A POSITIVE ALGORITHM FOR THE NONLINEAR COMPLEMENTARITY PROBLEM*

RENATO D. C. MONTEIRO†, JONG-SHI PANG‡, AND TAO WANG†

Abstract. In this paper, the authors describe and establish the convergence of a new iterative method for solving the (nonmonotone) nonlinear complementarity problem (NCP). The method utilizes ideas from two distinct approaches for solving this problem and combines them into one unified framework. One of these is the infeasible-interior-point approach that computes an approximate solution to the NCP by staying in the interior of the nonnegative orthant; the other approach is typified by the NE/SQP method which is based on a generalized Gauss–Newton scheme applied to a constrained nonsmooth-equations formulation of the complementarity problem. The new method, called a positive algorithm for the NCP, generates a sequence of positive vectors by solving a sequence of linear equations (as in a typical interior-point method) whose solutions (if nonzero) provide descent directions for a certain merit function that is derived from the NE/SQP iteration function modified for use in an interior-point context.

Key words. complementarity problems, interior-point methods, nonsmooth equations

AMS subject classifications. 90C30, 90C33, 49M37

1. Introduction. The idea of solving complementarity problems by staying in the interior of the feasible region can be traced to a paper published in 1980 by McLinden [18]. Although no explicit algorithm was formulated, the idea of tracing an interior path as a possible solution procedure was quite evident in this paper and the existence of the “central path” was demonstrated in the case of a complementarity problem with a maximal monotone multifunction. Unfortunately, this paper was not widely known. Of course, McLinden’s idea is central to many of today’s interior-point methods for solving a wide variety of mathematical programming problems.

In recent years, interior-point methods for solving complementarity problems have been the subject of many studies [4], [6]–[14], [19], [20], [26], [28], [31]–[35]. Among these, the monograph [9] presents a unified treatment of the original family of (feasible) interior-point methods for the linear complementarity problem (LCP) that requires all iterates to be strictly feasible; this volume also contains an extensive list of references for the interior-point methods up to the year 1990.

A proposal by Lustig [16] and the subsequent computational study [17] have led researchers to investigate the family of infeasible interior-point methods. The main feature of these methods for solving a complementarity problem is that the iterates are positive vectors, albeit not necessarily feasible to the problem, and have some desirable limiting properties. There are many papers dealing with these methods for solving linear programs; for the linear complementarity problem, we mention [31] and [34]. Most recently, the paper by Kojima, Noma, and Yoshise [15] presents a wide class
of infeasible interior point methods for solving a monotone nonlinear complementarity problem (NCP).

Inclusive of the early work of McLinden, all interior-point methods for solving complementarity problems to date have invariably relied on a certain monotonicity assumption (or more generally, a $P_0$-property) see [8], [9], [15] and [1, §5.9]. This fact helps to explain why the interior-point methods proposed so far for solving nonlinear programs are restricted to the class of linearly constrained convex programs [21]–[23]. In essence, such a monotonicity assumption (or $P_0$-property) is needed to ensure the nonsingularity of a key matrix that is used to define the main computational step of the methods. The major objective of this paper is to propose an interior-point method for solving a general, nonmonotone NCP. In a subsequent paper, we study the specialization of this method to the Karush–Kuhn–Tucker optimality conditions of a general nonlinear program formulated as a mixed NCP.

Our proposed method is an infeasible interior-point potential reduction algorithm; it involves two major ideas. One is to maintain the positivity of the iterates while a certain merit function is being decreased; this joint task is accomplished by means of the modified Armijo technique described in [22]. The other idea is to make use of the iteration function in the recent NE/SQP method [2], [27] to define a suitable merit function. The resulting algorithm does not rely on any monotonicity assumption of the problem. Instead, a key condition, called $s$-regularity, plays an important role in the convergence analysis.

This interior-point method, which we call a positive algorithm for the NCP to signify that the iterates are positive vectors but not necessarily feasible to the problem, consists of solving a sequence of linear equations each defined by a symmetric positive definite matrix; the unique solutions of these equations, if nonzero, provide descent directions for a special logarithmic merit function, which is a combination of the NE/SQP iteration function and the positivity conditions. The NE/SQP method also maintains the nonnegativity of the iterates; but it does so by imposing this requirement as constraints in the direction-finding subproblems that are then either (convex) quadratic [27] or linear programs [2]. Consequently, the positive algorithm may be considered as providing an interior-point approach to alleviate the direction-finding task in the NE/SQP method.

2. Preliminaries. Given a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, which is assumed to be continuously differentiable in an open set containing $\mathbb{R}^n_+$, the nonlinear complementarity problem, denoted NCP $(f)$, is to find a vector $x \in \mathbb{R}^n$ such that

$$x \geq 0, \quad f(x) \geq 0, \quad x^T f(x) = 0.$$  

The reader is referred to [5] for a comprehensive review of the theory and applications of this problem. In [27], the NE/SQP method was proposed as a solution procedure for this problem. Clearly, the NCP $(f)$ is equivalent to the following nonnegatively constrained system of nonsmooth equations:

$$H(x) := \min(x, f(x)) = 0, \quad x \geq 0,$$

where min denotes the componentwise minimum of two vectors. The NE/SQP method generates a sequence of nonnegative iterates by successively solving a sequence of nonnegatively constrained least-squares subproblems whose solutions provide descent directions for the merit function

$$\theta(x) = H(x)^T H(x).$$
Exploiting the idea of staying in the positive orthant, we describe an iterative algorithm for solving the NCP \((f)\) in which each direction-finding step requires the solution of a single system of linear equations.

It would be useful to summarize the key properties of the norm function \(\theta\). Clearly, \(\theta\) is nonnegative; its zeros are precisely the solutions of the NCP \((f)\). The function \(\theta\) is generally not Fréchet differentiable at an arbitrary vector, but it has a strong Fréchet derivative at all its zeros [25, Prop. 1]. Moreover, \(\theta\) is directionally differentiable everywhere with the directional derivative at a vector \(x\) along the direction \(d\) given by [24]

\[
\theta'(x, d) = \sum_{i:x_i<f_i(x)} x_id_i + \sum_{i:x_i=f_i(x)} x_i \min(d_i, \nabla f_i(x)^T d) + \sum_{i:x_i>f_i(x)} f_i(x) \nabla f_i(x)^T d.
\]

Motivated by this expression, we define three fundamental index sets associated with an arbitrary vector \(z \in \mathbb{R}^n\):

\[
I_e(z) = \{i : z_i < f_i(z)\}, \quad I_f(z) = \{i : z_i = f_i(z)\}, \quad I_f(z) = \{i : z_i > f_i(z)\}.
\]

For notational convenience, we let \(J_f(z) = I_e(z) \cup I_f(z)\). We note immediately that if \(x \in \mathbb{R}^n\) is nonnegative, then for all vectors \(d \in \mathbb{R}^n\),

\[
\theta'(x, d) \leq \sum_{i \in J_f(x)} x_id_i + \sum_{i \in I_f(x)} f_i(x) \nabla f_i(x)^T d.
\]

This inequality is the key to the descent step in the algorithm to be described later.

To write the inequality in a more compact form, we define the \(n \times n\) matrix \(A(x)\) whose \(i\)th column is given by

\[
A(x)_i = \begin{cases} 
  e^i & \text{if } i \in J_f(x), \\
  \nabla f_i(x) & \text{if } i \in I_f(x),
\end{cases}
\]

where \(e^i\) is the \(i\)th coordinate vector. In terms of this matrix, the above inequality becomes

\[(2) \quad \theta'(x, d) \leq H(x)^T A(x) d.
\]

It is important to note that the directional derivative \(\theta'(x, d)\) is generally not a continuous function in \(x\) for a fixed but arbitrary \(d\) and neither is the matrix-valued function \(A(x)\). However, the next result gives an important generalization of the inequality (2). A variant of this result can be found in [27, Lem. 4]. For the sake of completeness, we give a simpler and more direct proof of this result here.

**Proposition 2.1** Let \(x^* \in \mathbb{R}^n_+\) be arbitrary. Then for any sequence \(\{x^k\} \subseteq \mathbb{R}^n_+\) converging to \(x^*\) and any sequence \(\{h^k\} \subseteq \mathbb{R}^n \setminus \{0\}\) converging to \(0\), there holds

\[(3) \quad \limsup_{k \to \infty} \frac{\theta(x^k + h^k) - \theta(x^k) - H(x^k)^T A(x^k)^T h^k}{\|h^k\|} \leq 0.
\]

**Proof.** For each \(i = 1, \ldots, n\), let

\[
\begin{align*}
\Delta H_i(x^k, h^k) & \equiv \frac{1}{2} H_i^2(x^k + h^k) - \frac{1}{2} H_i^2(x^k) - H_i(x^k)[A(x^k)]_i^T h^k, \\
\Delta f_i(x^k, h^k) & \equiv \frac{1}{2} f_i^2(x^k + h^k) - \frac{1}{2} f_i^2(x^k) - f_i(x^k) \nabla f_i(x^k)^T h^k, \\
\Delta x_i(x^k, h^k) & \equiv \frac{1}{2} (x_i^k + h_i^k)^2 - \frac{1}{2} (x_i^k)^2 - x_i^k h_i^k.
\end{align*}
\]
We claim that for all $k$ sufficiently large,

\begin{equation}
\Delta H_i(x^k, h^k) \leq \max\{\Delta f_i(x^k, h^k), \Delta x_i(x^k, h^k)\}.
\end{equation}

Indeed, there are three cases to consider: whether $i \in I_f(x^*)$, $i \in I_x(x^*)$, or $i \in I_e(x^*)$. Assume first that $i \in I_f(x^*)$, that is $f_i(x^*) < x_i^*$. Using the fact that both sequences \(\{x^k\}\) and \(\{x^k + h^k\}\) converge to $x^*$ and a simple continuity argument, we obtain $f_i(x^k) < x_i^k$ and $f_i(x^k + h^k) < x_i^k + h_i^k$ for all $k$ sufficiently large. Using the definition of the functions $H(\cdot)$ and $A(\cdot)$, we then obtain

\[\Delta H_i(x^k, h^k) = \Delta f_i(x^k, h^k);\]

hence (4) holds. For the case in which $i \in I_x(x^*)$, that is $f_i(x^*) > x_i^*$, we can similarly show that

\[\Delta H_i(x^k, h^k) = \Delta x_i(x^k, h^k);\]

so (4) also holds. Consider now the case in which $i \in I_e(x^*)$, that is $f_i(x^*) = x_i^* > 0$. Then, for all $k$ sufficiently large, we have $f_i(x^k + h^k) > 0$ and $x_i^k + h_i^k > 0$. This implies that

\[H_i^2(x^k + h^k) = \min\{f_i^2(x^k + h^k), (x_i^k + h_i^k)^2\}.\]

Using this relation and considering whether $i \in I_f(x^k)$ or $i \in I_J(x^k)$, we can easily verify that (4) holds. We can now complete the proof of (3). Indeed, using (4) we obtain that

\[\limsup_{k \to \infty} \frac{\Delta H_i(x^k, h^k)}{\|h^k\|} \leq 0 \quad \forall i = 1, \ldots, n.\]

Since

\[\theta(x^k + h^k) - \theta(x^k) - H(x^k)^T A(x^k)^T h^k = \sum_{i=1}^{n} \Delta H_i(x^k, h^k),\]

relation (3) follows. \(\Box\)

3. Some important functions. The merit function to be used in the algorithm is defined as follows. Let $c > 0$ and $\zeta > n$ be given scalars. Define

\[\Omega \equiv \{x \in \mathbb{R}_+^n \mid \theta(x) > 0\}\]

and let

\[\psi_c(x) \equiv c \log \theta(x) + \zeta \log(\theta(x) + e^T x) - \sum_{i=1}^{n} \log x_i \quad \forall x \in \Omega,\]

where $e$ is the vector of all ones. The scalar $c$ is a penalty parameter that will be changed in the algorithm if it is deemed to be too small; unlike $c$, $\zeta$ is fixed throughout. The third term in the function $\psi_c$ is the logarithmic barrier function to prevent the iterates from reaching the boundary of the nonnegative orthant. The middle term is used to balance the third term; this is analogous to the potential function introduced in [30] for linear programs that have been used extensively in many primal-dual interior-point methods; see also [15] for a related merit function.
Clearly, we have
\[
\psi_c(x) \geq c \log \theta(x) + (\zeta - n) \log \theta(x) + e^T x \quad \forall x \in \Omega
\]
and
\[
\psi_c(x) \geq (c + \zeta) \log \theta(x) - \sum_{i=1}^{n} \log x_i \quad \forall x \in \Omega.
\]

These inequalities have two important implications that we summarize in the result below.

**Proposition 3.1.** For a fixed \(c > 0\), the following statements hold:

(a) If \(\{x^k\} \subseteq \Omega\) and \(\lim_{k \to \infty} \psi_c(x^k) = -\infty\), then \(\lim_{k \to \infty} \theta(x^k) = 0\);

(b) For any \(\alpha > 0\) and \(t \in \mathbb{R}\), there exist constants \(a > 0\) and \(b > 0\) such that

\[
[x \in \mathbb{R}^n_{++}, \psi_c(x) \leq t, \theta(x) \geq \alpha] \implies a \leq x_i \leq b \quad \forall i = 1, \ldots, n.
\]

The function \(\psi_c\) is directionally differentiable everywhere with the directional derivative at the vector \(x \in \Omega\) along the direction \(d \in \mathbb{R}^n\) given by

\[
\psi'_c(x, d) = \frac{c}{\theta(x)} \theta'(x, d) + \frac{\zeta}{\theta(x) + e^T x} (\theta'(x, d) + e^T d) - \sum_{i=1}^{n} \frac{d_i}{x_i}.
\]

Recalling the inequality (2), we define the forcing function

\[
\psi(x, d) = \frac{c}{\theta(x)} A(x)^T d + \frac{\zeta}{\theta(x)} (A(x)H(x) + e)^T d - \sum_{i=1}^{n} \frac{d_i}{x_i}
\]

whose role in the algorithm will become obvious momentarily. Clearly, by (2), we have

\[
\psi'_c(x, d) \leq \psi(x, d)
\]

for all \(x \in \mathbb{R}^n_{++}\), with \(\theta(x) > 0\) and all \(d \in \mathbb{R}^n\). Consequently, given such a vector \(x\), if we can find a vector \(d\) such that \(\psi(x, d) < 0\), then \(d\) is a descent direction for the function \(\psi_c\) at \(x\). To generate such a direction, we consider the system of linear equations

\[
(A(x)A(x)^T + X^{-2})d + w_c(x) = 0,
\]

where

\[
w_c(x) = \frac{c}{\theta(x)} A(x)H(x) + \frac{\zeta}{\theta(x)} (A(x)H(x) + e) - x^{-1},
\]

with \(X\) being the diagonal matrix with \(x_i\)'s on the diagonal and \(x^{-1}\) being the vector whose \(i\)th component is equal to \(1/x_i\). The system of linear equations (6) is equivalent to the least-squares problem

\[
\text{minimize} \quad (\|A(x)^T d\|_2^2 + \|X^{-1} d\|_2^2) + \psi(x, d),
\]

where the minimization is over all vectors \(d \in \mathbb{R}^n\) with the vector \(x\) fixed.
Noting that the matrix

\[
M(x) = A(x)A(x)^T + X^{-2}
\]
is symmetric positive definite, we let \(d_x\) be the unique solution of (6). Then we have

\[
z_c(x, d_x) = -(\|A(x)^T d_x\|^2 + \|X^{-1} d_x\|^2) \leq 0;
\]
moreover, \(z_c(x, d_x) < 0\) if and only if \(w_c(x) \neq 0\). Consequently, if \(w_c(x) \neq 0\), then \(d_x\) is a descent direction for the function \(\psi_c\) at the point \(x\). Nevertheless, if \(w_c(x)\) is equal to zero, then \(d_x = 0\) and we need to generate an alternate direction. In this case, we double the penalty constant \(c\) (actually, any scaling exceeding 1 suffices) and solve another system of the form (6) with the modified \(c\).

With a (nonzero) descent direction \(d_x\) successfully computed, we then perform a line search starting at \(x\) with the objective of decreasing the merit function \(\psi_c\) by a sufficient amount while preserving the positivity of the next iterate. Details of such a line search step can be found in [22]. The main procedure is then repeated with the new iterate replacing the old one if termination with regard to a prescribed rule still has not occurred.

We close this section with an immediate consequence of Proposition 2.1.

**Proposition 3.2.** Let \(x^* \in \Omega\) be arbitrary. Then for any sequence \(\{x^k\} \subseteq \mathbb{R}^n_+\), converging to \(x^*\), any sequence \(\{d^k\}\) with \(\{(X^k)^{-1}d^k\}\) bounded, and any sequence \(\{\lambda_k\}\) of positive scalars converging to zero, there holds

\[
\limsup_{k \to \infty} \frac{\psi_c(x^k + \lambda_k d^k) - \psi_c(x^k) - \lambda_k z_c(x^k, d^k)}{\lambda_k} \leq 0.
\]

**Proof.** Since \(\{(X^k)^{-1}d^k\}\) is bounded, \(\{\lambda_k\} \to 0\) and \(x^k + \lambda_k d^k = x^k(1 + \lambda_k d^k / x^k)\), it follows that \(x^k + \lambda_k d^k\) is positive for all \(k\) sufficiently large. Using the fact that

\[
\log(1 + s) = s + o(s), \quad \text{where } \lim_{s \to 0} \frac{o(s)}{s} = 0,
\]
we obtain that for each \(i = 1, \ldots, n,\)

\[
\lim_{k \to \infty} \frac{\log(x^k_i + \lambda_k d^k_i) - \log x^k_i - \lambda_k d^k_i / x^k_i}{\lambda_k} = \lim_{k \to \infty} \frac{\log(1 + \lambda_k d^k_i / x^k_i) - \lambda_k d^k_i / x^k_i}{\lambda_k} = 0.
\]

Using this limit, Proposition 2.1 and the definition of \(\psi_c(\cdot)\), we can now easily derive (8). \(\square\)

4. The positive algorithm. Summarizing the ideas outlined in the previous section, we now present the details of the long awaited algorithm for solving the NCP (f), where \(f\) is an arbitrary continuously differentiable function.

**Step 0.** (Initialization) Let \(\zeta > n, \delta > 0\), and \(\sigma, \alpha, \rho \in (0, 1)\) be given constants. Choose a scalar \(c_0 > 0\) and a vector \(x^0 > 0\) arbitrarily. Set \(k = 0\).

**Step 1.** (Direction generation) Compute \(w^k = w_{c_k}(x^k)\). If \(\|w^k\| \leq \delta\), set

\[
c_{k+1} = 2c_k \quad \text{and} \quad x^{k+1} = x^k.
\]
Replace \(k\) by \(k + 1\) and return to the beginning of this step. If \(\|w^k\| > \delta\), let \(d^k\) be the unique solution to the system of linear equations:

\[
M(x^k)d + w^k = 0.
\]
A POSITIVE ALGORITHM FOR THE NCP

**Step 2.** (Armijo line search) Determine the maximum stepsize

\[ \tau^0_k = \sup \{ \tau : x^k + \tau d^k \geq 0 \} \]

and let

\[ \hat{\tau}_k = \begin{cases} \alpha \tau^0_k & \text{if } \tau^0_k < \infty, \\ 1/\|d^k\| & \text{if } \tau^0_k = \infty. \end{cases} \]

Let \( m_k \) be the smallest nonnegative integer \( m \) such that

\[ \psi_{c_k}(x^k + \hat{\tau}_k \rho^m d^k) - \psi_{c_k}(x^k) < \sigma \hat{\tau}_k \rho^m z_{c_k}(x^k, d^k). \]

Set

\[ c_{k+1} = c_k \quad \text{and} \quad x^{k+1} = x^k + \hat{\tau}_k \rho^{m_k} d^k. \]

**Step 3.** (Termination check) If \( x^{k+1} \) fails a termination check (for example, if \( \theta(x^{k+1}) > \varepsilon \) for a prescribed tolerance \( \varepsilon > 0 \)), return to step 1 with \( k \) replaced by \( k + 1 \).

We make several remarks about the above algorithm. First, the use of the scalar \( \delta \) to guard against a zero vector \( w_{c_k}(x^k) \) is an extension of the discussion in the last section. As we shall see from the convergence analysis in §5, it is not enough to just check whether the vector \( w_{c_k}(x^k) \) is zero or not; we actually need to ensure that this vector is not too small in norm for the direction \( d^k \) to be useful. Second, the maximum stepsize \( \tau^0_k \) and the scalar \( \alpha \in (0, 1) \) together will ensure that the next iterate \( x^{k+1} \) remains a positive vector. Finally, a standard argument in an Armijo line search and the inequality (5) will ensure that the integer \( m_k \) can be determined in a finite number of trials; this proof is omitted. Moreover, in case \( m_k \geq 1 \), we must have

\[ \psi_{c_k}(x^k + \hat{\tau}_k \rho^{m_k-1} d^k) - \psi_{c_k}(x^k) \geq \sigma \hat{\tau}_k \rho^{m_k-1} z_{c_k}(x^k, d^k) \]

by the definition of \( m_k \).

It would be useful to compare the positive algorithm with the framework proposed in [15] for solving the NCP \((f)\). For this purpose, we consider this problem as being defined by the following conditions:

\[ y - f(x) = 0, \tag{12} \]

\[ x \geq 0, \tag{13} \]

\[ y \geq 0, \tag{14} \]

\[ x^T y = 0. \tag{15} \]

The positive algorithm generates a sequence of iterates \( \{x^k\} \), which induces a corresponding sequence \( \{y^k\} \) via the relation \( y^k = f(x^k) \) for all \( k \). Hence, the combined sequence \( \{(x^k, y^k)\} \) satisfies the conditions (12) and (13) but not necessarily (14) or (15); in fact, the latter two conditions are the goal of the positive algorithm. On the other hand, the methods described in [15] generate a sequence \( \{(x^k, y^k)\} \) that satisfies (13) and (14) but not necessarily (12) or (15), which is the goal of these other infeasible interior-point methods for solving the NCP \((f)\).
5. Convergence analysis. In this section, we analyze the limiting properties of an infinite sequence \( \{x^k\} \) generated by the positive algorithm. By the infinite nature of this sequence, we have \( \theta(x^k) > 0 \) for all \( k \). We divide the analysis into two cases, depending on whether the penalty constant \( c_k \) is updated infinitely often or only finitely many times. We first take up the latter case.

Finite update of \( c \). The next result analyzes the case in which \( \|w^k\| > \delta \) for all indices \( k \) sufficiently large.

**THEOREM 5.1.** Suppose that the penalty sequence \( \{c_k\} \) is updated finitely many times in the positive algorithm. Then,

\[
\lim_{k \to \infty} \theta(x^k) = 0.
\]

Consequently, every accumulation point of \( \{x^k\} \), if it exists, must be a solution of the NCP \( (f) \).

**Proof.** Since \( \{c_k\} \) is updated finitely many times, there exists an index \( k_0 > 0 \) and a constant \( c > 0 \) such that \( c_k = c \) for all \( k \geq k_0 \). Assume by contradiction that (16) does not hold. Then there exist a constant \( \varepsilon > 0 \) and a subsequence \( \{x^k\}_{k \in \mathcal{K}} \) such that

\[
\inf_{k \in \mathcal{K}} \theta(x^k) \geq \varepsilon.
\]

Inequality (10) and the fact that \( z_c(x^k, d^k) < 0 \), for all \( k \geq k_0 \), imply that \( \{\psi_c(x^k) : k \geq k_0\} \) is decreasing. Moreover, (17) and Proposition 3.1(a) imply that \( \{\psi_c(x^k) : k \in \mathcal{K}\} \) is bounded below. Hence, this sequence converges and, by (10), we have

\[
\lim_{k(\in \mathcal{K}) \to \infty} \tilde{r}_k \rho^m z_c(x^k, d^k) = 0.
\]

We next show that

\[
\inf_{k \in \mathcal{K}} \frac{|z_c(x^k, d^k)|}{\|d^k\|} > 0.
\]

Indeed, we know that \( \{x^k\}_{k \in \mathcal{K}} \subseteq \{x \in R^n_+ : \psi_c(x) \leq t, \theta(x) \geq \varepsilon\} \), where \( t \equiv \psi_c(x_0) \).

By Proposition 3.1(b), \( \{x^k\}_{k \in \mathcal{K}} \) is bounded and for all \( k \in \mathcal{K}, \)

\[
x^k_i \geq a, \quad i = 1, \ldots, n
\]

for some constant \( a > 0 \). Using this fact, we can easily show that

\[
\|M(x^k)\| \|M(x^k)^{-1}\| \leq T \quad \forall k \in \mathcal{K}
\]

for some constant \( T > 0 \). Hence, using the fact that \( \|w_c(x^k)\| \geq \delta \), we obtain that for all \( k \in \mathcal{K}, \)

\[
\frac{|z_c(x^k, d^k)|}{\|d^k\|} = \frac{|w_c(x^k)^T d^k|}{\|d^k\|} = \frac{\|w_c(x^k)^T M(x^k)^{-1} w_c(x^k)\|}{\|M(x^k)^{-1}\| \|w_c(x^k)\|} \geq \frac{\|w_c(x^k)\|}{\|M(x^k)^{-1}\| \|w_c(x^k)\|} \geq \frac{\|w_c(x^k)\|}{\|M(x^k)\| \|M(x^k)^{-1}\|} \geq \frac{\delta}{T}.
\]

Hence, (19) holds. Combining (18) and (19), we obtain

\[
\lim_{k(\in \mathcal{K}) \to \infty} \tilde{r}_k \rho^m \|d^k\| = 0.
\]
We next show that

\[ \inf_{k \in \mathcal{K}} \tilde{\tau}_k \|d^k\| > 0. \tag{23} \]

Indeed, if \( k \in \mathcal{K} \) is such that \( \tau^0_k = \infty \), then

\[ \tilde{\tau}_k \|d^k\| \geq 1 \]

by the definition of \( \tilde{\tau}_k \). On the other hand, if \( \tau^0_k < \infty \), then using (20) and the fact that \( x^k + \tau^0_k d^k \) has some component equal to zero, we can easily deduce that

\[ \tau^0_k \|d^k\| \geq a. \tag{24} \]

Consequently, (23) holds since \( \tilde{\tau}_k = \alpha \tau^0_k \) for all \( k \geq 0 \). Combining (22) and (23), we obtain

\[ \lim_{k \to \infty} \beta^k m_k^k = 0. \]

Hence, \( m_k \geq 1 \) for all \( k \in \mathcal{K} \) sufficiently large. Since \( \{x^k\}_{k \in \mathcal{K}} \) is bounded, we may take \( \mathcal{K}' \subseteq \mathcal{K} \) such that \( \lim_{k \to \infty} x^k = x^* \in R^{n_+}_+ \) and \( m_k \geq 1 \) for all \( k \in \mathcal{K}' \). By (11), we obtain

\[ \frac{\psi_c(x^k + \tilde{\tau}_k \rho_{m_k-1} d^k) - \psi_c(x^k)}{\tilde{\tau}_k \rho_{m_k-1} \|d^k\|} \geq \sigma \frac{z_c(x^k, d^k)}{\|d^k\|}, \]

or equivalently,

\[ \frac{\psi_c(x^k + \lambda_k p^k) - \psi_c(x^k) - \lambda_k w_c(x^k) T p^k}{\lambda_k} \geq -(1 - \sigma) \frac{z_c(x^k, d^k)}{\|d^k\|} \geq \frac{(1 - \sigma) \delta}{T}, \tag{25} \]

where \( \lambda_k \equiv \tilde{\tau}_k \rho_{m_k-1} \|d^k\| \) and \( p^k \equiv d^k/\|d^k\| \). Note that relation (22) implies that \( \lim_{k \to \infty} \lambda_k = 0 \). Also, it is easy to verify that \( \{(X^k)^{-1} p^k\} \) is bounded. Hence, by Proposition 3.2, we know that the lim sup of the left-hand side of (25) is nonpositive and this violates (25). We have thus obtained a contradiction and therefore (16) must hold.

We point out that other choices for the matrix \( M(x) \) used in the computation of the search direction are possible. In addition to the symmetry and positive definiteness of \( M(x) \), all that is required is that the condition number \( \text{cond}(M(x)) \) of the matrix \( M(x) \), defined as

\[ \text{cond}(M(x)) \equiv \|M(x)\| \|M(x)^{-1}\|, \]

be uniformly bounded on any compact subset of \( R^{n_+}_+ \); cf. (21). The above convergence proof of Theorem 5.1 (and thus the limit (16)) remains valid under this condition. Note that this condition is much weaker than the requirement that \( \text{cond}(M(x)) \) be uniformly bounded on the whole \( R^{n_+}_+ \). Our choice of \( M(x) \) in (7) satisfies the first condition but not the latter one.

It is important to point out that Theorem 5.1 is established under absolutely no assumption on the function \( f \) other than its continuous differentiability. Note also that the theorem does not require the boundedness of the sequence \( \{x^k\} \); indeed, as the following example shows, this sequence may be unbounded if no restriction is imposed on the function \( f \).

**Example.** Let

\[ f(x) = e^{-x}, \quad x \in R. \]
For $x > 0$ sufficiently large, it is easy to obtain the functions $H$ and $w_c$ as follows:

$$\begin{align*}
H(x) &= e^{-x} \\
w_c(x) &= -2c + \frac{1}{x^2} - x^{-1}.
\end{align*}$$

Note that $\lim_{x \to \infty} w_c(x) = -2c$; hence, provided that $2c > \delta$, we must have $|w_c(x)| > \delta$ for $x > 0$ sufficiently large. Consequently, corresponding to such an $x$, the search direction at $x$ is

$$d_x = \frac{-w_c(x)}{e^{-2x} + x^{-2}} > \frac{\delta}{e^{-2x} + x^{-2}} > 0.$$

Thus, if we initiate the positive algorithm with $x^0$ sufficiently large and the constant $c_0 > \delta/2$, the algorithm will generate an increasing sequence $\{x^k\}$ with $c_k$ remaining constant. This sequence cannot be bounded for otherwise its limit point would be a strictly positive solution of the NCP $(f)$; but the only solution to the NCP $(f)$ is $x = 0$.

**Boundedness of iterates.** The above example is not surprising because generally, if the NCP $(f)$ has no solution, then although the limiting value of the sequence $\{x(x^k)\}$ is zero, the sequence of iterates $\{x^k\}$ must be unbounded. Consequently, some conditions on $f$ must be needed for the latter sequence to be bounded. The following discussion pertains to this boundedness issue of $\{x^k\}$.

We recall some properties of a vector-valued mapping. A mapping $F : R^n \to R^n$ is norm-coercive on a set $X \subseteq R^n$ if

$$\lim_{\|x\| \to \infty, x \in X} \|F(x)\| = \infty;$$

$F$ is coercive on $X$ in the Hadamard sense if

$$\lim_{\|x\| \to \infty, x \in X} \frac{\|x * F(x)\|}{\|x\|} = \infty,$$

where $a * b$ denotes the Hadamard product of two vectors $a, b \in R^n$, i.e., $(a * b)_i = a_ib_i$ for all $i$. It is easy to see that a mapping $F$ is norm-coercive on $X$ if and only if for all $t \geq 0$, the level set

$$\{x \in X : \|F(x)\| \leq t\}$$

is bounded; moreover, coercivity in the Hadamard sense implies norm-coercivity. Examples of mappings that are coercive on $R^n_+$ in the Hadamard sense include the strongly copositive mappings and the uniform P-functions. The former are those mappings $F$ for which there exists a constant $\gamma_1 > 0$ such that

$$\max_{1 \leq i \leq n} x_i F_i(x) \geq \gamma_1 \|x\|^2$$

for all $x \in R^n_+$; and the latter are those mappings $F$ for which there exists a constant $\gamma_2 > 0$ such that

$$\max_{1 \leq i \leq n} (x_i - y_i)(F_i(x) - F_i(y)) \geq \gamma_2 \|x - y\|^2$$

for all $x, y \in R^n$.

Given a mapping $F : R^n \to R^n$, a **principal subfunction** of $F$ is defined as follows. For an arbitrary index set $\alpha \subseteq \{1, \ldots, n\}$ with cardinality $k$ and complement $\beta$ and an
arbitrary vector \( a_\beta \in \mathbb{R}^{n-k} \), the function \( G : \mathbb{R}^k \rightarrow \mathbb{R}^k \) defined by \( G(x_\alpha) = F_\alpha(x_\alpha, a_\beta) \) is a \( k \)-dimensional principal subfunction of \( F \).

**Corollary 5.2.** Suppose that the penalty parameter \( c_k \) is updated finitely many times in the positive algorithm. Then the sequence \( \{x^k\} \) is bounded under any one of the following conditions:

(a) There exists a scalar \( t > 0 \) such that the level set

\[
L_t := \{ x \in \mathbb{R}^n : 0(x) \leq t \}
\]

is bounded;

(b) \( f \) is \( (\text{globally}) \) Lipschitzian on \( \mathbb{R}^n \) and every \( k \)-dimensional principal subfunction \( f_\alpha(\cdot, a_\beta) \) of \( f \) is norm-coercive on \( \mathbb{R}^k_+ \) for every fixed vector \( a_\beta \in \mathbb{R}^{n-k}_+ \);

(c) \( f \) is \( (\text{globally}) \) Lipschitzian and coercive in the Hadamard sense on \( \mathbb{R}^n_+ \).

**Proof.** By Theorem 5.1, the sequence \( \{\theta(x^k)\} \) converges to zero. Hence, for all \( k \) sufficiently large, \( x^k \) is contained in the level set \( L_t \). Consequently, (a) implies the boundedness of \( \{x^k\} \). By the proof of [24, Lem. 4], it follows that (b) implies (a). Finally, we show that (c) implies (b). But this is an easy consequence of the identity

\[
a_\beta * f_\beta(x_\alpha, a_\beta) = a_\beta * f_\beta(a_\alpha, a_\beta) + a_\beta * (f_\beta(x_\alpha, a_\beta) - f_\beta(a_\alpha, a_\beta))
\]

and the Lipschitzian property of \( f \), which implies that

\[
\limsup_{\|x_\alpha\| \rightarrow \infty, x_\alpha \geq 0} \frac{\|a_\beta * (f_\beta(x_\alpha, a_\beta) - f_\beta(a_\alpha, a_\beta))\|}{\|x_\alpha\|} < \infty.
\]

Thus, by the coercivity of \( f \) on \( \mathbb{R}^n_+ \) in the Hadamard sense, it follows that the principal subfunction \( f_\alpha(\cdot, a_\beta) \) must be coercive on \( \mathbb{R}^k_+ \) in the Hadamard sense. As a consequence of an observation preceding this corollary, it follows that this subfunction is norm-coercive on \( \mathbb{R}^k_+ \). \( \square \)

Our next result concerns the LCP that corresponds to the NCP \((f)\) in which \( f \) is an affine mapping. The proof of this result requires a fundamental continuity property of the solution set of the LCP regarded as a multifunction of the constant vector of the problem [1, Thm. 7.2.1]. To explain the latter property, consider the LCP defined by the vector \( q \in \mathbb{R}^n \) and matrix \( M \in \mathbb{R}^{n \times n} \):

\[
x \geq 0, \quad q + Mx \geq 0, \quad x^T(q + Mx) = 0.
\]

We let \( \text{SOL}_M(q) \) denote the (possibly empty) solution set of this problem. As a multifunction in \( q \), \( \text{SOL}_M \) is locally upper Lipschitzian in the following sense: for a fixed but arbitrary \( q \), there exist a constant \( L > 0 \) and a neighborhood \( V \) of \( q \) such that for all \( q' \in V \),

\[
\text{SOL}_M(q') \subseteq \text{SOL}_M(q) + L\|q - q'\|_2 \mathcal{B},
\]

where \( \mathcal{B} \) denotes the (closed) unit ball in \( \mathbb{R}^n \) with the Euclidean norm. There are two immediate consequences of this result: one is that if \( \text{SOL}_M(q^k) \) is nonempty for a sequence of vectors \( \{q^k\} \) converging to \( q \), then \( \text{SOL}_M(q) \) is nonempty; moreover, if the latter solution set is bounded, then all solution sets \( \text{SOL}_M(q') \) with \( q' \) sufficiently close to \( q \) are uniformly bounded; see [1].

**Corollary 5.3.** Suppose that the penalty parameter \( c_k \) is updated finitely many times when the positive algorithm is applied to the LCP (26). Then \( \text{SOL}_M(q) \) must
be nonempty. Moreover, if this solution set is bounded, then there exists a constant $L' > 0$ such that for all $k$ sufficiently large,

$$d(x^k, \text{SOL}_M(q)) \leq L' \| \min(x^k, q + Mx^k) \|,$$

where $d(x, S)$ denotes the distance from the vector $x$ to the set $S$; in particular, the sequence $\{x^k\}$ is bounded.

**Proof.** Let $y^k = \min(x^k, q + Mx^k)$. Then Theorem 5.1 implies that the sequence $\{y^k\}$ converges to zero. The definition of $y^k$ implies that the vector $z^k = x^k - y^k \in \text{SOL}_M(q^k)$ where

$$q^k = q + My^k - y^k$$

which clearly converges to $q$. Hence, the desired conclusions follow easily from the aforementioned consequences of the locally upper Lipschitzian property of the solution set of an LCP.

**Remark.** The boundedness conclusion of the sequence $\{x^k\}$ in Corollary 5.3 does not follow from Corollary 5.2. The reason is that the solution set $\text{SOL}_M(q)$ is equal to the level set

$$\{x \in \mathbb{R}^n : \theta(x) \leq 0\} = \{x \in \mathbb{R}^n : \theta(x) = 0\},$$

which is different from any of the sets $L_t$ with $t \geq 0$. Clearly, the polyhedrality nature of the LCP has much to do with the validity of Corollary 5.3.

**Infinite update of $c$.** We return to the NCP $(f)$ and analyze the other case of the positive algorithm, namely, when $\|w_{c_k}(x^k)\| > \delta$ fails for infinitely many $k$. Our goal here is to demonstrate that if $x^*$ is the limit of any subsequence $\{x^k : k \in \kappa\}$ for which

$$\|w^k\| \leq \delta \quad \text{for all } k \in \kappa,$$

and if $x^*$ is an s-regular vector in the sense defined in [27], then $x^*$ solves the NCP $(f)$. For the sake of clarity, we review this concept in the definition below.

**DEFINITION.** A vector $x \geq 0$ is said to be $s$-regular if the following system of linear inequalities has a solution in the variable $d$:

$$x_i + d_i = 0 \quad \text{for } i \in I^\circ_-(x) \cup I^0_-(x),$$

$$f_i(x) + \nabla f_i(x)^T d = 0 \quad \text{for } i \in I^+_-(x),$$

$$x_i + d_i \geq 0 \quad \text{for } i \in I^0_-(x),$$

$$f_i(x) + \nabla f_i(x)^T d \geq 0 \quad \text{for } i \in I^+_+(x),$$

$$x_i + d_i \leq 0 \quad \text{for } i \in I^0_+(x),$$

$$f_i(x) + \nabla f_i(x)^T d \leq 0 \quad \text{for } i \in I^+_+(x),$$

where

$$I^+_-(x) = \{i : f_i(x) < x_i > 0\}, \quad I^+_+(x) = \{i : f_i(x) = x_i > 0\},$$

$$I^0_-(x) = \{i : f_i(x) = x_i = 0\}, \quad I^0_+(x) = \{i : f_i(x) = x_i = 0\}.$$

In [27, Prop. 3], a sufficient condition for a nonnegative vector $x$ to be $s$-regular was established in terms of certain matrix-theoretic properties of the Jacobian matrix.
\( \nabla f(x) \). We will postpone the discussion of this condition until the next subsection. In what follows, we proceed to establish a convergence result that complements Theorem 5.1. Given a subset \( X \subseteq \mathbb{R}^n \) and a point \( x \in X \), we recall the set of feasible directions at \( x \) with respect to \( X \):

\[
\mathcal{F}(x, X) \equiv \{ d \in \mathbb{R}^n \mid \exists \alpha > 0 \text{ such that } x + \alpha d \in X \text{ for all } \alpha \in [0, \bar{\alpha}] \}.
\]

**Lemma 5.4.** Suppose that the penalty sequence \( \{c_k\} \) is updated infinitely often and that \( x^* \) is the limit of a subsequence \( \{x^k : k \in \kappa\} \) for which

\[
\|w^k\| \leq \delta \quad \text{for all } k \in \kappa.
\]

Then, the sequence \( \{u^k : k \in \kappa\} \), where \( u^k \equiv A(x^k)H(x^k) \) has an accumulation point and any accumulation point \( u^\infty \) of this sequence satisfies

\[
d^T u^\infty \geq 0 \quad \forall d \in \mathcal{F}(x^*, R^+_n).
\]

**Proof.** Since \( x^k \xrightarrow{k \in \kappa} x^* \) and there is only a finite number of distinct index sets \( I_f(x^k) \), it is easy to see that the sequence \( \{u^k : k \in \kappa\} \) has an accumulation point. Assume that \( u^\infty \) is one such accumulation point and let us show that (27) holds. Indeed, let \( B \equiv \{i \mid x^*_i > 0\} \) and \( N \equiv \{i \mid x^*_i = 0\} \). Noting that \( \mathcal{F}(x^*, R^+_n) = \{d \in \mathbb{R}^n \mid d_N \geq 0\} \), we conclude that (27) is equivalent to the condition

\[
u^\infty_B = 0, \quad u^\infty_N \geq 0.
\]

We now show that (28) holds. Indeed, using the assumption that \( \|w^k\| \leq \delta \) for all \( k \in \kappa \) and the definition of \( w^k \), we obtain

\[
\left\|u^k + v^k - \frac{\theta(x^k)}{c_k}(x^k)^{-1}\right\| \leq \frac{\delta \theta(x^k)}{c_k} \quad \text{for all } k \in \kappa,
\]

where

\[
v^k \equiv \frac{\xi \theta(x^k)}{c_k(\theta(x^k) + e^T x^k)}(e + u^k) \quad \text{for all } k.
\]

Observe that the sequence \( v^k \xrightarrow{k \in \kappa} 0 \) since \( c_k \rightarrow \infty \). Hence, from the fact that \( x^k_B \xrightarrow{k \in \kappa} x_B^* > 0 \) and \( c_k \rightarrow \infty \), and relation (29), we obtain that \( u^\infty_B = \lim_{k \in \kappa} u^k_B = 0 \). Moreover, from (29) and the fact that \( x^k_N \xrightarrow{k \in \kappa} 0 \), we obtain that \( u^k_N + v^k_N \geq 0 \) for all \( k \in \kappa \) sufficiently large. Hence,

\[
u^\infty_N = \lim_{k \in \kappa}(u^k_N + v^k_N) \geq 0,
\]

and the result follows.

Observe that condition (27) can be viewed as a weak stationarity condition for the point \( x^* \). If the accumulation point \( x^* \) is nondegenerate; that is, it satisfies \( x^*_i \neq f_i(x^*) \) for all \( i = 1, \ldots, n \), then we can conclude that \( x^* \) is a stationary point for the function \( \theta(x) \) as the following corollary states.

**Corollary 5.5.** Let the assumptions of Lemma 5.4 hold and assume further that \( x^* \) is nondegenerate. Then, we have

\[
\theta'(x^*, d) \geq 0 \quad \text{for all } d \in \mathcal{F}(x^*, R^+_n).
\]
Proof. Using the fact that $x^*$ is nondegenerate, we conclude that $I_x(x^k) = I_x(x^*)$ and $I_f(x^k) = I_f(x^*)$ for all $k$ sufficiently large. It is now straightforward to see that the sequence $\{u^k : k \in \kappa\}$ converges and that its limit point $u^\infty$ satisfies $d^T u^\infty = \theta'(x^*, d)$ for all $d \in R^n$. The result now follows from Lemma 5.4.

The above two results say nothing about $x^*$ being a solution of NCP $(f)$ or, equivalently, that $x^*$ satisfies $\theta(x^*) = 0$. To conclude that $x^*$ is a solution of NCP $(f)$, we need to assume that $x^*$ is an $s$-regular vector. Before showing this fact, we state a preliminary result that gives several conditions that are related to $s$-regularity.

**Lemma 5.6.** Let $x \in R^n$ be given and for all $d \in R^n$, define

$$
g_x^{\min}(d) \equiv \theta'(x, d) = \sum_{i \in I_f(x)} f_i(x) \nabla f_i(x)^T d + \sum_{i \in I_e(x)} \min\{f_i(x) \nabla f_i(x)^T d, x_i d_i\} + \sum_{i \in I_s(x)} x_i d_i,
$$

and consider the following conditions on $x$:

(a) $x$ is an $s$-regular vector;

(b) there exists a $d \in \mathcal{F}(x, R^n)$ such that

$$
\max\{x_i [x_i + d_i], f_i(x) [f_i(x) + \nabla f_i(x)^T d]\} \leq 0 \quad \text{for all } i \in I_e(x),

$$

(c) there exist scalar $\gamma > 0$ and vector $d \in \mathcal{F}(x, R^n)$ such that

$$
\gamma \theta(x) + g_x(d) \leq 0;
$$

(d) if $\theta(x) > 0$ then there exists $d \in \mathcal{F}(x, R^n)$ such that $g_x(d) < 0$.

Then, the following implications hold:

$$(a) \implies (b) \implies (c) \iff (d).$$

**Proof.** The implication $(a) \implies (b)$ follows from the definition of $s$-regularity and the fact that $d \in \mathcal{F}(x, R^n)$ if and only if $d_i \geq 0$ for all $i$ such that $x_i = 0$. To show the second implication, assume that (b) holds. Summing the relations in (31) over all $i = 1, \ldots, n$, we obtain that

$$
\theta(x) + g_x^{\max}(d) \leq 0,
$$

which obviously implies (c) with $\gamma = 2$. The equivalence of (c) and (d) can be easily proved using the fact that $g_x$ is homogeneous of degree 1.

We are now ready to state the main result of this subsection.

**Theorem 5.7.** Suppose that the penalty sequence $\{c_k\}$ is updated infinitely often. If $x^*$ is the limit of a subsequence $\{x^k : k \in \kappa\}$ for which

$$
\|w^k\| \leq \delta \quad \text{for all } k \in \kappa
$$

then

$$
\lim_{k \to \infty} \theta(x^k) = 0.
$$
and \( x^* \) is \( s \)-regular, then \( \theta(x^*) = 0 \).

**Proof.** Assume for contradiction that \( \theta(x^*) \neq 0 \). Let \( u^\infty \) be an accumulation point of the sequence \( \{ u^k : k \in \kappa \} \) as defined in the statement of Lemma 5.4. Then, it is easily seen that the function \( g(x) \equiv d^T u^\infty \) satisfies the assumptions of Lemma 5.6. Hence, since \( x^* \) is an \( s \)-regular vector and \( \theta(x^*) \neq 0 \), it follows from Lemma 5.6(d) that there exists a vector \( d \in \mathcal{F}(x^*, \mathbb{R}^n_+) \) such that \( d^T u^\infty < 0 \). Since this contradicts Lemma 5.4, we must have \( \theta(x^*) = 0 \). \( \square \)

The class of \( s \)-regular functions. We may combine Theorems 5.1 and 5.7 to give a unifying convergence result for the positive algorithm. For this purpose, we define the following class of functions.

**Definition.** A function \( f : \mathbb{R}^n_+ \rightarrow \mathbb{R}^n \) is said to be \( s \)-regular if every nonnegative vector is \( s \)-regular.

A function \( f \) with the property that \( \nabla f(x) \) is a P-matrix for all \( x \geq 0 \) must be \( s \)-regular; in particular, any uniform P-function \( f \) on \( \mathbb{R}^n \) (and hence, any strongly monotone function) is \( s \)-regular. The proof of these observations hinges on the following result, which is a paraphrase of Proposition 3 in [27]. (The reader may want to consult [1] for discussion of the various matrix classes involved here.)

**Proposition 5.8.** Let \( x \) be a nonnegative vector. Suppose that (i) the principal submatrix

\[
\begin{bmatrix}
\nabla I^+_x f I^+_x (x) & \nabla I^+_x f I^+_x (x) - \nabla I^+_x f I^+_x (x) \\
\nabla I^+_x f I^+_x (x) & \nabla I^+_x f I^+_x (x) - \nabla I^+_x f I^+_x (x) \\
-\nabla I^+_x f I^+_x (x) & -\nabla I^+_x f I^+_x (x) + \nabla I^+_x f I^+_x (x)
\end{bmatrix}
\]

(32)

is nonsingular, and (ii) the Schur complement of this matrix in

\[
\begin{bmatrix}
\nabla I^+_x f I^+_x (x) & -\nabla I^+_x f I^+_x (x) \\
-\nabla I^+_x f I^+_x (x) & \nabla I^+_x f I^+_x (x)
\end{bmatrix}
\]

(33)

is an S-matrix.

(Here, the index sets are all evaluated at the given \( x \).) Then \( x \) is an \( s \)-regular vector.

We observe that if \( x \geq 0 \) is such that \( \nabla f(x) \) is a P-matrix, then conditions (i) and (ii) of Proposition 5.8 are satisfied. Indeed, if \( \nabla f(x) \) is a P-matrix then the matrices (32) and (33) are both P-matrices. In particular, (32) is nonsingular, showing that condition (i) holds. Moreover, since the Schur complement of any principal submatrix of a P-matrix is a P-matrix and since every P-matrix is an S-matrix, condition (ii) follows.

An interesting example of an \( s \)-regular function that is neither P nor monotone is the negative identity function. Indeed, if \( f(x) = -x \), then \( I^+_x (x) = I^+_x (x) = \emptyset \) for all \( x \geq 0 \). Consequently, any nonnegative vector is \( s \)-regular and \( f \) is an \( s \)-regular function. More generally, if \( f \) is a function for which these two index sets are empty for all nonnegative vectors and whose Jacobian matrix \( \nabla f(x) \) is nondegenerate for all \( x \geq 0 \), then \( f \) must be \( s \)-regular.

The following theorem is immediate from the previous results.

**Theorem 5.9.** If \( f \) is an \( s \)-regular function, then every accumulation point of a sequence of iterates produced by the positive algorithm is a solution of the NCP \( (f) \).

A referee of this paper correctly points out that the above theorem does not provide conditions under which a sequence of iterates produced by the positive algorithm will have at least one accumulation point. The difficulty with this deficiency of the
Theorem lies in the case where the scalar $c$ is updated infinitely often; indeed, in the present version of the algorithm, whenever $c_k$ is updated, we do not change the iterate $x^k$, and essentially, do nothing. It might be necessary to modify the algorithm to yield a more desirable result.

If the function $f$ has the property that the function $\theta$ has bounded level sets, then the only way for the sequence $\{x^k\}$ to have no accumulation point is that $\lim_{k \to \infty} \theta(x^k) = \infty$. Although we cannot rule out this possibility when $\{c_k\}$ is unbounded, it does seem rather unlikely to occur in practice. To substantiate this statement, we have implemented the positive algorithm for solving a variety of complementarity problems. The results are reported in the next section. In all the numerical tests we have conducted, the $\theta$-values at termination were consistently substantially smaller than the $\theta$-values at initiation, even in cases when the positive algorithm failed to solve a particular problem.

6. Numerical results. We have carried out some numerical experiments with the positive algorithm applied to two sets of complementarity problems: one experiment consists of NCPs arising from various equilibrium models that are documented in detail in [27]; the other is a set of randomly generated LCPs. For the first set of NCPs, we also compared the positive algorithm with the NE/SQP method described in [27] because the latter method was also based on the formulation (1) of the NCP and was highly successful in terms of robustness and speed. Our experiments show that the positive algorithm is faster than the NE/SQP method on the test problems but less robust. The improved speed is not surprising since in calculating each search direction, the positive algorithm solves only one system of linear equations whereas the NE/SQP method solves a convex quadratic programming with lower bound constraints; indeed, the simplicity of the direction generation is the single most important feature of the positive algorithm.

The positive algorithm was implemented in a FORTRAN-77 computer code and the experiments were conducted on a SUN SPARCStation IPX with 16 megabytes of memory and one CPU processor. Double precision arithmetic was employed in the calculations. In each iteration, we solved the system of linear equations (9) for the search direction by using the LU decomposition; the subroutines in [29] were used. We terminated the algorithm when the $\theta$-value was less than $10^{-12}$. The parameters of the algorithm were set as follows:

$$\zeta = n + 1, \quad \delta = 10^{-6}, \quad \sigma = \rho = 0.5, \quad \text{and} \quad \alpha = 0.95.$$ 

We set the initial penalty parameter $c_0 = 10^3$. The numerical results for the positive algorithm applied to the equilibrium problems are summarized in Table 1.

In Table 1, $n$ denotes the dimension of the NCP, niter the number of iterations, and aver.nls the average number of steps needed in the Armijo line search. In the column of niter, the numerators are the numbers of systems of linear equations solved by the positive algorithm, and the denominators are the numbers of quadratic programs solved by the NE/SQP method as reported in [27]. The column of $\theta$-values gives some indication of the speed of the positive algorithm at the tail of the iterations. In these runs, all the starting points were chosen to be the vector of ones, except for the Hansen-Koopmans problem where the NE/SQP method started from $(x_1, \ldots, x_{10}, y_1, y_2, z_1, z_2) = (0.3, \ldots, 0.3, 0, 0, 0, 0)$ while the positive algorithm started from $(x_1, \ldots, x_{10}, y_1, y_2, z_1, z_2) = (0.3, \ldots, 0.3, 0.1, 0.1, 0.1, 0.1)$; the reason for this deviation is that the positive algorithm must start from a positive vector.
The column of niter demonstrates the relative efficiency of the positive algorithm versus the NE/SQP method. Since one single system of linear equations was solved in the former algorithm, versus a quadratic program in the latter, the advantage of the positive algorithm in terms of speed should be evident. Indeed, we have compared these two algorithms on the largest of these problems, the traffic equilibrium problem, on the VAX 6000 computer at the Homewood Computing Facility Center at The Johns Hopkins University. The positive algorithm and the NE/SQP method used 3.94 and 8.35 CPU seconds, respectively. For the NE/SQP method, the FORTRAN package QPSOL [3] was used to solve each quadratic subprogram.

There are three test problems reported in [27] that were not included in the present experimentation. These are the PIES model, the Walrasian equilibrium problem with production, and the generalized von Thünen model. These problems are mixed NCPs as defined in [1]; the present version of the positive algorithm needs to be modified to deal with this class of problems. This is a topic for further study.

The set of LCPs to which the positive algorithm was applied can be classified into four types, each according to the properties of the matrix $M$. Specifically, the generation of $M$ was as follows. In each case, $M$ was completely dense.

1. $M = A^T A$. Each entry of the matrix $A$ was generated uniformly from the interval $(0, 50)$, with a probability of 0.5 for the entry to be given a negative sign. This matrix $M$ is symmetric positive semidefinite.

2. $M$ is diagonally dominant. We set $M = A$, where $A$ was generated in the same way as above except that each diagonal entry was set to be one plus the sum of the absolute values of the off-diagonal entries in the same row.

3. $M = A^2$. Here $A$ was generated as before. This matrix $M$ is indefinite.

4. $M$ is a positive matrix. Each entries of $M$ was generated uniformly from the interval $(1, 50)$.

For each matrix $M$ generated, we constructed a solvable LCP as follows. First we generated a random number from the interval $(0, 1)$; if this number was greater than 0.5, then we set $x_i^* = 0.0$, otherwise we generated a number from the interval $(0, 50)$ and set $x_i^*$ to be that number. Therefore, roughly half of the components of $x^*$ were zeros. We then formed the vector $q$ as follows. If $x^*_i > 0$, then $q_i = -M_{ii} x^*$, where $M_{ii}$ denotes the $i$th row of $M$; otherwise $q_i = -M_{ii} x^* + r$, where $r$ is a random number in $(0, 50)$ scaled by 0.3. Clearly the resulting LCP $(q, M)$ has at least one solution, namely, $x^*$. For $M$ of the first two types, $x^*$ is the unique solution.

After each LCP was generated, we scaled the problem in the following way: let $s$ be the sum of all entries of $M$ and $q$, we scaled $M$ and $q$ by the factor 50/$s$. The positive algorithm was then applied to the scaled LCP. It turned out that such a scaling was quite useful in ensuring the effectiveness of the algorithm.

The parameters of the algorithm were given the same values as before. We summarize the numerical results in Tables 2 and 3 for the cases $M = A^T A$ and $M$.

### Table 1

<table>
<thead>
<tr>
<th>Problem</th>
<th>$n$</th>
<th>niter</th>
<th>aver.nls</th>
<th>last three $\theta$—values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kojima-Shindo</td>
<td>4</td>
<td>8/7</td>
<td>2.000</td>
<td>8.01D–08/2.10D–12/5.05D–17</td>
</tr>
<tr>
<td>Nash–Cournot</td>
<td>10</td>
<td>9/9</td>
<td>1.000</td>
<td>4.62D–02/2.23D–06/6.07D–15</td>
</tr>
<tr>
<td>Hansen–Koopmans</td>
<td>14</td>
<td>22/10</td>
<td>1.667</td>
<td>1.18D–08/1.85D–12/2.01D–16</td>
</tr>
<tr>
<td>Spatial price</td>
<td>42</td>
<td>21/20</td>
<td>1.050</td>
<td>1.14D–11/1.57D–12/2.16D–13</td>
</tr>
</tbody>
</table>
Table 2

<table>
<thead>
<tr>
<th>Problem</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>niter</td>
<td>28</td>
<td>33</td>
<td>30</td>
<td>32</td>
<td>26</td>
</tr>
<tr>
<td>aver. nls</td>
<td>1.037</td>
<td>1.031</td>
<td>1.034</td>
<td>1.032</td>
<td>1.040</td>
</tr>
<tr>
<td>Initial θ</td>
<td>40308.4</td>
<td>35554.7</td>
<td>36078.4</td>
<td>43694.5</td>
<td>32757.5</td>
</tr>
<tr>
<td>Final θ</td>
<td>2.0D−14</td>
<td>3.8D−14</td>
<td>3.9D−14</td>
<td>1.5D−14</td>
<td>3.9D−13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>niter</td>
<td>41</td>
<td>35</td>
<td>23</td>
<td>31</td>
<td>25</td>
</tr>
<tr>
<td>aver. nls</td>
<td>1.025</td>
<td>1.029</td>
<td>1.045</td>
<td>1.033</td>
<td>1.041</td>
</tr>
<tr>
<td>Initial θ</td>
<td>31538.1</td>
<td>35382.4</td>
<td>29677.3</td>
<td>45676.6</td>
<td>31909.8</td>
</tr>
<tr>
<td>Final θ</td>
<td>1.0D−13</td>
<td>4.1D−14</td>
<td>6.5D−13</td>
<td>9.2D−15</td>
<td>7.0D−13</td>
</tr>
</tbody>
</table>

Table 3

<table>
<thead>
<tr>
<th>Problem</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>niter</td>
<td>11</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>aver. nls</td>
<td>1.300</td>
<td>1.125</td>
<td>1.222</td>
<td>1.444</td>
<td>1.300</td>
</tr>
<tr>
<td>Initial θ</td>
<td>46418.3</td>
<td>42027.2</td>
<td>47092.5</td>
<td>53837.2</td>
<td>36841.5</td>
</tr>
<tr>
<td>Final θ</td>
<td>1.6D−14</td>
<td>3.4D−13</td>
<td>2.6D−13</td>
<td>1.9D−14</td>
<td>3.1D−14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>niter</td>
<td>13</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>aver. nls</td>
<td>1.583</td>
<td>1.333</td>
<td>1.111</td>
<td>1.111</td>
<td>1.125</td>
</tr>
<tr>
<td>Initial θ</td>
<td>45390.3</td>
<td>44541.4</td>
<td>36948.6</td>
<td>48601.9</td>
<td>43534.8</td>
</tr>
<tr>
<td>Final θ</td>
<td>180D−14</td>
<td>2.3D−13</td>
<td>1.0D−14</td>
<td>2.5D−14</td>
<td>8.3D−13</td>
</tr>
</tbody>
</table>

diagonally dominant, respectively. The entries in the tables are self-explanatory (n is the dimension of M). As we have mentioned, the matrix M is completely dense; this is the reason why we have not attempted to solve problems of larger size in these two cases; in other words, data storage has imposed a restriction on our ability to use the Sun workstation for solving larger problems of this kind.

Observe that in the above two cases, all 10 problems in each group were successfully solved; more importantly, the computational statistics were very encouraging. In contrast, the results for the remaining two cases were not as good. In each of these cases, we ran LCPs of size n = 10, 20, 30. Ten problems were tested in each category. When M = A², the following results were obtained. For n = 10, seven problems were solved to satisfaction; for n = 20, one; and for n = 30, two. When M is positive, we obtained the following results. For n = 10, six problems were solved to satisfaction; for n = 20, four; and for n = 30, three. Invariably, when a successful run occurred, the results were good (i.e., small number of iterations, good speed at the tail, and small number of line searches). When an unsuccessful run occurred, it was due to the excessive number of iterations (80 was the maximum we set) and the small magnitude of the search directions; for these failed runs, the θ-values were consistently in the range of 10⁻⁴ and 10⁻⁸, which were small but not enough for successful termination according to our rule. There was good reason to believe that the iterates at termination of these unsuccessful runs were not s-regular vectors for the functions.

In summary, our computational results suggest that the positive algorithm holds promise in practice for solving complementarity problems that satisfy certain regularity conditions. For problems that do not necessarily satisfy the latter conditions, the algorithm requires further study and modification is needed.
7. Some concluding remarks. In this paper, we have presented and tested a positive algorithm for solving the general nonlinear complementarity problem. Some limiting properties of this algorithm were derived. At this time, although some theoretical issues remain with the algorithm, the computational experience we have gathered suggests that this algorithm is quite competitive with a previous algorithm on a set of equilibrium problems and has the potential for solving certain nonmonotone NCPs effectively.

REFERENCES


