w}_1 + \bar{z}_1, \cdots, \bar{w}_n + \bar{z}_n) \end{pmatrix},$$

where $A_{\cdot i} \in \{I_{\cdot i}, -M_{\cdot i}\}$ for $i = 1, \cdots, n$. $\det A$ is a principal minor of $-M$ and is thus nonzero, since M is nondegenerate by assumption. Further, since $|\bar{w}| + |\bar{z}| > 0$ and $\bar{w}$, $\bar{z}$ are complementary, $\bar{w}_i + \bar{z}_i \neq 0$ for $i = 1, \cdots, n$. Thus

$$\begin{aligned} \det DF(\bar{x}) &= \pm \det A \ \det \ \operatorname{diag}(\bar{w}_1 + \bar{z}_1, \cdots, \bar{w}_n + \bar{z}_n) \\ &= \pm \det A \prod_{i=1}^{n} (\bar{w}_i + \bar{z}_i) \\ &\neq 0, \end{aligned}$$

and $DF(\bar{x})$ is invertible. □

This absolute Newton iteration has been used for chemical equilibrium systems, which have a unique real positive solution. It has never been observed to fail for those systems with a random starting point $x^{(0)}$ [36]. The asymptotic behavior of this absolute Newton iteration is not understood, nor even the ordinary Newton iteration in complex Euclidean space C^{2n}, which is related to Julia sets and chaotic dynamical systems. Both the standard Newton iteration and the absolute Newton iteration were tried on $F(x) = 0$, where M was a P-matrix, and both completely failed for starting points distant from the solution. Why the absolute Newton method should be so successful on chemical equilibrium polynomial systems, and fail on LCP polynomial systems, is not clear.

9. Kojima–Saigal homotopy. This homotopy [25] uses the same nonlinear system as the absolute Newton method. Suppose that $w^{(0)}, z^{(0)} \in E^n$ have been obtained such that

$$w^{(0)} - Mz^{(0)} = q,$$
$$w^{(0)} > 0, \quad z^{(0)} > 0.$$

This can be done, for example, by applying Phase 1 of the simplex algorithm to the problem

$$w - Mz = q - e + Me,$$
$$w \geqq 0, \quad z \geqq 0$$

to get a feasible solution $(\bar{w}, \bar{z}) \geqq 0$. Then $w^{(0)} = \bar{w} + e > 0$ and $z^{(0)} = \bar{z} + e > 0$ will suffice. The homotopy map $K : [0,1) \times E^n \times E^n \to E^n$ is given by

$$K(\lambda, w, z) = \begin{pmatrix} w - Mz - q \\ w_1 z_1 - (1-\lambda) w_1^{(0)} z_1^{(0)} \\ \vdots \\ w_n z_n - (1-\lambda) w_n^{(0)} z_n^{(0)} \end{pmatrix}.$$

The following theorem shows that this is a reasonably good homotopy map, at least for P-matrices.

THEOREM 9.1. *Let $M \in E^{n\times n}$ be a P-matrix and let $q \in E^n$. Then there exist $w^{(0)}, z^{(0)} \in E^n$ such that*

$$w^{(0)} - M\, z^{(0)} = q, \quad w^{(0)} > 0, \quad z^{(0)} > 0.$$

Furthermore, there is a zero curve γ of $K(\lambda, w, z)$, along which the Jacobian matrix $DK(\lambda, w, z)$ has full rank (for $0 \leqq \lambda < 1$), emanating from $(0, w^{(0)}, z^{(0)})$ and reaching a point $(1, \bar{w}, \bar{z})$, where $\bar{z}$ solves the LCP (q, M). *λ is strictly increasing as a function of arc length s along γ ($d\lambda/ds > 0$).*

Proof. Since M is a P-matrix, the LCP $(q - e + Me, M)$ has a solution $(\hat{w}, \hat{z})$ by Theorem 6.5. Then $w^{(0)} = \hat{w} + e > 0$ and $z^{(0)} = \hat{z} + e > 0$ have the desired properties.

The Jacobian matrix of $K(\lambda, w, z)$ is

$$DK(\lambda, w, z) = \begin{pmatrix} 0 & I & -M \\ w_1^{(0)} z_1^{(0)} & & \\ \vdots & \operatorname{diag}(z_1, \cdots, z_n) & \operatorname{diag}(w_1, \cdots, w_n) \\ w_n^{(0)} z_n^{(0)} & & \end{pmatrix}.$$

Suppose $(w, z) > 0$ and consider the last $2n$ columns $D_{(w,z)}K$ of DK:

$$\begin{aligned} \det D_{(w,z)}K &= \det \begin{pmatrix} I & -M \\ \operatorname{diag}(z_1, \cdots, z_n) & \operatorname{diag}(w_1, \cdots, w_n) \end{pmatrix} \\ &= \det \begin{pmatrix} I & -M \\ 0 & \operatorname{diag}(w_1, \cdots, w_n) + \operatorname{diag}(z_1, \cdots, z_n)\, M \end{pmatrix} \\ &= \det\bigl(\operatorname{diag}(w_1, \cdots, w_n) + \operatorname{diag}(z_1, \cdots, z_n)\, M\bigr) \\ &> 0 \end{aligned}$$

since $\operatorname{diag}(w_1, \cdots, w_n) + \operatorname{diag}(z_1, \cdots, z_n) M$ is also a P-matrix (it is easily verified that the principal minors remain positive after multiplying by and adding a positive diagonal matrix). Thus rank $DK(\lambda, w, z) = 2n$ for $0 \leqq \lambda < 1$ and $w > 0$, $z > 0$. By the Implicit Function Theorem, there is a zero curve γ of K emanating from $(0, w^{(0)}, z^{(0)})$, and the Jacobian matrix $DK(\lambda, w, z)$ has full rank along γ for $0 \leqq \lambda < 1$ since $w > 0$, $z > 0$ along γ by continuity and the definition of K.

γ can be parametrized by arc length s, giving $\lambda = \lambda(s)$, $w = w(s)$, $z = z(s)$ along γ. Furthermore, the last $2n$ columns of $DK(\lambda(s), w(s), z(s))$ being independent means that $w = w(\lambda)$, $z = z(\lambda)$, and $d\lambda/ds > 0$ along γ (this is well known, see [65], for example). Thus $\lambda = \lambda(s)$ is strictly increasing along γ.

To prove that γ reaches $\lambda = 1$, it suffices to prove that γ is bounded. Let $\alpha = \max_A \{\|A\|_\infty, \|A^{-1}\|_\infty\}$, where the maximum is taken over all matrices $A \in E^{n\times n}$ with $A_{\cdot i} \in \{I_{\cdot i}, -M_{\cdot i}\}$ for $i = 1, \cdots, n$. α is well defined since each $\det A$ is a principal minor of $-M$, which is nonzero by assumption. Fix λ_0 in $(0,1)$, and let $\epsilon = \max_i (1 - \lambda_0) w_i^{(0)} z_i^{(0)}$. Then for $\lambda_0 < \lambda(s) \leqq 1$, either $w_i(s) < \epsilon$ or $z_i(s) < \epsilon$ along γ. For $i = 1, \cdots, n$, let y_i be $w_i(s)$ or $z_i(s)$, whichever is less than ϵ, and let $\bar{y}_i$ be the complementary variable. Write $w(s) - M\, z(s) = q$ as

$$A\,y + B\,\bar{y} = q.$$

Then

$$\|\bar{y}\|_\infty = \left\|B^{-1}(q - Ay)\right\|_\infty \leqq \left\|B^{-1}\right\|_\infty \left(\|q\|_\infty + \|A\|_\infty \|y\|_\infty\right) \leqq \alpha(\|q\|_\infty + \alpha\epsilon),$$

which says that $w(s)$ and $z(s)$ are bounded for $\lambda_0 < \lambda(s) \leqq 1$. $\square$

Note that the theorem does not include strictly semimonotone matrices since $\operatorname{diag}(w_1, \cdots, w_n) + \operatorname{diag}(z_1, \cdots, z_n) M$ can be singular for strictly semimonotone M. Thus while K is a better homotopy than Λ, Θ, and Υ, it is not as generally applicable as ρ_a or Ψ_a.

10. Numerical experiments. The homotopy maps from the previous sections were tested on several problems, chosen to illustrate certain features of the various homotopies. A complete description of the data, tables of numerical results, and a comparative discussion of the different homotopy maps and numerical results are in [70]. The main observations from those experiments are summarized here: The probability-one homotopies ρ_a and Ψ_a work for everything that the theory predicts. The computational complexity of ρ_a and Ψ_a, measured by the number of steps along the zero curve, is relatively insensitive to n. This is in direct contrast to pivoting methods, which can exhibit exponential complexity in the number of steps [47]. The homotopies Λ and Θ frequently fail, but when they work at all, may be more efficient than the homotopies ρ_a or Ψ_a. The expanded Lagrangian homotopy Υ without the optimization phase fails for most starting points, with the zero curves of Υ either going off to infinity or returning to another solution at $r = r_0$. Υ does work very well from sufficiently close starting points, but these are not random starting points (as are used for ρ_a and Ψ_a), and the homotopy algorithm based on Υ without optimization is certainly not globally convergent. The Kojima–Saigal homotopy requires Phase 1 of the simplex algorithm just to get a starting point, which is antithetical to the homotopy philosophy of global convergence from an easily obtainable starting point. Furthermore, K and Ψ_a both essentially relax complementarity, and Ψ_a is more generally applicable.

11. Conclusion. There are many reasonable ways to construct a homotopy map for the LCP, and only a few of the possibilities have been considered here. The homotopies here fall into three different classes: *artificial*, *natural*, and *interior*. (See the discussion of the words "artificial" and "natural" in relation to homotopies in [69].) Λ and Θ are "natural" homotopies in the sense that for each $\lambda \in [0,1]$ the equation $\Lambda(\lambda, z) = 0$ or $\Theta(\lambda, z) = 0$ corresponds to an LCP. Thus, the intermediate points (λ, z) on the zero curve of the homotopy map have interpretations as solutions to a related family of LCPs. In contrast, ρ_a and Ψ_a are "artificial" homotopies in that the homotopy equations $\rho_a(\lambda, z) = 0$ and $\Psi_a(\lambda, z) = 0$ do *not* correspond to an LCP for $0 < \lambda < 1$, and the points (λ, z) on the zero curves for $0 < \lambda < 1$ have no useful interpretations as LCP solutions. Υ and K would be considered "interior" methods, since they only generate points (λ, w, z) interior to the feasible region, i.e., $(w, z) > 0$ for $0 \leqq \lambda < 1$. These class distinctions are not always clear-cut, but are useful at a high conceptual level.

The theory of globally convergent probability-one homotopy maps can be applied to the LCP in several ways; the maps ρ_a and Ψ_a are two examples. The convergence theory for the homotopy maps ρ_a and Ψ_a is very satisfactory: global convergence from an arbitrary starting point is guaranteed for a wide class of LCPs. Theorems 3.1 and 6.5 are existence results, and as such are close to the best known existence results.

Our computational experience, reported in [70], indicates that Ψ_a is the best homotopy. It never failed, is indeed globally convergent, and was frequently more efficient than Λ and Θ, even on problems where Λ and Θ did well. ρ_a takes second place, since it also never failed, but tends to be very expensive (long homotopy zero curves). This is not surprising, since Ψ_a was crafted with the benefit of ten years experience since ρ_a was created. It is quite likely that a more efficient globally convergent homotopy map than Ψ_a can yet be constructed.

Λ and Θ failed badly on problems with many singularities (corresponding to the right-hand side passing through the face of a complementary cone) along the zero curves of the homotopy maps Λ and Θ. One might hope that the curve tracking algorithms would, by chance, miss hitting the singularities exactly and thereby step past them. This does happen, to some extent, but when there are a large number of singularities close together or highly rank deficient singularities (corresponding to the right-hand side passing through a lower dimensional face of a complementary cone), the numerical linear algebra is simply overwhelmed by the ill conditioning.

Overall, the natural homotopies Λ and Θ are much worse than the artificial homotopies ρ_a and Ψ_a. For particular problems, a natural homotopy may be very efficient, but their performance is unreliable and very much data dependent. The difficulties, both theoretical (cf. Propositions 4.2 and 5.2) and practical, of natural homotopies like Λ and Θ appear to remove them from further consideration (cf. the discussions in [39]–[41] and [69]).

The numerical experiments show that the expanded Lagrangian homotopy is unacceptable as a robust homotopy without solving the optimization problem (Step 1). The zero set of Υ contains loops (in $[0,1) \times E^{5n+1}$) starting and ending at $\lambda = 0$ as well as unbounded curves. Although the increased dimension is discouraging, we do note that $2n$ of the $5n+1$ equations result in diagonal matrices which can be exploited in the linear algebra. Furthermore, Υ does work well for fair starting points, and so

Υ may be useful for LCPs using an optimization phase to get a fairly good starting point. Although the expanded Lagrangian homotopy is an interior method based on a logarithmic barrier potential function similar in spirit to methods of Kojima et al. [26], [27], [38], it is not equivalent to any of those methods. The Kojima et al. methods converge to a solution in polynomial time from an arbitrary interior starting point (for a restricted class of LCPs), which is not true of the expanded Lagrangian homotopy method. However, generating a feasible interior starting point for K is tantamount to the optimization Step 1 for Υ, and neither K nor Υ can be considered a globally convergent homotopy for the LCP in the same sense as ρ_a and Ψ_a. Furthermore, the Kojima et al. homotopies without Phase 1 would be even less successful than the expanded Lagrangian homotopy is without Step 1.

The Kojima–Saigal homotopy is closely related to the continuous Newton homotopy of Smale. Both are theoretically interesting, but computational experience on real problems [67], [68] suggests that the globally convergent probability-one homotopies (like ρ_a and Ψ_a) are more robust and more general than the continuous Newton homotopies. Our numerical experience is that interior homotopies like Υ and K (lacking dynamic scaling) are very inefficient, but worthy of further study. At any rate, Ψ_a is more general than K (cf. Theorems 6.5 and 9.1). Similar comments apply to the polynomial-time homotopies of [26], [27], and [38], which are both less stable numerically and less generally applicable than probability-one homotopies like Ψ_a.

There are numerous fixed point iterative schemes for the LCP [2], [3], [8], [18], [35], [50], [51], [61], but they generally involve nonsmooth operators (e.g., $v+$ or $|v|$) or apply to a small class of matrices (e.g., symmetric positive definite M). Homotopy algorithms are more versatile than fixed point iteration algorithms, but whether they are competitive with fixed point iteration remains to be seen. A systematic comparison of complementary pivoting, fixed point iteration, and homotopy methods would be a worthwhile undertaking.

The LCP is a *linear* combinatorial problem. That the LCP should be reformulated as a *nonlinear* problem, which is in turn embedded in a complicated *nonlinear* homotopy, is counterintuitive. Nevertheless, a homotopy algorithm based on $\Psi_a(\lambda, z)$ is globally convergent for a wide class of LCPs, numerically robust, reasonably efficient, and (most encouraging) rather insensitive to the dimension of the problem.

12. Acknowledgment. The authors are indebted to Jong-Shi Pang, Katta Murty, and Romesh Saigal for useful comments and suggestions.

REFERENCES

[1] M. AGANAGIC, *Newton's method for linear complementarity problems*, Math. Programming, 28 (1984), pp. 349–362.

[2] B. H. AHN, *Solution of nonsymmetric linear complementarity problems by iterative methods*, J. Optim. Theory Appl., 33 (1981), pp. 175–185.

[3] ———, *Iterative methods for linear complementarity problems with upper bounds on primary variables*, Math. Programming, 26 (1983), pp. 295–315.

[4] Y. BARD, *Nonlinear Parameter Estimation*, Academic Press, New York, 1974.

[5] D. P. BERTAEKAS, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[6] R. W. Cottle and G. B. Danzig, *Complementary pivot theory of mathematical programming*, Linear Algebra Appl., 1 (1968), pp. 103–125.

[7] ———, *Monotone solutions of the parametric linear complementarity problem*, Math. Programming, 3 (1972), pp. 210–224.

[8] R. W. Cottle and J. S. Pang, *On the convergence of a block successive overrelaxation method for a class of linear complementarity problems*, Math. Programming, 17 (1982), pp. 126–138.

[9] C. Cryer, *The solution of a quadratic programming problem using systematic overrelaxation*, SIAM J. Control, 9 (1971), pp. 385–392.

[10] B. C. Eaves, *The linear complementarity problem in mathematical programming*, Management Sci., 17 (1971), pp. 612–634.

[11] ———, *Homotopies for computation of fixed points*, Math. Programming, 3 (1972), pp. 1–22.

[12] B. C. Eaves and R. Saigal, *Homotopies for computation of fixed points on unbounded regions*, Math. Programming, 3 (1972), pp. 225–237.

[13] M. Fiedler and V. Pták, *On matrices with non-positive off-diagonal elements and positive principal minors*, Czech. Math. J., 12 (1962), pp. 382–400.

[14] J. T. Fredricksen, L. T. Watson, and K. G. Murty, *A finite characterization of K-matrices in dimensions less than four*, Math. Programming, 35 (1986), pp. 17–31.

[15] D. Gale, *The Theory of Linear Economic Models*, McGraw-Hill, New York, 1960.

[16] T. Hansen and H. Scarf, *On the applications of a recent combinatorial algorithm*, Cowles Commission Discussion Paper No. 272, Yale Univ., New Haven, CT, 1969.

[17] A. W. Ingleton, *A problem in linear inequalities*, Proc. London Math. Soc. 3rd Series, 16 (1966), pp. 519–536.

[18] N. W. Kappel and L. T. Watson, *Iterative algorithms for the linear complementarity problem*, Internat. J. Computer Math., 19 (1986) pp. 273–297.

[19] S. Karamardian, *The complementarity problem*, Math. Programming, 2 (1972), pp. 107–129.

[20] L. M. Kelly and L. T. Watson, *Q-matrices and spherical geometry*, Linear Algebra Appl., 25 (1979), pp.175–190.

[21] ———, *Erratum: Some perturbation theorems for Q-matrices*, SIAM J. Appl. Math., 34 (1978), pp. 320–321.

[22] M. Kojima and R. Saigal, *On the number of solutions to a class of complementarity problems*, Math. Programming, 21 (1981), pp. 190–203.

[23] ———, *A study of PC^1 homeomorphisms on subdivided polyhedrons*, SIAM J. Math. Anal., 10 (1979), pp. 1299–1312.

[24] ———, *On the number of solutions for a class of linear complementarity problems*, Math. Programming, 17 (1979), pp. 136–139.

[25] ———, *Private communication*, October, 1987.

[26] M. Kojima, S. Mizuno, and A. Yohishe, *A polynomial-time algorithm for a class of linear complementarity problems*, Math. Programming, to appear.

[27] M. Kojima, S. Mizuno, and T. Noma, *A new continuation method for complementarity problems with uniform P-functions*, Math. Programming, to appear.

[28] M. M. Kostreva, *Elasto-hydrodynamic lubrication: A nonlinear complementarity problem*, Mathematics Dept., GM Research Laboratories, Warren, MI, 1982.

[29] C. E. Lemke, *Bimatrix equilibrium points and mathematical programming*, Management Sci., 11 (1965), pp. 681–689.

[30] ———, *On complementary pivot theory*, in Mathematics of the Decision Sciences Part 1, G. B. Danzig and A. F. Veinott, Jr., eds., Amer. Math. Soc., Providence, RI, 1968, pp. 95–114.

[31] ———, *Recent results on complementarity problems*, in Nonlinear Programming, J. B. Rosen, O. L. Mangasarian and K. Ritter, eds., Academic Press, New York, 1970, pp. 349–384.

[32] C. E. Lemke and J. T. Howson, *Equilibrium points of bimatrix games*, SIAM J. Appl. Math., 12 (1964), pp. 413–423.

[33] O. L. Mangasarian, *Equivalence of the complementarity problem to a system of nonlinear equations*, SIAM J. Appl. Math., 31 (1976), pp. 89–92.

[34] ———, *Linear complementarity problems solvable by a single linear program*, Math. Programming, 10 (1976), pp. 263–270.

[35] ———, *Solution of symmetric linear complementarity problems by iterative methods*, J. Optimization Theory Appl., 22 (1977), pp. 465–485.

[36] K. MEINTJES AND A. P. MORGAN, *A methodology for solving chemical equilibrium systems*, Appl. Math. Comput., 22 (1987), pp. 333–361.

[37] O. H. MERRILL, *Applications and extensions of an algorithm that computes fixed points of certain non-empty, convex, upper semi-continuous point to set mappings*, Dept. of Industrial Engineering, Univ. of Michigan Tech. Report No. 71-7, 1971.

[38] S. MIZUNO, A. YOHISHE, AND T. KIKUCHI, *Practical polynomial time algorithms for linear complementarity problems*, Dept. of Management Sci. and Eng., Tokyo Institute of Technology, Tech. Report No. 13, Tokyo, Japan, April, 1988.

[39] A. P. MORGAN, *Solving polynomial systems using continuation for scientific and engineering problems*, Prentice-Hall, Englewood Cliffs, NJ, (1987).

[40] A. P. MORGAN AND A. J. SOMMESE, *A homotopy for solving general polynomial systems that respects m-homogeneous structures*, Appl. Math. Comput., 24 (1987), pp. 101–113.

[41] ———, *Computing all solutions to polynomial systems using homotopy continuation*, Appl. Math. Comput., 24 (1987), pp. 115–138.

[42] K. G. MURTY, *On the parametric complementarity problem*, Engineering Summer Conference notes, University of Michigan, Ann Arbor, MI, 1971.

[43] ———, *On a characterization of P-matrices*, SIAM J. Appl. Math., 20 (1971), pp. 378–384.

[44] ———, *On the number of solutions to the complementarity problem and spanning properties of complementary cones*, Linear Algebra Appl., 5 (1972), pp. 65–108.

[45] ———, *Note on a Bard-type scheme for solving the complementarity problem*, Opsearch, 11 (1974) pp. 123–130.

[46] ———, *Linear and Combinatorial Programming*, John Wiley, New York, 1976, pp. 481–519.

[47] ———, *Computational complexity of complementary pivot methods*, Math. Programming Study 7 (1978), pp. 61–73.

[48] ———, *On the linear complementarity problem*, Methods Oper. Res., 31 (1978), pp. 425–439.

[49] ———, *Linear Complementarity, Linear and Nonlinear Programming*, Helderman-Verlag, West Berlin, 1986.

[50] J. S. PANG, *On the convergence of a basic iterative method for the implicit complementarity problem*, J. Optim. Theory Appl., 37 (1982), pp. 139–162.

[51] ———, *Necessary and sufficient conditions for the convergence of iterative methods for the linear complementarity problem*, J. Optim. Theory Appl., 42 (1984), pp. 1–18.

[52] J. S. PANG AND R. CHANDRASEKHARAN, *Linear complementarity problems solvable by a polynomially bounded pivoting algorithm*, Math. Programming Study 25 (1985), pp. 13–27.

[53] A. B. POORE AND D. SORIA, *Continuation algorithms for linear programming*, in preparation.

[54] A. B. POORE AND Q. AL-HASSAN, *The expanded Lagrangian system for constrained optimization problems*, SIAM J. Control Optim., 26 (1988), pp. 417–427.

[55] R. SAIGAL, *A note on a special linear complementarity problem*, Opsearch, 7 (1970), pp. 175–183.

[56] ———, *On the class of complementary cones and Lemke's algorithm*, SIAM J. Appl. Math., 23 (1972), pp. 46–60.

[57] ———, *On the convergence rate of algorithms for solving equations that are based on methods of complementary pivoting*, Math. Operations Res., 22 (1977), pp. 108–124.

[58] H. SAMUELSON, R. M. THRALL AND O. WESLER, *A partitioning theorem for Euclidean n-space*, Proc. Amer. Math. Soc., 9 (1958), pp. 805–807.

[59] H. SCARF, *The Computation of Economic Equilibria*, Yale Univ. Press, New Haven, CT, 1973.

[60] A. C. STICKNEY AND L. T. WATSON, *Digraph models of Bard-type algorithms for the linear complementarity problem*, Math. Operations Res., 3 (1978), pp. 322–333.

[61] W. M. G. VAN BOKHOVEN, *A class of linear complementarity problems is solvable in polynomial time*, unpublished paper, Dept. of Electrical Eng., Univ. of Technology, The Netherlands, 1980.

[62] L. T. WATSON, *A variational approach to the linear complementarity problem*, Doctoral Dissertation, Dept. of Math., Univ. of Michigan, Ann Arbor, MI, 1974.

[63] ———, *Some perturbation theorems for Q-matrices*, SIAM J. Appl. Math., 31 (1976), pp. 379–384.

[64] ———, *An algorithm for the linear complementarity problem*, Internat. J. Computer Math., 6 (1978), pp. 319–325.

[65] ———, *A globally convergent algorithm for computing fixed points of C^2 maps*, Appl. Math. Comput., 5 (1979), pp. 297–311.

[66] ———, *Solving the nonlinear complementarity problem by a homotopy method*, SIAM J. Control Optim., 17 (1979), pp. 36–46.

[67] ———, *Computational experience with the Chow-Yorke algorithm*, Math. Programming, 19 (1980), pp. 92–101.

[68] ———, *Numerical linear algebra aspects of globally convergent homotopy methods*, SIAM Rev., 28 (1986), pp. 529–545.

[69] L. T. WATSON, S. C. BILLUPS, AND A. P. MORGAN, HOMPACK: *A suite of codes for globally convergent homotopy algorithms*, ACM Trans. Math. Software, 13 (1987), pp. 281–310.

[70] L. T. WATSON, J. P. BIXLER, AND A. B. POORE, *Continuous homotopies for the linear complementarity problem*, Tech. Report TR 87-38, Dept. of Computer Sci., VPI & SU, Blacksburg, VA, 1987.