

Complementarity Problems

Stephen C. Billups^{a,1} Katta G. Murty^b

^a*Department of Mathematics, University of Colorado, Denver, Colorado
80217-3364*

^b*Department of Industrial and Operations Engineering, University of Michigan,
Ann Arbor, Michigan 48109-2117*

Abstract

This paper provides an introduction to complementarity problems, with an emphasis on applications and solution algorithms. Various forms of complementarity problems are described along with a few sample applications, which provide a sense of what types of problems can be addressed effectively with complementarity problems. The most important algorithms are presented along with a discussion of when they can be used effectively. We also provide a brief introduction to the study of matrix classes and their relation to linear complementarity problems. Finally, we provide a brief summary of current research trends.

Key words: complementarity problems, variational inequalities, matrix classes

1 Introduction

The distinguishing feature of a complementarity problem is the set of **complementarity conditions**. Each of these conditions requires that the product of two or more nonnegative quantities should be zero. (Here, each quantity is either a decision variable, or a function of the decision variables). Complementarity conditions made their first appearance in the optimality conditions for continuous variable nonlinear programs involving inequality constraints, which were derived by Karush in 1939. But the significance of complementarity conditions goes far beyond this. They appear prominently in the study of equilibria problems and arise naturally in numerous applications from economics, engineering and the sciences. There is therefore a great deal of practical interest in the development of robust and efficient algorithms for solving complementarity problems.

¹ Research partially supported through NSF grant DMS-9973321.

The early motivation for studying the **Linear Complementarity Problem (LCP)** was because the KKT optimality conditions for linear and **Quadratic Programs (QP)** constitute an LCP of the form (1), or a mixed LCP of the form (2) (see next section for statements of these problems). However, the study of LCP really came into prominence in 1963 when Howson, in his Ph.D. thesis, and Lemke and Howson [10] showed that the problem of computing a **Nash equilibrium point** of a **bimatrix game** can be posed as an LCP of the form (1), and developed an elegant and efficient constructive procedure (the **complementary pivot method**) for solving it. In 1968, the unification of linear and quadratic programs and bimatrix games under the LCP format by Cottle and Dantzig was seen as a fundamental breakthrough, and the study of complementarity problems suddenly blossomed.

The **Nonlinear Complementarity Problem (NCP)** was introduced by Cottle in his Ph.D. thesis in 1964, and the closely related **Variational Inequality Problem (VIP)** was introduced by Hartman and Stampacchia in 1966, primarily with the goal of computing stationary points for nonlinear programs. While these problems were introduced soon after the LCP, most of the progress in developing algorithms for these problems did not begin until the late 1970s.

Well over a thousand articles and several books have been published on the subject of complementarity problems. We limit the scope of this paper. First we describe what complementarity problems are and try to give a sense of what types of problems can be addressed effectively within this framework. This includes a description of the various types of complementarity problems in Section 2, as well as a discussion of applications in Section 3. In Sections 4 and 5 we describe the most important computational algorithms for solving complementarity problems and discuss when these methods are most likely to be successful. In Section 6, we give a brief introduction to the study of matrix classes, which represents a very rich field within LCP. Finally, in Section 7, we discuss some of the current trends in complementarity research. For a detailed comprehensive treatment of the LCP, we refer the reader to the books by Murty [12] and Cottle, Pang, and Stone [3]. For a general treatment of NCP and VIP we recommend the reference [8]. Additional references are given in [2], as well as in the references mentioned above.

2 The Various Complementarity Problems

The simplest and most widely studied of the complementarity problems is the **LCP**, which has often been described as a *fundamental problem* because the first order necessary optimality conditions for QP involving inequality constraints in nonnegative variables form an LCP: given $M \in R^{n \times n}$, $q \in R^n$,

find $w = (w_j) \in R^n, z = (z_j) \in R^n$ satisfying

$$w - Mz = q; \quad w, z \geq 0; \quad w^T z = 0. \quad (1)$$

We denote this LCP by the symbol (q, M) . The name comes from the third condition, the *complementarity condition* which requires that at least one variable in the pair (w_j, z_j) should be equal to 0 in the solution of the problem, for each $j = 1$ to n . This pair is therefore known as the *jth complementary pair* in the problem, and for each j , the variable w_j is known as the *complement* of z_j and vice versa. The LCP (q, M) is said to be *monotone* if the matrix M is **positive semidefinite (PSD)**.

A slight generalization of the LCP is the **mixed LCP (mLCP)**: given $A \in R^{n \times n}, B \in R^{m \times m}, C \in R^{n \times m}, D \in R^{m \times n}, a \in R^n, b \in R^m$, find $u \in R^n, v \in R^m$ satisfying

$$a + Au + Cv = 0, \quad b + Du + Bv \geq 0, \quad v \geq 0, \quad v^T(b + Du + Bv) = 0. \quad (2)$$

The mLCP is a mixture of the LCP and a system of linear equations, which correspond to the unrestricted variables u . The first order necessary optimality conditions for a quadratic program involving some equality and some inequality constraints are in this form. In (2), if A is nonsingular, then u can be eliminated from it using $u = -A^{-1}(a + Cv)$ and then (2) becomes the LCP $(b - DA^{-1}a, B - DA^{-1}C)$. This mLCP is said to be monotone if the matrix

$$\begin{pmatrix} A & C \\ D & B \end{pmatrix} \text{ in (2) is PSD.}$$

Another generalization of the LCP is the **horizontal LCP** or **hLCP**: given $N \in R^{n \times n}, M \in R^{n \times n}, q \in R^n$ find $w \in R^n, z \in R^n$ satisfying

$$Nw - Mz = q; \quad w, z \geq 0; \quad w^T z = 0 \quad (3)$$

The hLCP (3) becomes the standard LCP if $N = I$. Also, if N is nonsingular, then (3) is equivalent to the LCP $(N^{-1}q, N^{-1}M)$. The hLCP is said to be monotone if for any two pairs of points (w^1, z^1) and (w^2, z^2) satisfying $Nw - Mz = q$ we have $(w^1 - w^2)^T(z^1 - z^2) \geq 0$. Note that if $N = I$, this is equivalent to the matrix M being PSD.

For each $i = 1$ to n let m_i be a positive integer, and $m = \sum_{i=1}^n m_i$. The **Vertical LCP** or **VLCP** is another generalization of the LCP for which the

input data are $M \in R^{m \times n}$, $q \in R^m$ partitioned as follows:

$$M = \begin{pmatrix} M^1 \\ \vdots \\ M^n \end{pmatrix}, q = \begin{pmatrix} q^1 \\ \vdots \\ q^n \end{pmatrix}$$

where for each $i = 1$ to n , $M^i \in R^{m_i \times n}$, $q^i \in R^{m_i}$. Given this data, the VLCP is to find $z = (z_i) \in R^n$ satisfying

$$q + Mz \geq 0, \quad z \geq 0, \quad z_i \prod_{j=1}^{m_i} (q^i + M^i z)_j = 0, \quad i = 1, \dots, n. \quad (4)$$

If $m_i = 1$ for all i , then this VLCP becomes the standard LCP.

In the spirit of the VLCP, we can define a **general horizontal linear complementarity problem (HLCP)** involving a vector $q \in R^n$, a square matrix $N \in R^{n \times n}$, a rectangular matrix $M \in R^{n \times m}$ where $m \geq n$, and a partition of the vector of variables $z = (z^1, \dots, z^n)^T \in R^m$ where each z^i is again a vector consisting of one or more variables. Given this data, the problem is to find a $w = (w_i) \in R^n$ and a $z = (z^1, \dots, z^n)^T \in R^m$ satisfying $Nw - Mz = q$, $w \geq 0$, $z \geq 0$, and for each i at least one variable among $\{w_i, z^i\}$ is 0. Clearly (3) is a special case of this problem.

Then there is the **generalized LCP (GLCP)** with data $A, B \in R^{m \times n}$, $C \in R^{m \times d}$ and $q \in R^m$. The problem is to find $(x \in R^n, s \in R^n, z \in R^d)$ satisfying $Ax + Bs + Cz = q$, $(x, s, z) \geq 0$, $x^T s = 0$.

Now we will present some nonlinear generalizations of the LCP. The most important of these is the **NCP**: given a mapping $F(z) = (F_i(z)) : R^n \rightarrow R^n$, find a $z \in R^n$ satisfying

$$z \geq 0, \quad F(z) \geq 0, \quad z^T F(z) = 0. \quad (5)$$

If $F(z)$ is the affine function $q + Mz$, then (5) becomes the LCP (q, M) .

A further generalization of the NCP is the **VIP**: given a mapping $F(z) = (F_i(z)) : R^n \rightarrow R^n$, and $\emptyset \neq K \subset R^n$, find a $z^* \in K$ satisfying

$$(y - z^*)^T F(z^*) \geq 0 \quad \text{for all } y \in K \quad (6)$$

denoted by $VI(K, F)$. If $K = \{z : z \geq 0\}$, then the z^* solving (6) also solves (5). Also, if K is polyhedral and F is affine, it can be verified that $VI(K, F)$ is

an LCP. When K is a rectangular region defined by $K := \prod_{i=1}^n [l_i, u_i]$, $-\infty \leq l_i < u_i \leq \infty$, $i = 1, \dots, n$, this is called the **Box Constrained VIP (BVIP)**, which is also commonly referred to as the **(nonlinear) Mixed Complementarity Problem (MCP)**.

For any subset $K \subset R^n$, its polar cone, denoted by K^* , is defined to be $\{y \in R^n : x^T y \geq 0 \text{ for all } x \in K\}$. Another generalization of the NCP is the **complementarity problem over a cone**: given a mapping $F(z) = (F_i(z)) : R^n \rightarrow R^n$ and a cone K in R^n , find a $z \in K$ satisfying $F(z) \in K^*$, and $z^T F(z) = 0$. This problem, denoted by $CP(K, F)$, reduces to NCP (5) if $K = R_+^n$. Also, since K is a cone, $CP(K, F)$ and $VI(K, F)$ have the same solution set.

3 Example Applications

Complementarity problems arise naturally in the study of many phenomena in economics and engineering. A comprehensive and excellent treatment of applications of complementarity problems is provided in [6]. Additionally, a large collection of problems from a variety of application areas can be found in the MCPLIB library of test problems [4]. Applications of complementarity from the field of economics include general Walrasian equilibrium, spatial price equilibria, invariant capital stock, and game-theoretic models. In engineering, complementarity problems arise in contact mechanics, structural mechanics, obstacle and free boundary problems, elastohydrodynamic lubrication, and traffic equilibrium.

As a rule of thumb, the complementarity framework should be considered whenever the system being studied involves complementary pairs of variables (that is, where one or the other member of each pair must be at its bound). For example, in contact mechanics, the force between two objects is complementary to the distance between the two objects; there is no force unless the distance between the objects is zero. As another example, in Walrasian equilibrium problems, the price of a commodity is complementary to excess supply of the commodity; if there is excess supply, the price will fall either until the demand rises to eliminate the excess supply, or until the price is zero.

3.1 Piecewise Linear Equations

Consider the LCP (q, M) . For each $z \in R^n$ define $h_i(z) = \min\{z_i, (q + Mz)_i\}$, and let $h(z) = (h_i(z))$. Then $h(z) : R^n \rightarrow R^n$ is a piecewise linear concave function, and clearly, solving the LCP (q, M) is equivalent to solving the system of

piecewise linear equations $h(z) = 0$. Conversely, under a mild nonsingularity assumption, any piecewise linear system of equations can be reformulated as a linear complementarity problem.

In the same way, the VLCP (4) is equivalent to the system of piecewise linear equations $H(z) = 0$ where $H(z) = (H_i(z))$ and $H_i(z) = \min\{z_i, (q^i + M^i z)_1, \dots, (q^i + M^i z)_{m_i}\}$ for $i = 1$ to n .

3.2 An Application Involving a Smallsize Convex QP Model

Since the optimality conditions for convex QP form an LCP or mLCP, any application involving convex QP offers an application for LCP or mLCP. We present a recent application, described by Murty [13], in supply chain management, which is becoming increasingly popular.

An important issue in supply chain management is to forecast the demand for each item and to determine when to place orders for it and the order quantities. Classical analysis in the inventory management literature assumes that the distribution of demand is known and typically assumes this distribution to be normal. This assumption confers many theoretical advantages, the principal among which is the fact that the normal distribution is fully characterized by only two parameters, the mean and the standard deviation. So, when the distribution changes, one just has to change the values of these two parameters in the models.

In recent times, in the computer and electronics industries and many other manufacturing industries, the rapid rate of technological change is resulting in new products replacing the old periodically. The result is that product life cycles are shortened. The short life cycle itself is partitioned into three distinct periods. A *growth period* at the beginning of life sees the demand for the item growing due to gradual market penetration, reaching its peak by the end of this period. This is followed by a short *stable period* during which the demand for the item is relatively stable. This is followed by the final *decline period* during which the demand for the item undergoes a steady decline until it is replaced at the end by a technologically superior one. Because of this constant change, classical models based on a single stable demand distribution are not suitable. We need to use models that frequently and periodically update the demand distribution based on recent data.

Approximating demand distributions by something like the normal or gamma distributions, which are characterized by two or fewer parameters, allows us the freedom to change only those few parameters when updating the demand distribution. This appears quite inadequate to capture all the dynamic changes occurring in the shapes of demand distributions. A better strategy

is to approximate the demand distribution by its histogram from past data. In this approximation, called the *discretized demand distribution*, the range of variation of the demand is divided into a convenient number (about 10 to 25 in practice) demand intervals, and the probability associated with each interval in the initial distribution is taken to be its relative frequency among historical data.

Let I_1, \dots, I_n be the demand intervals and $p = (p_1, \dots, p_n)^T$ the vector of probabilities associated with them in the present distribution. For $i = 1$ to n , let r_i be the relative frequency in I_i over the most recent k periods (if the period is a day for example, k could be about 50) at the time of updating. Let $x = (x_1, \dots, x_n)^T$ denote the unknown current (i.e., updated) probability vector. $r = (r_1, \dots, r_n)^T$ is an estimate of x , but it is based on too few (only k) observations. We can take an estimate of x to be the \bar{y} which is the optimum solution of the following quadratic program. This \bar{y} will be used in place of p in the next planning period.

$$\begin{aligned} \min \quad & \beta \sum_{i=1}^n (p_i - y_i)^2 + (1 - \beta) \sum_{i=1}^n (r_i - y_i)^2 \\ \text{subject to} \quad & \sum_{i=1}^n y_i = 1, \quad \text{and all } y_i \geq 0 \end{aligned}$$

where $0 < \beta < 1$ is a weight. Typically $\beta = 0.9$ works well. The reason for choosing the weight of the second term in the objective function to be small is because the relative frequency vector r is based on a small number of observations. Since the quadratic objective function in this model is the weighted sum of squared forecast errors over all demand intervals, it has the effect of tracking gradual changes in the demand distribution when used at every ordering period. Optimal ordering policies based on discretized demand distributions are discussed in [13].

The optimum solution for the quadratic program above is $\bar{y} = (\bar{z}_1, \dots, \bar{z}_{n-1}, \bar{w}_n)^T$ where $\bar{w} = (\bar{w}_j)^T$, $\bar{z} = (\bar{z}_j)^T$ is the solution of the LCP with data

$$q = \begin{pmatrix} -1 - \beta(p_1 - p_n) - (1 - \beta)(r_1 - r_n) \\ -1 - \beta(p_2 - p_n) - (1 - \beta)(r_2 - r_n) \\ \vdots \\ -1 - \beta(p_{n-1} - p_n) - (1 - \beta)(r_{n-1} - r_n) \\ 1 \end{pmatrix}, \quad M = \begin{pmatrix} 2 & 1 & \dots & 1 & 1 \\ 1 & 2 & \dots & 1 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & \dots & 2 & 1 \\ -1 & -1 & \dots & -1 & 0 \end{pmatrix}$$

This LCP is of small order, and it can be solved very conveniently using the complementary pivot algorithm, or any other pivoting algorithm for the LCP mentioned above. Simple codes for the complementary pivot algorithm using

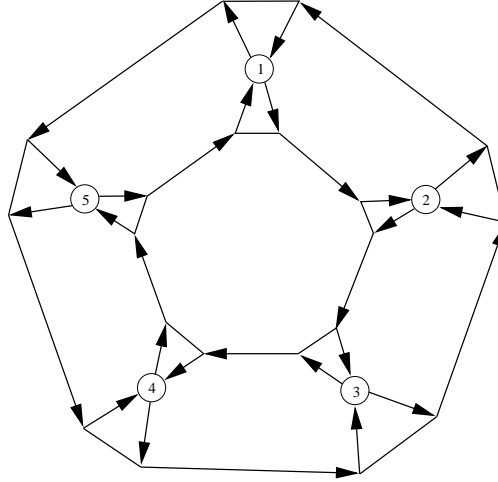


Fig. 1. A Traffic Network.

the explicit inverse of the basis, available from several sources, work very well on problems of this size. In supply chain management, such a model is solved for each item at every ordering period, providing an excellent source of application for the pivoting algorithms for the LCP.

3.3 Traffic Equilibrium

The following traffic equilibrium example from MCPLIB [4] illustrates the connection between equilibrium and complementarity.

The problem involves five cities, numbered 1 through 5, connected by a network of one-way roads called links, (see Figure 1). Each city i must ship a quantity d_i of a commodity to the third city clockwise from itself. For example, city 1 ships to city 4, city 2 to city 5, and so on. Naturally, the goal is to ship the commodity in the shortest time possible. However, the time to ship along a given path is determined by the total flow of traffic on the links making up that path.

From the figure, it is clear that for each city, there are only two possible shortest paths: shipping counterclockwise along the outside loop or clockwise along the inside loop. Let x_i represent the amount shipped from city i to city $i + 3$ (modulo 5) along the outside path, and let y_i represent the amount shipped along the inside path. We say that a set of flows $x = (x_i), y = (y_i)$ is feasible if it satisfies the demands $x + y \geq d$ and $x, y \geq 0$, where $d = (d_i)$.

From the flow vectors x and y , it is possible to determine the traffic on each link of the network. For example, the outside link between cities 3 and 4 will have flow given by $x_1 + x_2 + x_3$. The delay on a link k is determined by the total traffic on the link and is assumed to be a convex function of the traffic

flow. The delay for a given path is then given as the sum of the delays of all the links making up that path. From this discussion, it follows that for each city i , the delay O_i along the outside path is determined by the flow vector x , and the delay I_i along the inside path is determined by the flow vector y . This can be encapsulated by defining the two functions $O(x) = (O_i(x))$ and $I(y) = (I_i(y))$.

We define the *effective delay* between two cities to be the maximum delay among paths with nonzero flow between the two cities. Each city chooses a shipping strategy in order to minimize its effective delay subject to the shipping strategies of the other cities remaining constant. This minimum is achieved either when the city ships everything along the path with shortest delay, or when both the inside and the outside paths have equal delay. To see this, note that if the delay for the inside path is less than the delay for the outside path, the shipper can improve the effective delay by shipping more along the inside path. This reduces the traffic on the outside path, which reduces the delay on the outside path thereby reducing the effective delay.

An equilibrium traffic pattern emerges when all five cities are shipping optimally subject to the shipping strategies of the other cities remaining constant. From the above discussion, this is equivalent to the complementarity conditions

$$\begin{aligned} 0 &\leq O(x) - u, & x &\geq 0, & x^T(O(x) - u) &= 0 \\ 0 &\leq I(y) - u, & y &\geq 0, & y^T(I(y) - u) &= 0 \end{aligned}$$

where we have introduced the additional variable $u \in \mathbb{R}^5$ to represent the effective delay. Notice that u is complementary to the demand constraint. In particular, there can only be excess supply if the effective delay is zero.

The conditions described above lead to the NCP (5), with $z := (x, y, u)$ and

$$F \begin{pmatrix} x \\ y \\ u \end{pmatrix} := \begin{pmatrix} O(x) - u \\ I(y) - u \\ x + y - d \end{pmatrix}.$$

It should be noted, however, that this problem can be solved more efficiently using a generalization of the NCP. In particular, this problem can be reformulated as a BVIP involving only five variables, instead of the fifteen used above. Details of this are provided in [4].

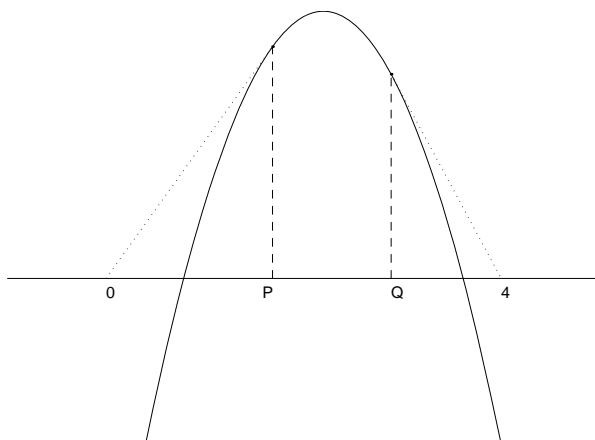


Fig. 2. An elastic string stretched over an obstacle.

3.4 Obstacle and Free Boundary Problems

The obstacle problem consists of finding the equilibrium position of an elastic membrane that is held at a fixed position on its boundary and which lies over an obstacle.

Consider stretching an elastic string fixed at the endpoints $(0, 0)$, and $(4, 0)$ over an obstacle defined by a function f (in this example, we use $f(x) = 1 - (x - 2.2)^2$ —see Figure 2). Notice that the position of the string will be defined by $f(x)$ for x between the unknown points P and Q , and that in the intervals $0 \leq x \leq P$, and $Q \leq x \leq 4$, the string will lie along straight line segments connecting $(0, 0)$ to $(P, f(P))$ and $(Q, f(Q))$ to $(4, 0)$, respectively. If we represent the equilibrium position of the string by the function u , then u must satisfy the following conditions:

$$\begin{aligned} u(0) &= 0, & u(4) &= 0, \\ u'(P) &= f'(P), & u'(Q) &= f'(Q), \\ u(x) &= f(x) & \text{for } P \leq x \leq Q, \\ u''(x) &= 0 & \text{for } 0 < x < P \text{ or } Q < x < 4. \end{aligned}$$

This representation of the problem is complicated by the presence of the free boundaries P and Q . The complementarity framework allows a simpler representation, which does not require free boundaries. First, note that since there is no downward force on the string, $u''(x) \leq 0$ for all x , except possibly at $x = P$ or $x = Q$ where u'' may be discontinuous. Also, note that $u(x) \geq f(x)$

everywhere. Finally, at each point x , either $u''(x) = 0$ or $u(x) = f(x)$. Thus, if we ignore momentarily the discontinuity of u'' at P and Q , we see that u must satisfy the conditions

$$\begin{aligned} u(0) &= u(4) = 0 \\ u(x) &\geq f(x) \\ u''(x) &\leq 0 \\ (u(x) - f(x)) u''(x) &= 0 \end{aligned} \quad 0 \leq x \leq 4.$$

This system can be solved numerically using a finite difference or finite element scheme. For example, using a central difference scheme on a regular mesh with step size $h = 4/n$, u is approximated by the vector $u = (u_0, u_1, \dots, u_n)$, where $u_i := f(x_i)$, $x_i := x_0 + ih$, $i = 0, \dots, n$, and $x_0 = 0$. The above system is then approximated by

$$\begin{aligned} u_0 &= u_n = 0 \\ \frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} &\leq 0 \\ u_i - f(x_i) &\geq 0 \\ (u_i - f(x_i)) \frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} &= 0 \end{aligned} \quad i = 1, \dots, n-1.$$

Using the simple change of variables $z_i := u_i - f(x_i)$, this system is equivalent to the linear complementarity problem (q, M) , where M is an $(n-1) \times (n-1)$ matrix and q is an $(n-1)$ -vector defined by

$$M = \begin{pmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & \cdots & & 0 & -1 & 2 \end{pmatrix} \quad q := \begin{pmatrix} -2f(x_1) + f(x_2) \\ f(x_1) - 2f(x_2) + f(x_3) \\ \vdots \\ f(x_{n-3}) - 2f(x_{n-2}) + f(x_{n-1}) \\ f(x_{n-2}) - 2f(x_{n-1}) \end{pmatrix}.$$

The solution $z = (z_1, \dots, z_{n-1})$ of this LCP then gives the discrete approximation to u at the interior grid points by the relation $u_i = z_i + f(x_i)$, $i = 1, \dots, n-1$.

4 Algorithms for the LCP

The fascination of the subject stems from the fact that it exhibits enormous diversity. Depending on the properties of the data matrix M , the LCP can be so nice at one end that it admits an extremely simple greedy type algorithm for its solution, or be an intractable NP-hard problem at the other end.

Chung (published in 1989, but result discussed in [12] also) has shown that the LCP (q, M) with general integer data is NP-hard. The only known algorithms that are guaranteed to process the LCP (q, M) with no restrictions on the data are enumerative algorithms.

A *complementary vector* of variables in (1) is a vector $y = (y_1, \dots, y_n)^T$ where $y_j \in \{w_j, z_j\}$ for all $j = 1$ to n . A complementary vector is said to be *basic* if the set of column vectors associated with them in (1) form a nonsingular matrix, and it is said to be *basic feasible* if the basic solution of (1) associated with it is nonnegative. Clearly, the basic solution associated with a basic feasible complementary vector is a solution of the LCP.

4.1 Pivotal Methods

The first class of methods to be developed for the LCP are the pivotal methods which try to obtain a basic feasible complementary vector through a series of pivot steps. These methods are variants of the complementary pivot method of Lemke and Howson [10]. The mathematical principle used by Lemke and Howson in the complementary pivot method has been applied by Scarf in 1967 to develop a method for computing fixed points, and for solving systems of nonlinear equations, using partitions of R^n into sets called primitive sets. In the 1970's and early 1980's a lot of researchers have extended this work and developed a variety of methods for computing fixed points and for solving nonlinear equations and complex equilibrium problems using triangulations of R^n . These methods are now called **simplicial methods**, or **piecewise linear methods** because they employ piecewise linear approximations of maps, or also **complementary pivot methods**. Prime candidates among these are the methods of Merrill, and Eaves and Saigal, developed in 1972.

The simplicial methods start at a solution of an artificially set up simple system, and trace a path through the n -dimensional simplices of the triangulation, which, when the method works, terminates with a simplex that contains an easily computed approximate solution of the original system. Using a homotopy interpretation of this path, a variety of other homotopy and path tracing algorithms have been developed for solving systems of nonlinear equations.

Almost all the pivotal methods are guaranteed to process the LCP (q, M) when the matrix M is PSD; this class of LCPs is equivalent to the class of convex quadratic programs. All these algorithms are finite procedures when applied to the classes of problems for which they are guaranteed to work, and in practice these algorithms are quite efficient on problems of reasonable size. However, in 1978, Murty showed that in the worst case, two of the most important among them, Lemke's method and Murty's least-index method, require $O(2^n)$ pivot steps to solve monotone LCPs of order n ; this same result has been extended to other pivotal methods for the LCP.

The most famous among the pivotal methods for the LCP is Lemke's method, which we now describe in some detail. Given an LCP (q, M) , define the feasible region to be the set

$$\mathcal{F} := \{z \mid w := Mz + q \geq 0, z \geq 0\}.$$

Elements of \mathcal{F} are called *feasible solutions*, and extreme points (or vertices) of \mathcal{F} are called *basic feasible solutions*. A feasible solution z is said to be *complementary* if $z_i w_i = 0$ for all i and *almost complementary* if $z_i w_i = 0$ for all but one i , where $w := Mz + q$. Clearly, a point z is a complementary feasible solution if and only if z solves (q, M) .

For simplicity of discussion, we assume that the LCP is *nondegenerate*, which means that for every basic feasible solution z , the vector (z, w) has exactly n nonzero components. Under this nondegeneracy assumption, every complementary feasible solution is an extreme point of \mathcal{F} , and every almost complementary solution lies on an edge of \mathcal{F} , where we define an edge to be the intersection of $n - 1$ linearly independent hyperplanes of the form $z_i = 0$ or $w_j = 0$. If every point on the edge is almost complementary, we call the edge an *almost complementary edge*. If the edge is unbounded, it is called an *almost complementary ray*.

It can be shown that every almost complementary (but not complementary) extreme point (satisfying, for example, $w_j z_j = 0$, for $j = 2, \dots, n$) is incident with exactly two almost complementary edges of \mathcal{F} , every point on which also satisfies the same conditions ($w_j z_j = 0$, for $j = 2, \dots, n$). Hence, the set of all such solutions is a collection of paths. Lemke's method traces exactly one such path in this collection, beginning with an almost complementary extreme point incident to an almost complementary ray. Such a path must terminate, either with a complementary basic feasible solution, or with a secondary almost complementary ray. In the first case, a solution to the LCP is found. In the second case, the method has failed to produce a solution to the problem.

To get the algorithm started, it is necessary to construct an almost complementary extreme point that is incident to an almost complementary ray. Since

this may be difficult to find for the original LCP (q, M) , we instead solve the augmented LCP (\tilde{q}, \tilde{M}) , where $\tilde{M} \in \mathbb{R}^{(n+1) \times (n+1)}$ and $\tilde{q} \in \mathbb{R}^{n+1}$ are defined by

$$\tilde{M} := \begin{pmatrix} 1 & 0 \\ e & M \end{pmatrix} \quad \tilde{q} = \begin{pmatrix} 0 \\ q \end{pmatrix},$$

where $e \in \mathbb{R}^{n+1}$ is the vector of all ones. Notice that a point (z_0, z) is a solution to (\tilde{q}, \tilde{M}) if and only if $z_0 = 0$ and z is a solution to (q, M) . Notice further that by choosing $\delta = \max\{|q_i| \mid q_i < 0\}$, the point $(z_0, z) = (\delta, 0)$ is an almost complementary extreme point of \mathcal{F} that is incident to the almost complementary ray $\{(z_0, z) = (\delta, 0) + t(1, 0) \mid t \geq 0\}$. Thus, a starting point for Lemke's method can easily be generated for this augmented LCP.

When the LCP is degenerate, the above method can be modified using a degeneracy resolution technique such as lexicographical ordering. For details of such techniques, we refer the reader to the monographs [12, 3].

4.2 Interior Point Methods

The other important class of methods for the LCP are the interior point methods. Interior point methods originated from an algorithm introduced by Karmarkar in 1984 for solving linear programs. The most successful interior point methods follow a path in $\mathcal{F}^0 = \{(w, z) : w - Mz = q, w > 0, z > 0\}$ (hence the name interior point methods) in an effort to reduce $w^T z$ to 0. One such method defines this path as the set of solutions to the following parameterized system

$$\begin{aligned} w - Mz &= q, \\ w_i z_i &= \mu, \quad i = 1, \dots, n \\ w &> 0, z > 0, \end{aligned} \tag{7}$$

where each choice of the parameter μ yields a different point along the path. This path is followed by generating a sequence of iterates $\{(w^k, z^k)\}$, starting from a feasible point (w^0, z^0) . Each step is calculated by solving the system

$$\begin{pmatrix} -I & M \\ Z^k & W^k \end{pmatrix} \begin{pmatrix} \Delta w^k \\ \Delta z^k \end{pmatrix} = \begin{pmatrix} w^k - Mz^k - q \\ -W^k Z^k e + \mu_k e \end{pmatrix},$$

where W^k and Z^k are the diagonal matrices whose diagonal components are defined by $W_{ii}^k = w_i^k$ and $Z_{ii}^k = z_i^k$, and $\mu_k = (z^k)^T w^k / n$. Note that this system is just Newton's method applied to (7).

The next iterate (w^{k+1}, z^{k+1}) is then determined by

$$(w^{k+1}, z^{k+1}) = (w^k, z^k) + \alpha_k (\Delta w, \Delta z),$$

where the steplength α_k is chosen to ensure that the iterates don't get too close to the boundary of the positive orthant.

The mLCP and hLCP are solved using essentially a similar strategy. In fact, in a 1995 paper, Guler showed that a monotone hLCP can be reduced to a monotone LCP in the same variables. This allows any interior point method for monotone LCPs to be generalized to solve monotone hLCPs. Moreover, the iterates generated by the interior point method in solving the hLCP correspond directly to iterates that would be produced by applying the interior point method to the corresponding LCP. Thus, any convergence results that hold true for the monotone LCP also hold true for the monotone hLCP.

Interior point methods for convex QP essentially use the above methods on the KKT conditions, which form an LCP or mLCP. These methods have polynomial time worst case complexity for monotone complementarity problems, which correspond to convex quadratic programs (Convex QPs).

4.3 Computational Applications and Limitations

Convex QP models like the one discussed in Section 3 involving not too large a number of variables and constraints, appear very commonly in many sciences and are a major source of application for the pivoting algorithms for the LCP. These are the algorithms of choice when the number of variables + the number of constraints in a convex QP is of the order of 100 or less. However, when a convex QP is a large scale problem (i.e., when the number of variables + the number of constraints is $\gg 100$) computational tests indicate that the active set methods of nonlinear programming are much better suited to solve it than the pivot methods applied to the LCP or mLCP formed by its first order optimality constraints. The newly developed interior point methods for convex QP also compete with active set methods for solving large scale problems.

Convex QPs appear very prominently in the **Sequential (or Recursive) QP Approach (SQP)** for solving nonlinear programs. The SQP approach solves nonlinear programs (either convex or nonconvex) using a series of steps; each step consisting of solving a convex QP to find a search direction and then a

line search to find an optimum step length in that direction. Most publicly distributed nonlinear programming software packages based on SQP do not use the LCP based algorithms for solving the convex QP in each step because the authors of these codes assume that the users will apply them to solve large scale nonlinear programs. The convex QP solvers in these codes are usually based on some type of active set approach.

Methods based on the LCP or mLCP formed by the first order necessary optimality conditions (i.e., the KKT conditions) are not suitable for handling nonconvex QPs. This is because all these methods focus only on finding a KKT point for the problem and never even compute the objective value in any step or track how it is changing over the steps. Even if these methods obtain a KKT point, there are no efficient methods known to check whether that KKT point is even a local minimum for the original nonconvex QP as shown by Murty and Kabadi in 1987. For handling nonconvex QP, descent methods, which try to decrease the objective value in each step, are definitely to be preferred in practice.

In the same way, the complementary pivot or simplicial methods for computing fixed points and solving nonlinear equations, using triangulations of R^n , do not have any measure to track the progress of the algorithm from one step to the next. For this reason, these methods are currently not popular.

5 Methods for the NCP

We now turn our focus to methods for solving NCPs. For simplicity, we discuss only the standard NCP defined by (5); although it is straightforward to implement these methods in the context of some more general formulations.

The first methods we consider are sequential LCP methods. These methods generate a sequence of iterates $\{z^k\}$, such that z^{k+1} is a solution to a linear complementarity problem (q^k, M^k) , where q^k and M^k are chosen to approximate F near x^k . Depending on the choice of M^k and q^k , various algorithms can be generated, each of which is analagous to a standard iterative method for solving nonlinear systems of equations. Among these are Newton, quasi-Newton, Jacobi, successive overrelaxation, symmetrized Newton, and projection methods. Details of how to choose M^k and q^k for each of these methods can be found in [8]. Here we focus on Newton's method, which corresponds to the choices $M^k := \nabla F(z^k)$ and $q^k = F(z^k) - M^k z^k$. Notice here that $F^k(z) := M^k z + q^k$ is the first order Taylor approximation to F , making this method analagous to Newton's method for nonlinear equations. Josephy showed in 1979 that, in a neighborhood of a solution z^* to the NCP, the iterates produced by this method are well-defined and converge quadratically

to z^* provided that ∇F is locally Lipschitzian at z^* and that a certain strong regularity assumption is satisfied.

As with Newton's method for nonlinear equations, it is desirable to employ a globalizing strategy to increase the domain of convergence of the method. One such strategy for nonlinear equations is a backtracking linesearch, in which a step is chosen along the Newton direction $d^k := z^{k+1} - z^k$ so as to ensure sufficient descent of a merit function at every iteration. Unfortunately, attempts to apply this strategy to the NCP have been largely unsuccessful and have generally required very strong assumptions. The difficulty lies in the fact that for reasonable choices of merit functions, the Newton direction is not guaranteed to be a descent direction, even when z^k is not a stationary point. Instead, a backtracking strategy that is not restricted to the Newton direction is needed.

Such a strategy was proposed by Ralph in 1994. He devised a path search algorithm in which global convergence is achieved by searching along a piecewise linear path connecting z^k to the solution of the LCP (q^k, M^k) . This path is generated by a complementary pivot algorithm, similar to Lemke's method, which is used to solve the LCP subproblem. This path search strategy is the basis for the highly successful PATH algorithm developed by Dirkse and Ferris in 1995.

Another class of methods for NCP involves reformulating the problem as a system of nonlinear equations. This involves constructing a function $H : \mathbb{R}^n \rightarrow \mathbb{R}^n$ with the property that zeros of H correspond to solutions of the NCP. Such a function H is called an NCP-function. Perhaps the simplest example is to define H by

$$H_i(z) := \min(F_i(z), z_i)$$

Many other NCP functions have been studied in the literature [7]. Interestingly, while smooth NCP functions exist, they are generally not in favor computationally since they have singular Jacobian matrices at degenerate solutions (a degenerate solution is a point z^* such that $F_i(z^*)$ and z_i are both zero for some index i). Thus, the NCP functions of interest are usually only piecewise differentiable.

Once the NCP function H has been constructed, a generalized Newton method can be used to find a solution to $H(z) = 0$. Generalized Newton methods are similar to Newton's method except that the Newton equation

$$\nabla F(z^k)d^k = -F(z^k),$$

which is used to calculate the search direction d^k , is replaced by the equation

$$V^k d^k = -F(z^k),$$

where V^k is an element of the Clarke subdifferential or the B -subdifferential of $F(z^k)$.

An alternative approach to applying a generalized Newton method is to approximate the nonsmooth system by a family of smooth functions. This is the fundamental idea behind the so-called *smoothing methods*. An excellent review of these techniques can be found in [7]. The basic idea of these techniques is to approximate the function H by a family of smooth approximations H_μ parameterized by the scalar μ . Under suitable assumptions, the solutions $z(\mu)$ to the perturbed systems $H_\mu(z) = 0$ form a smooth trajectory parameterized by μ , leading to a zero of H as $\mu \downarrow 0$. The smoothing methods generate a sequence of iterates that follow this trajectory, and in that sense are similar to interior point methods. Numerous smoothers have been proposed in the literature for complementarity problems.

Another class of algorithms for monotone NCP are interior point algorithms. In similar fashion to the interior point techniques for LCP, these methods follow the central path defined by

$$w = F(z), \quad (w, z) > 0, \quad w_i z_i = \mu, \quad (8)$$

which leads to a solution as $\mu \downarrow 0$.

The final class of methods we discuss are continuation or homotopy methods. Like the smoothing and interior point algorithms, these methods work by introducing an additional variable μ and then following a path which leads to a solution. However, unlike the smoothing and interior point methods, the continuation methods do not assume that μ decreases monotonically along this path.

6 The Geometry of LCP, Matrix Classes

For any matrix $D = (d_{ij})$ of order $m \times n$ we let $D_{.j}$ denote its j th column; for any $J \subset \{1, \dots, n\}$ we let $D_{.J}$ denote the $m \times |J|$ matrix consisting of columns $D_{.j}$ for $j \in J$; and for any $P \subset \{1, \dots, m\}$, $J \subset \{1, \dots, n\}$ we let $D_{P,J}$ denote the $|P| \times |J|$ matrix $(d_{ij} : i \in P, j \in J)$.

Consider the LCP (1). Let $y = (y_1, \dots, y_n)^T$ be a complementary vector for it. Let $A_{.j}$ be the column associated with y_j in (1). Hence $A_{.j} \in \{I_{.j}, -M_{.j}\}$ for

each $j = 1$ to n . The cone $\text{Pos}\{A_{.1}, \dots, A_{.n}\} = \{x \in R^n : x = \sum_{j=1}^n \alpha_j A_{.j}, \alpha_j \geq 0 \ \forall j\}$ is known as the **complementary cone** associated with y for (1). Clearly, there are 2^n complementary cones, and the LCP (1) has a solution iff q belongs to some complementary cone. Hence the LCP (1) is equivalent to finding a complementary cone containing q . The geometric study of the LCP using complementary cones has been initiated by Murty in 1968 in his Ph.D. thesis.

It can be verified that if the matrix $M = I$, the unit matrix of order n , the complementary cones become the orthants of R^n . In an earlier paper not focused on the LCP, Samelson, Thrall, and Wesler defined in 1958 the complementary cones with respect to a square matrix M as a generalization of orthants, and investigated the question of what conditions on the matrix M would guarantee that these complementary cones form a partition of R^n . They established that the required condition is that M must be a **P-matrix**, i.e., a square matrix all of whose principal subdeterminants are > 0 . For the LCP (1), this result leads to the theorem that (1) has a unique solution for all $q \in R^n$ iff M is a P-matrix.

The geometric study of the LCP has been the object of enduring study in the literature ever since. This study is purely mathematical in nature, and not motivated by immediate practical application.

We briefly summarize some of the other major results in this geometric investigation. This research has identified a wide variety of classes of square matrices that correspond to certain properties related to the LCP. Let $S(q, M) = \{z : (w := Mz + q, z) \text{ is a solution of the LCP } (q, M)\}$.

The w -part of the solution of the LCP (q, M) is unique $\forall q \in R^n$ iff M is a **column adequate matrix** (this requires that M must be a P_0 -matrix (i.e., all its principal subdeterminants are ≥ 0), and for any $J \subset \{1, \dots, n\}$ M_{JJ} is singular iff the columns of $M_{.J}$ are themselves linearly dependent).

For all $q \in R^n$, every solution of the LCP (q, M) is locally unique iff M is a **nondegenerate matrix** (M belongs to this class iff all its principal subdeterminants are nonzero).

The LCP (q, M) has a unique solution $\forall q > 0$ iff M is a **semimonotone matrix** (class denoted by E_0) (M belongs to this class if $\forall J \subset \{1, \dots, n\}$, the system $M_{JJ}x_J < 0, x_J \geq 0$ has no solution).

The LCP (q, M) has a unique solution $\forall q \geq 0$ iff M is a **strictly semimonotone matrix** (class denoted by E) (M belongs to this class if $\forall J \subset \{1, \dots, n\}$, the system $M_{JJ}x_J \leq 0, x_J \geq 0, x_J \neq 0$ has no solution).

These are some of the main ones, but there are so many other classes of

matrices identified, we refer the reader to [3, 12, 1] for a summary and some recent work in the area. Theoretical studies on the geometry of complementary cones continues to be very actively pursued.

Given a square matrix M of order n , to check whether M is PSD or positive definite (PD) requires at most n Gaussian pivot steps, and hence can be carried out very efficiently. However, for many of the matrix classes defined above, checking membership is intractable, as shown in [12]

The study of the mathematical aspects of the geometry of LCP, matrix classes, and establishing connections between LCP and other branches of mathematics such as degree theory through the study of the degree of piecewise linear equation formulations of the LCP continue to be pursued very actively.

7 Generalizations and Current Trends

Complementarity problems and variational inequalities remain a vigorous area of research. While excellent algorithms have been developed, there is still much attention being devoted to developing new algorithms. Much of this interest lies in expanding the classes of functions for which algorithms can be proven effective. In the realm of LCPs, this, in part, has motivated the study of matrix classes. In the NCP and VI arenas, new algorithms are constantly being introduced. Some of these are based on new merit or NCP functions [5] leading to variants of damped Newton-type methods. Others are variants of path-following algorithms, including homotopy, smoothing [5], interior point [9], and, recently, non-interior point methods. Also various globalizing strategies, such as regularizing methods [7], tunneling and filled functions, trust region methods and proximal perturbation strategies have recently been explored.

Another trend is focused on developing algorithms that do not require Jacobian evaluations. These include projection methods, quasi-Newton methods, and derivative-free methods [7].

Finally, we mention mathematical programs with equilibrium constraints (MPECs). These can be defined in the form

$$\begin{aligned} & \min && f(x, y) \\ & \text{subject to} && (x, y) \in Z \subset \mathbb{R}^{n+m} \\ & && y \text{ solves VI}(C(x), F(x, \cdot)) \end{aligned}$$

where $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ are the decision variables, Z is a closed set, C is a set-valued mapping, and $f : \mathbb{R}^{m+n} \rightarrow \mathbb{R}$ and $F : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^{m+n}$ are given

functions. Here the constraints state that the variable y must be a solution to a variational inequality that is parameterized by x . Numerous applications of MPECs have been identified. These include misclassification minimization in machine learning; robotics; continuous network design; discrete transit planning; optimal design of mechanical structures; and Stackelberg leader-follower games, which have numerous applications in economics, such as oligopolistic market analysis.

The MPEC is an extremely difficult problem to solve. This is due largely to the fact that the feasible region $\{(x, y) \in Z \mid y \text{ solves VI}(C(x), F(x, \cdot))\}$ is not convex, and in some cases, is not even closed. Nevertheless, a number of reasonable algorithms exist for solving MPECs, and research in this area remains vigorous. The reader is referred to [11] for a detailed treatment on MPECs.

References

- [1] S. Bandyopadhyay. *On Lipschitzian, INS and Connected Matrices in LCP*. PhD thesis, Indian Statistical Institute, Delhi, 1998.
- [2] S. C. Billups and K. G. Murty. Complementarity problems. UCD/CCM Report No. 147, Department of Mathematics, University of Colorado at Denver, Denver, Colorado, 1999.
- [3] R. W. Cottle, J.-S. Pang, and R. E. Stone. *The Linear Complementarity Problem*. Academic Press, Boston, 1992.
- [4] S. P. Dirkse and M. C. Ferris. MCPLIB: A collection of nonlinear mixed complementarity problems. *Optimization Methods and Software*, 5:319–345, 1995.
- [5] M. C. Ferris and J.-S. Pang, editors. *Complementarity and Variational Problems: State of the Art*, Philadelphia, Pennsylvania, 1997. SIAM Publications.
- [6] M. C. Ferris and J.-S. Pang. Engineering and economic applications of complementarity problems. *SIAM Review*, 39:669–713, 1997.
- [7] M. Fukushima and L. Qi, editors. *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*. Kluwer Academic Publishers, 1998.
- [8] P. T. Harker and J.-S. Pang. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical Programming*, 48:161–220, 1990.
- [9] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise. *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems*, volume 538 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1991.
- [10] C. E. Lemke and J. T. Howson. Equilibrium points of bimatrix games. *SIAM Journal of Applied Mathematics*, 12:413–423, 1964.

- [11] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996.
- [12] K. G. Murty. *Linear Complementarity, Linear and Nonlinear Programming*. Heldermann Verlag, Berlin, 1988. Free public access at <http://www-personal.engin.umich.edu/~murty/book/LCPbook/index.html>.
- [13] K. G. Murty. Supply chain management in the computer industry. Manuscript, Department of IOE, University of Michigan, Ann Arbor, 1998.