Mathematical Programming manuscript No.
(will be inserted by the editor)

Michael C. Ferris* · Christian Kanzow** · Todd S. Munson*

Feasible Descent Algorithms for Mixed Complementarity Problems

March 31, 1998 [revised November 11, 1998]

Abstract. In this paper we consider a general algorithmic framework for solving nonlinear mixed complementarity problems. The main features of this framework are: (a) it is well-defined for an arbitrary mixed complementarity problem, (b) it generates only feasible iterates, (c) it has a strong global convergence theory, and (d) it is locally fast convergent under standard regularity assumptions. This framework is applied to the PATH solver in order to show viability of the approach. Numerical results for an appropriate modification of the PATH solver indicate that this framework leads to substantial computational improvements.

Key words. Mixed complementarity problems – global convergence – superlinear convergence – feasible descent methods

1. Introduction

In the past few years, interest in formulating and solving large scale nonlinear mixed complementarity problems (MCP) has become significant and continues to grow. Many theoretical algorithms have been postulated with numerical successes being reported for a variety of them on certain problem classes. Much current research focusses on nonsmooth Newton and smoothing-type methods, which typically have strong theoretical foundations and perform well on many test examples. On the other hand, the most widely used algorithms remain those based on successive linearization, which solve linear complementarity subproblems using generalizations of the Lemke pivoting algorithm. These linearization algorithms were proven locally quadratically convergent, and in practice seem to work extremely well. However, they typically require strong theoretical assumptions in order to guarantee that the subproblems are solvable.

The aim of this paper is twofold. First, we give a theoretical framework for the globalization of linearization and related algorithms. Second, we apply

Michael Ferris: University of Wisconsin – Madison, Computer Sciences Department, 1210 West Dayton Street, Madison, WI 53706, USA, e-mail: `ferris@cs.wisc.edu`

Christian Kanzow: University of Hamburg, Institute of Applied Mathematics, Bundesstrasse 55, D-20146 Hamburg, Germany, e-mail: `kanzow@math.uni-hamburg.de`

Todd Munson: University of Wisconsin – Madison, Computer Sciences Department, 1210 West Dayton Street, Madison, WI 53706, USA, e-mail: `tmunson@cs.wisc.edu`

* The research of this author was partially supported by National Science Foundation Grants CCR-9619765 and CDA-9726385.

** The research of this author was supported by the DFG (Deutsche Forschungsgemeinschaft).

our theory to one particular linearization method, namely the PATH solver, to demonstrate viability. In fact, our modification of the PATH code has better numerical behaviour than all previous versions of this code.

Merit functions are used extensively in the development of globalization theory and the implementation of robust algorithms. Broadly speaking, merit functions summarize how close the current iterate is to a solution of the problem under consideration with a single number. In complementarity problems and nonlinear systems of equations, the merit functions are normally nonnegative, and zero precisely at a solution to the original problem. Each merit function is typically used in a globalization strategy that involves searching between the current iterate and the Newton point (the solution of the linearization).

The classical example of a merit function in nonlinear equation solving is the square of the two-norm residual that measures the sum of squares of the errors in satisfying the equations. This merit function has one additional property to those listed above: namely, it is everywhere differentiable provided that the equation itself is everywhere differentiable.

In complementarity, the two classical merit functions are based on the natural residual [25] and the normal map [32]. Both the natural residual and the normal map provide reformulations of the complementarity problem as a system of equations; unfortunately, the systems and corresponding residual merit functions are nonsmooth. Even with this drawback, Ralph [30] showed how to construct an extension of the line search procedure for smooth nonlinear equations that enables fast local convergence of linearization methods under conditions that are exact generalizations of those required in smooth systems. This procedure has been implemented and successfully used in the PATH code [10,15].

A key implementational difficulty remains what to do when the linearization subproblem has no solution. The theory assumes this situation does not happen. In practice, this occurs frequently, particularly if the user of the code does not provide a good initial starting point. In nonlinear equations, an algorithm can resort to taking a steepest descent direction for the merit function, guaranteeing progress toward a stationary point of the merit function. Since the merit function used in the PATH solver is not guaranteed to be differentiable, heuristics need to be implemented to overcome these cases. While these heuristics are quite successful in practice, this situation is nonetheless unsatisfactory and prone to failure. This paper is an attempt to provide a theoretically justifiable escape mechanism by using a completely different merit function for the mixed complementarity problem in conjunction with the direction generated by a linearization method.

In order to outline the details of the paper, we first recall the definition of the mixed complementarity problem. Let $l_i \in \mathbb{R} \cup \{-\infty\}$ and $u_i \in \mathbb{R} \cup \{\infty\}$ be given lower and upper bounds with $l_i < u_i$ for all $i \in I$, where I is used throughout this paper to denote the index set $\{1, \dots, n\}$. Let l and u be the n-dimensional vectors with components l_i and u_i and assume that $F : [l, u] \to \mathbb{R}^n$ is a given function, continuously differentiable in a neighbourhood of the feasible set $[l, u]$. The mixed complementarity problem consists of finding a vector $x^* \in [l, u]$ such that exactly one of the following holds:

$$x_i^* = l_i \text{ and } F_i(x^*) > 0,$$
$$x_i^* = u_i \text{ and } F_i(x^*) < 0,$$
$$x_i^* \in [l_i, u_i] \text{ and } F_i(x^*) = 0.$$

The first component of this paper, described in Section 2, is a reformulation of the mixed complementarity problem, based on the Fischer-Burmeister function [17]. This results in an equivalent nonsmooth system of equations $\Phi(x) = 0$ where the corresponding merit function

$$\Psi(x) := \frac{1}{2}\Phi(x)^T\Phi(x) = \frac{1}{2}\|\Phi(x)\|^2$$

is continuously differentiable. This fact, along with other pertinent properties of Ψ is the subject of Section 3.

In the special case of nonlinear complementarity problems, both Φ and Ψ are used to design unconstrained algorithms for the solution of the problem. Unfortunately, in many practical situations, the imposed bounds, l and u, on the variables of the mixed complementarity problem are important not only for the problem definition but also because the complementarity function F (or its derivative) may not be defined outside of these bounds. For example, applications that include fractional powers can cause severe difficulties if the function is evaluated outside the feasible region.

The basic algorithmic framework of this paper does not consider Φ directly, but instead attempts to solve the bound constrained optimization reformulation

$$\min \Psi(x) \quad \text{s.t.} \quad x \in [l, u].$$

Section 3 also shows that a constrained stationary point is already a solution of the mixed complementarity problem under exactly the same assumptions that are used in order to prove a similar result for unconstrained stationary points of Ψ.

In Section 4 we present our algorithmic framework and prove that it is well-defined as well as globally and locally fast convergent under very weak assumptions. The theory only assumes a pre-existing feasible method that is locally well-defined and superlinearly convergent; it is not limited to linearization methods or to the PATH solver. Our algorithmic framework generates iterates that lie within the bounds, resorting to a projected gradient step for the bound constrained problem whenever the pre-existing method fails to provide sufficient decrease. We show, however, that our method remains locally superlinearly convergent.

To demonstrate the practicality of our theory, we give a brief description of an implementation and computational results for a particular instance of our class of algorithms in Section 5. The implementation is based on a modification of the PATH solver; the results indicate that the modified PATH solver is more robust than all previous versions of PATH.

Before proceeding, we give a few words about our notation. If F is any vector-valued function, we denote its Jacobian at a point x by $F'(x)$ and let $\nabla F(x)$ signify the transposed Jacobian. The gradient of a real-valued function f will be denoted by ∇f and will always be viewed as a column vector.

2. Equation Reformulation of MCP

In this section, we first define the mapping Φ, which produces a reformulation of the mixed complementarity problem as a nonlinear system of equations $\Phi(x) = 0$, and then investigate the properties of this mapping. These properties are extensions of some known ones for the standard nonlinear complementarity problem where $l_i = 0$ and $u_i = +\infty$ for all $i \in I$ (see, in particular, [9,14]). Our generalizations will be important in the analysis of subsequent sections.

Let us first define the mapping $\phi : \mathbb{R}^2 \to \mathbb{R}$ by

$$\phi(a, b) := \sqrt{a^2 + b^2} - a - b.$$

This function was introduced by Fischer [17] (and attributed to Burmeister) and is widely used in the context of nonlinear complementarity problems. The function's most interesting property is the fact that it is an NCP-function, i.e.,

$$\phi(a, b) = 0 \Longleftrightarrow a \geq 0,\ b \geq 0,\ ab = 0.$$

To use this NCP-function for the solution of the more general mixed complementarity problem, we first introduce a partition of I:

$$\begin{aligned} I_l &:= \{i \in I \mid -\infty < l_i < u_i = +\infty\}, \\ I_u &:= \{i \in I \mid -\infty = l_i < u_i < +\infty\}, \\ I_{lu} &:= \{i \in I \mid -\infty < l_i < u_i < +\infty\}, \\ I_f &:= \{i \in I \mid -\infty = l_i < u_i = +\infty\}, \end{aligned}$$

i.e., I_l, I_u, I_{lu} and I_f denote the set of indices $i \in I$ where there are finite lower bounds only, finite upper bounds only, finite lower and upper bounds and no finite bounds on the variable x_i, respectively.

We now follow an idea of Billups [2,3] and define the operator $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ componentwise as follows:

$$\Phi_i(x) := \begin{cases} \phi(x_i - l_i, F_i(x)) & \text{if } i \in I_l, \\ -\phi(u_i - x_i, -F_i(x)) & \text{if } i \in I_u, \\ \phi(x_i - l_i, \phi(u_i - x_i, -F_i(x))) & \text{if } i \in I_{lu}, \\ -F_i(x) & \text{if } i \in I_f. \end{cases}$$

The reader may ask why we use the minus sign for indices $i \in I_u$ or $i \in I_f$. In fact, all results of this paper would be true without the minus sign. However, the following lemma motivates why we use it in the definition of Φ_i for $i \in I_u \cup I_f$. A different motivation has recently been given by Billups [3].

Lemma 1. *Let $x \in \mathbb{R}^n$ be fixed. Then the following limits hold:*

(a) $\lim_{l_i \to -\infty} \phi(x_i - l_i, \phi(u_i - x_i, -F_i(x))) = -\phi(u_i - x_i, -F_i(x))$.
(b) $\lim_{u_i \to \infty} \phi(x_i - l_i, \phi(u_i - x_i, -F_i(x))) = \phi(x_i - l_i, F_i(x))$.
(c) $\lim_{l_i \to -\infty} \lim_{u_i \to \infty} \phi(x_i - l_i, \phi(u_i - x_i, -F_i(x))) = -F_i(x)$.

Proof. Let $\{a^k\} \subseteq \mathbb{R}$ be any sequence converging to ∞ and let $b \in \mathbb{R}$ be any fixed number. Then

$$\begin{aligned}
\phi(a^k, b) &= \sqrt{(a^k)^2 + b^2} - a^k - b \\
&= \frac{\left(\sqrt{(a^k)^2 + b^2} - (a^k + b)\right)\left(\sqrt{(a^k)^2 + b^2} + (a^k + b)\right)}{\sqrt{(a^k)^2 + b^2} + (a^k + b)} \\
&= \frac{-2a^k b}{\sqrt{(a^k)^2 + b^2} + (a^k + b)} \\
&= \frac{-2b}{\sqrt{1 + (b/a^k)^2} + 1 + b/a^k} \\
&\to -b.
\end{aligned}$$

From this observation, the three statements follow immediately by simple continuity arguments. □

To prove the following characterization of the mixed complementarity problem is straightforward. The proof is a simple extension of that given in Billups [2, Proposition 3.2.7].

Proposition 1. *$x^* \in \mathbb{R}^n$ is a solution of the mixed complementarity problem if and only if x^* solves the nonlinear system of equations $\Phi(x) = 0$.*

The function Φ is not differentiable everywhere. However, it is locally Lipschitzian and therefore has a nonempty generalized Jacobian in the sense of Clarke [8]. We next present an overestimation of this generalized Jacobian (see Billups [2, Lemma 3.2.10]).

Proposition 2. *We have*

$$\partial \Phi(x)^T \subseteq \{D_a(x) + \nabla F(x) D_b(x)\},$$

where $D_a(x) \in \mathbb{R}^{n\times n}$ and $D_b(x) \in \mathbb{R}^{n\times n}$ are diagonal matrices whose diagonal elements are defined as follows:

(a) If $i \in I_l$, then if $(x_i - l_i, F_i(x)) \neq (0,0)$,

$$(D_a)_{ii}(x) = \frac{x_i - l_i}{\|(x_i - l_i, F_i(x))\|} - 1,$$
$$(D_b)_{ii}(x) = \frac{F_i(x)}{\|(x_i - l_i, F_i(x))\|} - 1$$

but if $(x_i - l_i, F_i(x)) = (0,0)$,

$$((D_a)_{ii}(x), (D_b)_{ii}(x)) \in \{(\xi - 1, \rho - 1) \in \mathbb{R}^2 \,|\, \|(\xi, \rho)\| \le 1\}.$$

(b) If $i \in I_u$*, then if* $(u_i - x_i, -F_i(x)) \neq (0,0)$,

$$(D_a)_{ii}(x) = \frac{u_i - x_i}{\|(u_i - x_i, -F_i(x))\|} - 1,$$
$$(D_b)_{ii}(x) = \frac{-F_i(x)}{\|(u_i - x_i, -F_i(x))\|} - 1$$

but if $(u_i - x_i, -F_i(x)) = (0,0)$,

$$((D_a)_{ii}(x), (D_b)_{ii}(x)) \in \{(\xi - 1, \rho - 1) \in \mathbb{R}^2 \,|\, \|(\xi, \rho)\| \leq 1\}.$$

(c) If $i \in I_{lu}$*, then*

$$(D_a)_{ii}(x) = a_i(x) + b_i(x)c_i(x), \qquad (D_b)_{ii}(x) = b_i(x)d_i(x).$$

Here, if $(x_i - l_i, \phi(u_i - x_i, -F_i(x))) \neq (0,0)$,

$$a_i(x) = \frac{x_i - l_i}{\|(x_i - l_i, \phi(u_i - x_i, -F_i(x)))\|} - 1,$$
$$b_i(x) = \frac{\phi(u_i - x_i, -F_i(x))}{\|(x_i - l_i, \phi(u_i - x_i, -F_i(x)))\|} - 1$$

but if $(x_i - l_i, \phi(u_i - x_i, -F_i(x))) = (0,0)$,

$$(a_i(x), b_i(x)) \in \{(\xi - 1, \rho - 1) \in \mathbb{R}^2 \,|\, \|(\xi, \rho)\| \leq 1\}.$$

Further, if $(u_i - x_i, -F_i(x)) \neq (0,0)$*, then*

$$c_i(x) = \frac{x_i - u_i}{\|(u_i - x_i, -F_i(x))\|} + 1, \quad d_i(x) = \frac{F_i(x)}{\|(u_i - x_i, -F_i(x))\|} + 1$$

but if $(u_i - x_i, -F_i(x)) = (0,0)$,

$$(c_i(x), d_i(x)) \in \{(\xi + 1, \rho + 1) \in \mathbb{R}^2 \,|\, \|(\xi, \rho)\| \leq 1\}.$$

(d) If $i \in I_f$*, then* $(D_a)_{ii}(x) = 0$, $(D_b)_{ii}(x) = -1$.

Note that the statement of Proposition 2 is rather lengthy because we have to take into account the definition of Φ using the four different index sets I_l, I_u, I_{lu} and I_f; its proof is straightforward and therefore omitted here. However, Proposition 2 is extremely important for our subsequent analysis and will be used several times within the proofs of some important results established in this and the next section.

The remainder of this section is devoted to proving Theorem 1. The motivation for this result is to establish the local convergence of the algorithm that we propose in Section 4. To that end, let $x^* \in \mathbb{R}^n$ be a solution of the mixed complementarity problem and let us introduce another partition of I:

$$\alpha := \{i \,|\, l_i < x_i^* < u_i, F_i(x^*) = 0\},$$
$$\beta := \{i \,|\, x_i^* \in \{l_i, u_i\}, F_i(x^*) = 0\},$$
$$\gamma := \{i \,|\, x_i^* \in \{l_i, u_i\}, F_i(x^*) \neq 0\}.$$

Then we obtain the following result as a simple consequence of Proposition 2.

Lemma 2. *Let $x^* \in \mathbb{R}^n$ be a solution of the mixed complementarity problem. Furthermore, let $H \in \partial\Phi(x^*)$ be any fixed matrix, $H = D_a(x^*) + D_b(x^*)F'(x^*)$ with diagonal matrices $D_a(x^*)$ and $D_b(x^*)$ as specified in Proposition 2. Then these diagonal matrices have the following properties:*

(a) $(D_a)_{ii}(x^*) = 0$ *and* $(D_b)_{ii}(x^*) = -1$ *for all* $i \in \alpha$.
(b) $(D_a)_{ii}(x^*) \le 0, (D_b)_{ii}(x^*) \le 0$, *and* $(D_a)_{ii}(x^*) + (D_b)_{ii}(x^*) < 0$ *for all* $i \in \beta$.
(c) $(D_a)_{ii}(x^*) = -1$ *and* $(D_b)_{ii}(x^*) = 0$ *for all* $i \in \gamma$.

Proof. If $i \in \alpha$, then we immediately obtain statement (a) from Proposition 2 by considering the four possible cases $i \in I_l, i \in I_u, i \in I_{lu}$ and $i \in I_f$ separately.

Next consider statement (c), i.e., assume that $i \in \gamma$. Then we either have $x_i^* = l_i$ and $F_i(x^*) > 0$ or we have $x_i^* = u_i$ and $F_i(x^*) < 0$.

First assume that $x_i^* = l_i$ and $F_i(x^*) > 0$. Then the index i necessarily belongs to I_l or to I_{lu}. If $i \in I_l$, we obtain from Proposition 2 that $(D_a)_{ii}(x^*) = -1$ and $(D_b)_{ii}(x^*) = 0$. On the other hand, if $i \in I_{lu}$, we get from Proposition 2, together with the observation that $\phi(a,b) > 0$ outside the nonnegative orthant, that

$$(D_a)_{ii}(x^*) = a_i(x^*) + b_i(x^*)c_i(x^*) = -1 + 0 \cdot c_i(x^*) = -1$$

and

$$(D_b)_{ii}(x^*) = b_i(x^*)d_i(x^*) = 0 \cdot d_i(x^*) = 0.$$

The case $x_i^* = u_i$ and $F_i(x^*) < 0$ can be proven in a similar manner. Furthermore, statement (b) also follows by using an identical argument. □

We next restate a useful characterization of Robinson's [31] strong regularity condition in the context of mixed complementarity problems. A proof may be found in [13]. We stress that, in the case of a nonlinear complementarity problem (i.e., $l_i = 0$ and $u_i = \infty$ for all $i \in I$), this characterization reduces to a standard characterization from Robinson [31].

Proposition 3. *The following two statements are equivalent:*

(a) x^* *is a strongly regular solution of the mixed complementarity problem.*
(b) The submatrix $F'(x^*)_{\alpha\alpha}$ *is nonsingular, and the Schur-complement*

$$F'(x^*)_{\alpha\cup\beta,\alpha\cup\beta}(x^*)/F'(x^*)_{\alpha\alpha} := F'(x^*)_{\beta\beta} - F'(x^*)_{\beta\alpha}F'(x^*)_{\alpha\alpha}^{-1}F'(x^*)_{\alpha\beta}$$

is a P-matrix.

In order to establish a nonsingularity result for the generalized Jacobian $\partial\Phi(x^*)$ at a strongly regular solution of the mixed complementarity problem, we also need the following result whose proof can be found in [22, Proposition 2.7]; see [19] for several extensions.

Proposition 4. *A matrix of the form*

$$D_a + D_b M$$

is nonsingular for all negative semidefinite diagonal matrices $D_a, D_b \in \mathbb{R}^{m\times m}$ such that $D_a + D_b$ is negative definite if and only if $M \in \mathbb{R}^{m\times m}$ is a P-matrix.

Based on the previous results, we are able to prove the main result of this section.

Theorem 1. *If x^* is a strongly regular solution of the mixed complementarity problem, then all elements $H \in \partial\Phi(x^*)$ are nonsingular.*

Proof. Let $H \in \partial\Phi(x^*)$. By Proposition 2, there exist diagonal matrices $D_a(x^*), D_b(x^*) \in \mathbb{R}^{n\times n}$ such that

$$H = D_a(x^*) + D_b(x^*)F'(x^*). \tag{1}$$

Hence, if we write

$$D_a(x^*) = \begin{pmatrix} (D_a)_{\alpha\alpha}(x^*) & 0 & 0 \\ 0 & (D_a)_{\beta\beta}(x^*) & 0 \\ 0 & 0 & (D_a)_{\gamma\gamma}(x^*) \end{pmatrix},$$

$$D_b(x^*) = \begin{pmatrix} (D_b)_{\alpha\alpha}(x^*) & 0 & 0 \\ 0 & (D_b)_{\beta\beta}(x^*) & 0 \\ 0 & 0 & (D_b)_{\gamma\gamma}(x^*) \end{pmatrix}$$

and

$$F'(x^*) = \begin{pmatrix} F'(x^*)_{\alpha\alpha} & F'(x^*)_{\alpha\beta} & F'(x^*)_{\alpha\gamma} \\ F'(x^*)_{\beta\alpha} & F'(x^*)_{\beta\beta} & F'(x^*)_{\beta\gamma} \\ F'(x^*)_{\gamma\alpha} & F'(x^*)_{\gamma\beta} & F'(x^*)_{\gamma\gamma} \end{pmatrix}$$

and if we take into account Lemma 2, the homogeneous linear system $Hd = 0$ can be rewritten as

$$F'(x^*)_{\alpha\alpha}d_\alpha + F'(x^*)_{\alpha\beta}d_\beta + F'(x^*)_{\alpha\gamma}d_\gamma = 0_\alpha, \tag{2}$$

$$(D_a)_{\beta\beta}(x^*)d_\beta + (D_b)_{\beta\beta}(x^*)\left[F'(x^*)_{\beta\alpha}d_\alpha + F'(x^*)_{\beta\beta}d_\beta + F'(x^*)_{\beta\gamma}d_\gamma\right] = 0_\beta, \tag{3}$$

$$-d_\gamma = 0_\gamma. \tag{4}$$

Since $d_\gamma = 0$ by (4) and $F'(x^*)_{\alpha\alpha}$ is nonsingular by assumption and Proposition 3, we obtain from (2):

$$d_\alpha = -F'(x^*)_{\alpha\alpha}^{-1}F'(x^*)_{\alpha\beta}d_\beta. \tag{5}$$

Substituting (4) and (5) into (3), we obtain after some rearrangements:

$$\left[(D_a)_{\beta\beta}(x^*) + (D_b)_{\beta\beta}(x^*)(F'(x^*)_{\alpha\cup\beta,\alpha\cup\beta}/F'(x^*)_{\alpha\alpha})\right]d_\beta = 0_\beta. \tag{6}$$

Since the Schur complement $F'(x^*)_{\alpha\cup\beta,\alpha\cup\beta}/F'(x^*)_{\alpha\alpha}$ is a P-matrix by assumption and Proposition 3 and since, by Lemma 2 (b), the diagonal matrices $(D_a)_{\beta\beta}(x^*)$ and $(D_b)_{\beta\beta}(x^*)$ are negative semidefinite with a negative definite sum, it follows from Proposition 4 that the coefficient matrix in (6) is nonsingular. Hence we obtain $d_\beta = 0_\beta$. This, in turn, implies $d_\alpha = 0_\alpha$ by (5). Reference to (4) shows that $d = 0$, so that H is nonsingular. □

3. Smooth Merit Function for MCP

We now investigate the properties of the residual merit function

$$\Psi(x) = \frac{1}{2}\Phi(x)^T\Phi(x)$$

associated with the equation operator Φ.

Despite the fact that Φ is nondifferentiable in general, it turns out that the merit function Ψ is continuously differentiable everywhere. More precisely, we have the following result.

Proposition 5. *The function Ψ is continuously differentiable with gradient $\nabla\Psi(x) = H^T\Phi(x)$ for an arbitrary $H \in \partial\Phi(x)$.*

Proof. The proof is essentially the same as that given for Proposition 3.4 by Facchinei and Soares [14] for the special case of a nonlinear complementarity problem. □

We next provide a stationary point result for the unconstrained reformulation

$$\min \Psi(x), \qquad x \in \mathbb{R}^n,$$

of the mixed complementarity problem. To this end, we need the following characterization of P_0-matrices, see [7] as well as [19] for some generalizations (note the difference between this result and the related statement in Proposition 4).

Proposition 6. *A matrix of the form*

$$D_a + D_b M$$

is nonsingular for all negative definite diagonal matrices $D_a, D_b \in \mathbb{R}^{m\times m}$ if and only if $M \in \mathbb{R}^{m\times m}$ is a P_0-matrix.

Proposition 6 enables us to prove the first major result of this section. In this result and in the remaining part of this section, we use the short-hand notation $\nabla F(x^*)_{ff}$ to denote the submatrix $\nabla F(x^*)_{I_f I_f}$. A similar notation is used for submatrices and subvectors defined by other index sets.

Theorem 2. *Let $x^* \in \mathbb{R}^n$ be a stationary point of Ψ. Assume that*

(a) the principal submatrix $\nabla F(x^*)_{ff}$ is nonsingular, and
(b) the Schur complement $\nabla F(x^*)/\nabla F(x^*)_{ff}$ is a P_0-matrix.

Then x^* is a solution of the mixed complementarity problem.

Proof. Let x^* be a stationary point of Ψ. Then, by Proposition 5, we have

$$H^T\Phi(x^*) = \nabla\Psi(x^*) = 0 \tag{7}$$

for an arbitrary $H \in \partial\Phi(x^*)$. By Proposition 2, there exist diagonal matrices $D_a(x^*), D_b(x^*) \in \mathbb{R}^{n\times n}$ such that

$$H = D_a(x^*) + D_b(x^*)F'(x^*).$$

Therefore, (7) becomes

$$[D_a(x^*) + \nabla F(x^*)D_b(x^*)]\,\Phi(x^*) = 0. \tag{8}$$

Writing

$$D_a(x^*) = \begin{pmatrix} (D_a)_{ff}(x^*) & 0 \\ 0 & (D_a)_{\bar f\bar f}(x^*) \end{pmatrix},$$

$$D_b(x^*) = \begin{pmatrix} (D_b)_{ff}(x^*) & 0 \\ 0 & (D_b)_{\bar f\bar f}(x^*) \end{pmatrix}$$

and

$$\nabla F(x^*) = \begin{pmatrix} \nabla F(x^*)_{ff} & \nabla F(x^*)_{f\bar f} \\ \nabla F(x^*)_{\bar f f} & \nabla F(x^*)_{\bar f\bar f} \end{pmatrix},$$

where $I_{\bar f} := I \setminus I_f$, and taking into account that

$$\begin{aligned} (D_a)_{ii}(x^*) &= 0 \qquad \forall i \in I_f, \\ (D_b)_{ii}(x^*) &= -1 \qquad \forall i \in I_f \end{aligned}$$

by Proposition 2, we can rewrite (8) as

$$-\nabla F(x^*)_{ff}\Phi(x^*)_f + \nabla F(x^*)_{f\bar f}(D_b)_{\bar f\bar f}(x^*)\Phi(x^*)_{\bar f} = 0_f, \tag{9}$$

$$(D_a)_{\bar f\bar f}(x^*)\Phi(x^*)_{\bar f} - \nabla F(x^*)_{\bar f f}\Phi(x^*)_f + \nabla F(x^*)_{\bar f\bar f}(D_b)_{\bar f\bar f}(x^*)\Phi(x^*)_{\bar f} = 0_{\bar f}. \tag{10}$$

Due to the assumed nonsingularity of $\nabla F(x^*)_{ff}$, we obtain from (9):

$$\Phi(x^*)_f = \nabla F(x^*)_{ff}^{-1}\nabla F(x^*)_{f\bar f}(D_b)_{\bar f\bar f}(x^*)\Phi(x^*)_{\bar f}. \tag{11}$$

Substituting this expression into (10) and rearranging terms gives

$$\left[(D_a)_{\bar f\bar f}(x^*) + (\nabla F(x^*)/\nabla F(x^*)_{ff})(D_b)_{\bar f\bar f}(x^*)\right]\Phi(x^*)_{\bar f} = 0_{\bar f}. \tag{12}$$

Since

(i) the diagonal matrices $(D_a)_{\bar f\bar f}(x^*)$ and $(D_b)_{\bar f\bar f}(x^*)$ have nonpositive entries,
(ii) a diagonal element of $(D_a)_{\bar f\bar f}(x^*)$ or $(D_b)_{\bar f\bar f}(x^*)$ can be zero only if the corresponding component of $\Phi(x^*)_{\bar f}$ is zero, and
(iii) the diagonal matrices $(D_a)_{\bar f\bar f}(x^*)$ and $(D_b)_{\bar f\bar f}(x^*)$ are always postmultiplied by $\Phi(x^*)_{\bar f}$ in the system (9), (10),

we can assume without loss of generality that all diagonal entries of $D_a(x^*)$ and $D_b(x^*)$ are negative. But then Proposition 6 and the assumption of our theorem show that the coefficient matrix in (12) is nonsingular. Hence we get $\Phi(x^*)_{\bar{f}} = 0_{\bar{f}}$ from (12). But then (11) implies $\Phi(x^*)_f = 0_f$. Hence $\Phi(x^*) = 0$, i.e., x^* is a solution of the mixed complementarity problem by Proposition 1. □

Our next aim is to provide a sufficient condition for a stationary point of the constrained reformulation of the mixed complementarity problem

$$\min \Psi(x) \quad \text{s.t.} \quad x \in [l,u] \tag{13}$$

to be a global minimum. In fact, this result is of much more importance for this paper than the unconstrained stationary point result given in Theorem 2. However, Theorem 2 is of interest on its own and will also be used in order to establish our constrained stationary point result. Before doing this, we first prove another technical result.

Lemma 3. *Let $x \in \mathbb{R}^n$ be arbitrary and $H \in \partial\Phi(x), H = D_a(x) + D_b(x)F'(x)$ with diagonal matrices $D_a(x), D_b(x) \in \mathbb{R}^{n\times n}$ as defined in Proposition 2. Then the following two statements hold:*

(a) For all $i \in I$, $[D_a(x)\Phi(x)]_i[D_b(x)\Phi(x)]_i \geq 0$.
(b) For each $i \notin I_f$, $[D_a(x)\Phi(x)]_i = 0 \Longleftrightarrow [D_b(x)\Phi(x)]_i = 0 \Longleftrightarrow \Phi_i(x) = 0$.

Proof. (a) By considering the four possible cases $i \in I_l, i \in I_u, i \in I_{lu}$ and $i \in I_f$, it is easy to see that $(D_a)_{ii}(x) \leq 0$ and $(D_b)_{ii}(x) \leq 0$ for all $i \in I$. Hence

$$[D_a(x)\Phi(x)]_i[D_b(x)\Phi(x)]_i = (D_a)_{ii}(x)(D_b)_{ii}(x)\Phi_i(x)^2 \geq 0$$

for all $i \in I$.

(b) If $\Phi_i(x) = 0$, we immediately have

$$[D_a(x)\Phi(x)]_i = 0 \quad \text{and} \quad [D_b(x)\Phi(x)]_i = 0.$$

Conversely, assume that $[D_a(x)\Phi(x)]_i = 0$ for some index $i \notin I_f$ (the proof is analogous if $[D_b(x)\Phi(x)]_i = 0$). Then

$$(D_a)_{ii}(x) = 0 \quad \text{or} \quad \Phi_i(x) = 0.$$

In the latter case, there is nothing to show. So suppose that $(D_a)_{ii}(x) = 0$. Due to the definition of $D_a(x)$, we distinguish three cases.

Case 1: $i \in I_l$.
If $(x_i - l_i, F_i(x)) = (0,0)$, then $\Phi_i(x) = 0$ follows immediately from the definition of the operator Φ. Otherwise Proposition 2 (a) gives

$$0 = (D_a)_{ii}(x) = \frac{x_i - l_i}{\|(x_i - l_i, F_i(x))\|} - 1.$$

This implies $x_i - l_i > 0$ and $F_i(x) = 0$, so that $\Phi_i(x) = 0$ in view of the very definition of Φ and the NCP-property of the function ϕ.

Case 2: $i \in I_u$.
The proof of this case is very similar to the one given for Case 1 and we therefore omit the details.

Case 3: $i \in I_{lu}$.
If $(x_i - l_i, \phi(u_i - x_i, -F_i(x)) = (0,0)$, we are done. So assume that $(x_i - l_i, \phi(u_i - x_i, -F_i(x)) \neq (0,0)$. Then Proposition 2 (c) gives

$$0 = (D_a)_{ii}(x) = a_i(x) + b_i(x)c_i(x) \tag{14}$$

with certain numbers $a_i(x), b_i(x)$ and $c_i(x)$ specified in Proposition 2 (c). Since it follows immediately from this Proposition that

$$a_i(x) \leq 0, \ \ b_i(x) \leq 0 \quad \text{and} \quad c_i(x) \geq 0,$$

the right-hand side of (14) is the sum of two nonpositive expressions which can therefore be equal to zero only if $a_i(x) = 0$. This, however, implies $x_i - l_i > 0$ and $\phi(u_i - x_i, -F_i(x)) = 0$ in view of the definition of $a_i(x)$ given in Proposition 2 (c). Hence $\Phi_i(x) = 0$ by the NCP-property of the function ϕ. □

Note that Lemma 3 (b) does not hold for indices $i \in I_f$ since $(D_b)_{ii}(x) = -1$ for all $i \in I_f$ in view of Proposition 2 (d).

We are now able to prove the main result of this section.

Theorem 3. *Let $x^* \in \mathbb{R}^n$ be a stationary point of the constrained reformulation (13) of the mixed complementarity problem. Assume that*

(a) the principal submatrix $\nabla F(x^*)_{ff}$ is nonsingular, and
(b) the Schur complement $\nabla F(x^*)/\nabla F(x^*)_{ff}$ is a P_0-matrix.

Then x^* is a solution of the mixed complementarity problem.

Proof. Since x^* is a stationary point of the reformulation (13), it satisfies the following conditions (which themselves form a mixed complementarity problem):

$$\begin{aligned} x_i^* = l_i &\Longrightarrow [\nabla \Psi(x^*)]_i \geq 0, \\ x_i^* = u_i &\Longrightarrow [\nabla \Psi(x^*)]_i \leq 0, \\ x_i^* \in (l_i, u_i) &\Longrightarrow [\nabla \Psi(x^*)]_i = 0. \end{aligned} \tag{15}$$

The main part in proving that x^* is already a solution of the mixed complementarity problem consists in showing that we actually have $[\nabla \Psi(x^*)]_i = 0$ for all $i \in I$.

The proof is by contradiction, so assume that $\nabla \Psi(x^*) \neq 0$. Since $\nabla \Psi(x^*)$ can be written as

$$\nabla \Psi(x^*) = H^T \Phi(x^*) = D_a(x^*)\Phi(x^*) + \nabla F(x^*) D_b(x^*) \Phi(x^*)$$

for a matrix $H \in \partial\Phi(x^*)$ and certain diagonal matrices $D_a(x^*), D_b(x^*) \in \mathbb{R}^{n\times n}$ by Propositions 2 and 5, and since we necessarily have

$$[\nabla\Psi(x^*)]_f = 0_f$$

because of (15), we can follow the argument used in the proof of Theorem 2 in order to show that

$$\Phi(x^*)_f = \nabla F(x^*)_{ff}^{-1}\nabla F(x^*)_{f\bar{f}}(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}.$$

Substituting this into the expression for $[\nabla\Psi(x^*)]_{\bar{f}}$ and rearranging terms, we obtain

$$[\nabla\Psi(x^*)]_{\bar{f}} = (D_a)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}} + (\nabla F(x^*)/\nabla F(x^*)_{ff})(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}. \tag{16}$$

Since $\nabla\Psi(x^*) \neq 0$ by assumption, there exists an index $i \in I_{\bar{f}}$ such that either

$$x_i^* = l_i \quad \text{and} \quad [\nabla\Psi(x^*)]_i > 0 \tag{17}$$

or

$$x_i^* = u_i \quad \text{and} \quad [\nabla\Psi(x^*)]_i < 0. \tag{18}$$

Now it follows easily from Proposition 2 that if $x_i^* = l_i$, then

$$[(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_i \leq 0$$

and that if $x_i^* = u_i$,

$$[(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_i \geq 0.$$

Therefore, if we premultiply $[\nabla\Psi(x^*)]_i$ in (17) and (18) by $[(D_b)_{\bar{f}\bar{f}}\Phi(x^*)_{\bar{f}}]_i$ and substitute the ith component from the expression (16) for $[\nabla\Psi(x^*)]_i$, we obtain in both cases that

$$\begin{aligned}[(D_a)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_i[(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_i +{}& \\ [(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_i[(\nabla F(x^*)/\nabla F(x^*)_{ff})(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_i &\leq 0. \end{aligned} \tag{19}$$

Note that this inequality holds for all indices $i \in I_{\bar{f}}$ such that $[\nabla\Psi(x^*)]_i \neq 0$. In addition, we can show in a similar way that the equality

$$\begin{aligned}[(D_a)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_i[(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_i +{}& \\ [(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_i[(\nabla F(x^*)/\nabla F(x^*)_{ff})(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_i &= 0 \end{aligned} \tag{20}$$

holds for all indices $i \in I_{\bar{f}}$ with $[\nabla\Psi(x^*)]_i = 0$.

Since $\nabla\Psi(x^*) \neq 0$ by assumption and since we already know that $[\nabla\Psi(x^*)]_f = 0$, it follows immediately from (16) and Lemma 3 (b) that $(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}$ is a nonzero vector. Therefore, since the Schur complement $\nabla F(x^*)/\nabla F(x^*)_{ff}$ is a P_0-matrix by assumption, there exists an index $i_0 \in I_{\bar{f}}$ such that

$$[(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_{i_0} \neq 0 \quad \text{and}$$
$$[(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_{i_0}[(\nabla F(x^*)/\nabla F(x^*)_{ff})(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_{i_0} \geq 0. \quad (21)$$

Now Lemma 3 (a), (19), (20) and (21) imply that

$$0 = [(D_a)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_{i_0}[(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_{i_0}$$

and therefore $[(D_b)_{\bar{f}\bar{f}}(x^*)\Phi(x^*)_{\bar{f}}]_{i_0} = 0$ by Lemma 3 (b). This, however, contradicts the choice of the index i_0 in (21).

Hence we must have $\nabla\Psi(x^*) = 0$, so that Theorem 2 gives the desired result that x^* is a solution of the mixed complementarity problem. □

We note that, if we apply the main results of this section to the standard nonlinear complementarity problem, then we obtain some known properties of the merit function Ψ, see [9,14,18]. We also believe that appropriate modifications of the above theory will enable us to use other merit functions such as those described in [6,29,35].

4. Algorithmic Framework

In this section, we present our class of algorithms for the solution of the mixed complementarity problem and the corresponding global and local convergence theory. In our class of methods, we assume we have a basic algorithm, let us call it **Algorithm A**, with the following two properties:

(a) Given any point $x^k \in [l,u]$, if Algorithm A is able to compute a search direction $d^k \in \mathbb{R}^n$, then this direction satisfies $x^k + d^k \in [l,u]$;
(b) Given any sequence $\{x^k\}$ converging to a strongly regular solution x^* of the mixed complementarity problem, Algorithm A is able to compute a search direction d^k for all x^k sufficiently close to x^*, and this direction has the property that $\|x^k + d^k - x^*\| = o(\|x^k - x^*\|)$.

Property (a) is a very weak assumption; it does not even assume that Algorithm A is able to do anything at an arbitrary given point x^k (e.g., Algorithm A might not be well-defined due to inconsistent subproblems). However, if Algorithm A is able to compute a search direction, we assume that it computes a search direction such that, if we take the full step, then the new point $x^k + d^k$ stays in the feasible set $[l,u]$. Note that, due to the convexity of our feasible set $[l,u]$, this guarantees that all the points $x^k + t_k d^k, t_k \in [0,1]$ are feasible, too. On the other hand, property (b) states that, under Robinson's strong regularity condition, Algorithm A is locally well-defined and generates a locally superlinearly convergent search direction. Hence we view Algorithm A as a feasible and locally superlinearly convergent method for the solution of the mixed complementarity problem. Note, however, that the above two conditions say nothing about the way in which we generate the sequence $\{x^k\}$.

Several methods satisfy the above two conditions. For example, one may take the Josephy-Newton method as Algorithm A, see Josephy [20,26]. Alternatively, the method suggested by Ralph [30] has these two properties. Note that this method is the basis for the PATH solver by Dirkse and Ferris [10], to which we will come back in our numerical section. The NE/SQP method by Pang and Gabriel [28] is another possible candidate for Algorithm A as is the inexact QP-based solver by Kanzow [21] (these two methods have been used to solve the standard complementarity problem only, but it is not difficult to extend both methods to mixed complementarity problems, see, e.g., [2,4]).

Our class of algorithms globalizes Algorithm A as follows. We use the merit function Ψ to measure any progress. If the point generated by Algorithm A has a function value of Ψ sufficiently smaller than the previous one, it is accepted as the new iterate. Otherwise a projected gradient step for our smooth merit function Ψ is taken. In this way, we guarantee that all iterates stay in the feasible set $[l, u]$. In effect, our class of methods is an algorithmic framework for the solution of the box constrained optimization problem

$$\min \Psi(x) \quad \text{s.t.} \quad x \in [l, u]. \tag{22}$$

We now give a detailed statement of our class of methods, where the projection of an arbitrary point $z \in \mathbb{R}^n$ on the feasible set $[l, u]$ is denoted by $[z]_+$.

Algorithm 1 *(General Descent Framework)*

(S.0) (Initialization)
Choose $x^0 \in [l, u], s > 0, \beta \in (0,1), \gamma \in (0,1)$ *and set* $k := 0$.

(S.1) (Termination Criterion)
If x^k *is a stationary point of (22): STOP.*

(S.2) (Compute Fast Search Direction)
Use Algorithm A to compute a search direction d^k. *If this is not possible or if the condition*

$$\Psi(x^k + d^k) \le \gamma \Psi(x^k) \tag{23}$$

is not satisfied, go to Step (S.4), else go to Step (S.3).

(S.3) (Accept Fast Search Direction)
Set $x^{k+1} := x^k + d^k, k \leftarrow k + 1$, *and go to Step (S.1).*

(S.4) (Take Projected Gradient Step)
Compute $t_k = \max\{s\beta^\ell \mid \ell = 0, 1, 2, \dots\}$ *such that*

$$\Psi(x^k(t_k)) \le \Psi(x^k) - \sigma \nabla \Psi(x^k)^T (x^k - x^k(t_k)), \tag{24}$$

where $x^k(t) := [x^k - t\nabla\Psi(x^k)]_+$. *Set* $x^{k+1} := x^k(t_k), k \leftarrow k + 1$, *and go to Step (S.1).*

Other methods have attempted to use projected gradient steps in conjunction with steps that give fast local convergence. See for example, Ferris and Ralph [16]. Unfortunately, these hybrid algorithms are difficult to implement and numerical testing has therefore only been carried out on small test examples. A key difference in the approach outlined here is that the implementation can be achieved as a minor modification of an existing code.

We now investigate the convergence properties of our class of methods 1. To this end, we always assume implicitly that Algorithm 1 does not terminate in a finite number of steps, i.e., none of the iterates x^k is a stationary point of (22).

We begin with the global convergence analysis that consists of two parts. We first show that the algorithm is well-defined, and then we prove that any accumulation point of a sequence $\{x^k\}$ generated by Algorithm 1 is at least a stationary point of the bound constrained optimization problem

$$\min \Psi(x) \quad \text{s.t.} \quad x \in [l, u]. \tag{25}$$

Recall that Theorem 3 gives a relatively mild condition for a stationary point of (25) to be a solution of the mixed complementarity problem.

Theorem 4. *Algorithm 1 is well-defined for an arbitrary mixed complementarity problem with a continuously differentiable function F defined on an open set containing the rectangle $[l, u]$.*

Furthermore, every accumulation point of a sequence $\{x^k\}$ generated by Algorithm 1 is a stationary point of (25).

Proof. To prove the algorithm is well-defined, we only have to show that the projected gradient step can be carried out at each iteration, i.e., that we can always find a finite steplength $t_k > 0$ satisfying condition (24). However, since we assume that none of the iterates x^k is a stationary point of (22), this follows, e.g., from Proposition 2.3.3 (a) in Bertsekas [1].

For the second part of the theorem, let x^* be a stationary point of the sequence $\{x^k\}$, and assume that $\{x^k\}_K$ is a subsequence converging to x^*. Suppose there are infinitely many $k \in K$ such that x^{k+1} is generated by using a projected gradient step for all these k. Since the iterates x^k belong to the feasible set $[l, u]$ for all $k \in \mathbb{N}$ and since the sequence $\{\Psi(x^k)\}$ is monotonically decreasing, it is not difficult to see that the proof of Proposition 2.3.3 (b) in Bertsekas [1] can be adapted in a straightforward manner to establish that x^* is a stationary point of the constrained reformulation (25).

Hence we can assume without loss of generality that all iterates $k \in K$ satisfy the descent condition (23). Due to the monotone decrease of the sequence $\{\Psi(x^k)\}$, this implies that the entire sequence $\{\Psi(x^k)\}$ converges to 0. In particular, in view of the definition of our merit function, we see that the accumulation point x^* is a solution of the mixed complementarity problem and hence also a stationary point of problem (25). $\square$

We next want to show that Algorithm 1 is locally Q-superlinearly convergent under Robinson's [31] strong regularity condition.

The proof is in two parts. We first show that the entire sequence generated by Algorithm 1 converges to a solution x^* if this solution satisfies the strong regularity assumption. The critical tool to establish this result is the following proposition of Moré and Sorensen [27].

Proposition 7. *Assume that x^* is an isolated accumulation point of a sequence $\{x^k\}$ (not necessarily generated by Algorithm 1) such that $\{\|x^{k+1} - x^k\|\}_K \to 0$ for any subsequence $\{x^k\}_K$ converging to x^*. Then the whole sequence $\{x^k\}$ converges to x^*.*

Once convergence is established, we then determine the rate of convergence. The basic device used to show Q-superlinear convergence of our class of methods given in 1 is to prove that eventually there are no projected gradient steps, so our method inherits the local convergence properties of the locally superlinearly convergent Algorithm A used in Step (S.2) of Algorithm 1. To simplify the proof, we invoke the following proposition from [23] (see also [14] for a similar result).

Proposition 8. *Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitzian, $x^* \in \mathbb{R}^n$ with $G(x^*) = 0$ be such that all elements in $\partial G(x^*)$ are nonsingular, and assume that there are two sequences $\{x^k\} \subseteq \mathbb{R}^n$ and $\{d^k\} \subseteq \mathbb{R}^n$ (not necessarily generated by Algorithm 1) with $\{x^k\} \to x^*$ and $\|x^k + d^k - x^*\| = o(\|x^k - x^*\|)$. Then $\|G(x^k + d^k)\| = o(\|G(x^k)\|)$.*

We are now in the position to state our main local convergence result for Algorithm 1. The convergence rate established here depends critically on the main result of Section 2, namely Theorem 1.

Theorem 5. *Let $\{x^k\} \subseteq \mathbb{R}^n$ be a sequence generated by Algorithm 1. Assume that this sequence has an accumulation point x^* which is a strongly regular solution of the mixed complementarity problem. Then the entire sequence $\{x^k\}$ converges to this point, and the rate of convergence is Q-superlinear.*

Proof. To establish convergence, we first note that a strongly regular solution is an isolated solution of the mixed complementarity problem, see [31]. Since Algorithm 1 generates a decreasing sequence $\{\Psi(x^k)\}$ and x^* is a solution of the mixed complementarity problem, the entire sequence $\{\Psi(x^k)\}$ converges to zero. Hence every accumulation point of the sequence $\{x^k\}$ must be a solution of the mixed complementarity problem. Therefore, the assumed strong regularity of x^* implies that x^* is an isolated accumulation point of the sequence $\{x^k\}$.

Now let $\{x^k\}_K$ denote any subsequence converging to x^*. Assume first that we take a projected gradient step for all $k \in K$. Then we obtain, using the nonexpansive property of the projection operator:

$$
\begin{aligned}
\|x^{k+1} - x^k\| &= \|x^k(t_k) - x^k\| \\
&= \|[x^k - t_k \nabla\Psi(x^k)]_+ - x^k\| \\
&= \|[x^k - t_k \nabla\Psi(x^k)]_+ - [x^k]_+\| \\
&\le \|t_k \nabla\Psi(x^k)\| \\
&\le s\|\nabla\Psi(x^k)\| \\
&\to 0
\end{aligned}
\tag{26}
$$

since x^* solves the mixed complementarity problem so that x^* is a global minimizer and hence a stationary point of our merit function Ψ.

On the other hand, if we calculate the search direction d^k by using Algorithm A for infinitely many $k \in K$, our assumptions on Algorithm A and the assumed strong regularity imply that $\{d^k\} \to 0$ on this infinite subsequence, so that the updating rules from Algorithm 1 show that

$$\|x^{k+1} - x^k\| = \|d^k\| \to 0 \tag{27}$$

on this subsequence. Combining (26) and (27), we immediately obtain that

$$\{\|x^{k+1} - x^k\|\}_K \to 0.$$

Hence the assumptions of Proposition 7 are satisfied, and convergence follows from that result.

In order to establish the Q-superlinear rate, we note that the strong regularity of the solution x^* and Theorem 1 show that all elements in $\partial\Phi(x^*)$ are nonsingular. Since, in view of our assumptions about the search directions d^k generated by Algorithm A, we have $\|x^k + d^k - x^*\| = o(\|x^k - x^*\|)$ for these search directions, Proposition 8 implies that

$$\|\Phi(x^k + d^k)\| = o(\|\Phi(x^k)\|)$$

and therefore

$$\Psi(x^k + d^k) = o(\Psi(x^k)).$$

This shows that the descent condition

$$\Psi(x^k + d^k) = \gamma\Psi(x^k)$$

is eventually satisfied in Step (S.2) of Algorithm 1, i.e., for all $k \in \mathbb{N}$ sufficiently large, Algorithm 1 does not take any projected gradient steps. Hence Algorithm 1 has the same local convergence properties as Algorithm A. Since $x^{k+1} = x^k + d^k$, this means that $\{x^k\}$ converges Q-superlinearly to x^*. $\square$

Obviously, if the basic Algorithm A is locally Q-quadratically convergent, the class of Algorithms given in 1 is also locally Q-quadratically convergent. Typically, this holds if we assume in addition that the Jacobian of F is locally Lipschitzian.

5. Computational Results

The computational results in this section were carried out by extending the PATH 3.0 solver to allow use of a different merit function and projected gradient steps. The coding was done in ANSI-C and enabled these extensions based on setting the option `merit_function fischer`.

The PATH solver is described in detail in [10,15]. The code is extensively used by economists for solving general equilibrium problems and is well known to be robust and efficient on the majority of the mixed complementarity problems it encounters. The algorithm successively linearizes the normal map [32] associated with the MCP, defined by

$$F([y]_+) + y - [y]_+$$

where $[y]_+$ represents the projection of y onto $[l, u]$ in the Euclidean norm, thereby generating a sequence of linear mixed complementarity problems. These subproblems are solved by generating a path between the current iterate and the solution of the linear subproblem; the precise details of this generation scheme can be found for example in [10]. A nonmonotone backtracking search is performed on this path to garner sufficient decrease in its merit function, the norm of the residual of the normal map. It is known that the solutions of the subproblem will eventually provide descent for the merit function and that local superlinear or quadratic convergence will occur under appropriate conditions. A crash procedure [12] is used to quickly identify an approximation to the active set at the solution; this is based on a projected Newton step for the normal map, but the direction produced is not known to be a descent direction for the merit function used.

The PATH 3.0 code was modified to incorporate the extensions that are outlined in this paper. Two implementational points are of interest. Firstly, a backtracking search was implemented instead of the simple test given as (23). This search inspected points that form the following arc, parametrized by $t \in (0, 1]$:

$$[x^k + t(x^N - x^k)]_+$$

where x^N is the solution of the linearized normal map. Note that, in general, this is not the line segment joining x^k to x^N. A projected gradient step was only taken if suitable descent did not occur for some minimum steplength allowed. This line search was chosen to be consistent with that used in the projected gradient step given as (24). Secondly, the gradient of the merit function required for (24) was calculated using the formulas detailed in Propositions 5 and 2 and the method described in [2].

In the following two tables, we give the number of successes and failures of our new code, PATH 4.0, and the PATH 3.0 code from all starting points in the MCPLIB collection of test problems [11]. Two tables are presented by splitting the problems into standard MCP models (Table 1) and models that were generated using the MPSGE preprocessor [33,34] in GAMS (Table 2). In order to condense the information in Table 1, we have grouped several similar models together whenever this grouping results in no loss of information; for example, problems `colvdual` and `colvnlp` are grouped together as example `colv*` in Table 1.

It was noted in [15] that restarting PATH 3.0 from the user provided starting point after the algorithm failed to find a solution on the first attempt significantly improved the robustness of the code. In the results reported in Table 1, we use exactly the same restart parameter settings in both codes (see [15]), with the exception that the PATH 4.0 code is able to carry out at most 5 projected

Problem	Without Restarts		With Restarts	
	PATH 3.0	PATH 4.0	PATH 3.0	PATH 4.0
asean9a,hanson	3(0)	3(0)	3(0)	3(0)
badfree,degen,qp	3(0)	3(0)	3(0)	3(0)
bert_oc	4(0)	4(0)	4(0)	4(0)
bertsekas,gafni	9(0)	9(0)	9(0)	9(0)
billups	0(3)	0(3)	0(3)	0(3)
bishop	1(0)	1(0)	1(0)	1(0)
bratu,obstacle	9(0)	9(0)	9(0)	9(0)
choi,nash	5(0)	5(0)	5(0)	5(0)
colv*	10(0)	10(0)	10(0)	10(0)
cycle,explcp	2(0)	2(0)	2(0)	2(0)
duopoly	0(1)	0(1)	1(0)	1(0)
ehl_k*	6(6)	8(4)	11(1)	12(0)
electric	0(1)	0(1)	1(0)	1(0)
force*	2(0)	2(0)	2(0)	2(0)
freebert	7(0)	7(0)	7(0)	7(0)
games	16(9)	19(6)	23(2)	25(0)
hanskoop	10(0)	10(0)	10(0)	10(0)
hydroc*,methan08	3(0)	3(0)	3(0)	3(0)
jel,jmu	3(0)	3(0)	3(0)	3(0)
josephy,kojshin	15(1)	16(0)	16(0)	16(0)
lincont	1(0)	1(0)	1(0)	1(0)
mathi*	11(2)	13(0)	13(0)	13(0)
ne-hard	1(0)	1(0)	1(0)	1(0)
opt_cont*	5(0)	5(0)	5(0)	5(0)
pgvon*	6(4)	5(5)	6(4)	6(4)
pies	1(0)	1(0)	1(0)	1(0)
powell*	12(0)	12(0)	12(0)	12(0)
scarf*	12(0)	11(1)	12(0)	12(0)
shubik	38(10)	39(9)	48(0)	48(0)
simple-*	1(1)	1(1)	1(1)	1(1)
sppe,tobin	7(0)	7(0)	7(0)	7(0)
tinloi	56(8)	61(3)	63(1)	64(0)
trafelas	2(0)	2(0)	2(0)	2(0)
Total	261(46)	273(34)	295(12)	299(8)

Table 1. Results for regular models

gradient steps when the direction provided by the linear subproblem is poor. This factor can be changed by setting the `gradient_step_limit` option to a suitable constant. Since the merit function in PATH 3.0 is nonsmooth, there is no possibility of carrying out a similar scheme in this code. We note that several heuristic procedures were described and tested in [12]; all of these appear not to be beneficial to PATH.

MPSGE models are generated differently by GAMS [5], and thus a solver can distinguish them from general MCP models. We have used this information to choose different default options for MPSGE models both in PATH 3.0 and PATH 4.0. For PATH 4.0, by default there is no crash and we use a linesearch. The first restart for the MPSGE models is to crash using the projected Newton technique and use a linesearch. The remaining restart procedures (that are not used for any problem on our test set) are identical to those documented in [15].

As can be seen from the results of Tables 1 and 2, PATH 4.0 is considerably better than PATH 3.0 when the codes are not allowed to restart. The projected

Problem	Without Restarts		With Restarts	
	PATH 3.0	PATH 4.0	PATH 3.0	PATH 4.0
box	340(21)	359(2)	355(6)	361(0)
denmark	31(9)	40(0)	40(0)	40(0)
eppa	8(0)	8(0)	8(0)	8(0)
eta2100	1(0)	1(0)	1(0)	1(0)
mrtmge	0(1)	1(0)	0(1)	1(0)
multi-v	3(0)	3(0)	3(0)	3(0)
olg	0(1)	1(0)	1(0)	1(0)
romer	0(1)	1(0)	0(1)	1(0)
trade12	2(0)	2(0)	2(0)	2(0)
uruguay	7(0)	7(0)	7(0)	7(0)
Total	392(33)	423(2)	417(8)	425(0)

Table 2. Results for MPSGE models

gradient steps allow progress to be made in this code, which cannot occur in PATH 3.0. We note that the total number of projected gradient steps over the total of 732 runs was 349, indicating our preference to take such steps only as a last resort (i.e. only after nonmonotone search and the watchdog strategy fail). When restarts are allowed, it still seems that PATH 4.0 is more robust; however the margin is significantly smaller, in part due to the fact that most of the problems are already solved by PATH 3.0. On the MPSGE models, however, restarts allow us to solve all the models in the test collection.

To conclude this section, Table 3 provides more complete information for the algorithm PATH 4.0 on a subset of the test problems. These problems were selected by taking every 10th run in an alphabetical ordering of all the test problems considered. However, we only report the first three instances of any given problem. We believe this is an unbiased sample of our test results. A complete listing of the results is available from

$$\texttt{http://www.cs.wisc.edu/math-prog/tech-reports}$$

The columns of Table 3 indicate the number of starting points (SP), number of major iterations (MI), crash iterations (CI), restarts (R) and projected gradient steps (PG) taken. The final column of this table reports the time for PATH 4.0 in seconds, with the time for PATH 3.0 added in parentheses. All runs were carried out on a Sun UltraSparc 300 MHz processor with 256MB RAM. A "*" indicates failure of the method.

It is hard to draw firm conclusions from Table 3. It indicates that the solution times of both algorithms are comparable, with some smaller times for PATH 4.0 and some for PATH 3.0. There are only 4 problems in this subset which use projected gradient steps and restarts, the vast majority of the problems being solved without invoking these strategies.

Overall, the theoretical extensions outlined in this paper result in an improvement in robustness of PATH 3.0 without any noticeable change in the accuracy or speed of the code. Further testing on even more challenging problems is required to fully determine the effects of different merit functions within the PATH code. This will be the subject of future research.

Problem	SP	MI	CI	R	PG	T
bert_oc	3	0	3	0	0	0.6 (0.4)
billups	3	38	2	3	4	* (*)
box	9	3	0	0	0	0.0 (0.0)
box	19	4	0	0	0	0.0 (0.0)
box	29	3	0	0	0	0.0 (0.0)
colvnlp	2	3	1	0	0	0.0 (0.0)
denmark	4	11	0	0	0	9.9 (25.2)
denmark	14	3	0	0	0	2.0 (25.1)
denmark	24	11	0	0	0	9.2 (10.3)
ehl_k40	3	80	1	1	6	3.4 (*)
electric	1	128	1	1	6	2.0 (1.3)
explcp	1	1	1	0	0	0.0 (0.0)
gafni	1	3	1	0	0	0.0 (0.0)
games	8	5	1	0	0	0.1 (0.0)
games	18	29	1	0	0	0.2 (0.3)
hanskoop	3	6	1	0	0	0.0 (0.0)
hydroc06	1	4	1	0	0	0.0 (0.0)
josephy	6	5	1	0	0	0.0 (0.0)
kojshin	8	1	1	0	0	0.0 (0.0)
mathisum	3	7	1	0	0	0.0 (0.0)
nash	1	5	1	0	0	0.0 (0.0)
obstacle	6	0	7	0	0	1.3 (1.3)
pgvon105	2	13	1	0	0	0.2 (0.1)
powell	1	6	1	0	0	0.0 (0.0)
powell_mcp	5	7	1	0	0	0.0 (0.0)
scarfasum	2	3	1	0	0	0.0 (0.0)
shubik	4	3	0	0	0	0.0 (0.0)
shubik	14	13	1	0	0	0.0 (0.1)
shubik	24	41	1	1	5	0.2 (0.1)
tinloi	1	1	1	0	0	0.0 (0.0)
tinloi	11	1	2	0	0	0.1 (0.0)
tinloi	21	1	2	0	0	0.0 (0.0)
trafelas	1	7	22	0	0	6.2 (6.3)

Table 3. Selected full results

Acknowledgements. The authors would like to thank Roman Snajder for pointing out the relation of some of our results with reference [19].

References

1. D.P. Bertsekas: *Nonlinear Programming.* Athena Scientific, Belmont, MA, 1995.
2. S.C. Billups: *Algorithms for complementarity problems and generalized equations.* Ph.D. Thesis, Computer Sciences Department, University of Wisconsin, Madison, WI, August 1995.
3. S.C. Billups: *A homotopy based algorithm for mixed complementarity problems.* Technical Report, Department of Mathematics, University of Colorado, Denver, CO, February 1998.
4. S.C. Billups and M.C. Ferris: *QPCOMP: A quadratic program based solver for mixed complementarity problems.* Mathematical Programming 76, 1997, pp. 533–562.
5. A. Brooke, D. Kendrick, and A. Meeraus: *GAMS: A User's Guide.* The Scientific Press, South San Francisco, CA, 1988.
6. B. Chen, X. Chen, and C. Kanzow: *A penalized Fischer-Burmeister NCP-function: Theoretical investigation and numerical results.* Preprint 126, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, September 1997 (revised May 1998).

7. B. Chen and P.T. Harker: *A noninterior continuation method for quadratic and linear programming.* SIAM Journal on Optimization 3, 1993, pp. 503–515.
8. F.H. Clarke: *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York, NY, 1983 (reprinted by SIAM, Philadelphia, PA, 1990).
9. T. De Luca, F. Facchinei and C. Kanzow: *A semismooth equation approach to the solution of nonlinear complementarity problems.* Mathematical Programming 75, 1996, pp. 407–439.
10. S.P. Dirkse and M.C. Ferris: *The PATH solver: A non-monotone stabilization scheme for mixed complementarity problems.* Optimization Methods and Software 5, 1995, pp. 123–156.
11. S.P. Dirkse and M.C. Ferris: *MCPLIB: A collection of nonlinear mixed complementarity problems.* Optimization Methods and Software 5, 1995, pp. 319–345.
12. S.P. Dirkse and M.C. Ferris: *Crash techniques for large-scale complementarity problems.* In: M.C. Ferris and J.-S. Pang (eds.): *Complementarity and Variational Problems: State of the Art.* SIAM, Philadelphia, PA, 1997, pp. 40–61.
13. F. Facchinei, A. Fischer, and C. Kanzow: *A semismooth Newton method for variational inequalities: The case of box constraints.* In: M.C. Ferris and J.-S. Pang (eds.): *Complementarity and Variational Problems: State of the Art.* SIAM, Philadelphia, 1997, pp. 76–90.
14. F. Facchinei and J. Soares: *A new merit function for nonlinear complementarity problems and a related algorithm.* SIAM Journal on Optimization 7, 1997, pp. 225–247.
15. M.C. Ferris and T.S. Munson: *Interfaces to PATH 3.0: Design, implementation and usage.* Computational Optimization and Applications, to appear.
16. M.C. Ferris and D. Ralph: *Projected gradient methods for nonlinear complementarity problems via normal maps.* In: D. Du, L. Qi and R. Womersley (eds.): *Recent Advances in Nonsmooth Optimization.* World Scientific Publishers, 1995, pp. 57–87.
17. A. Fischer: *A special Newton-type optimization method.* Optimization 24, 1992, pp. 269–284.
18. A. Fischer: *A new constrained optimization reformulation for complementarity problems.* Journal of Optimization Theory and Applications 97, 1998, pp. 105–117.
19. M.S. Gowda and R. Snajder: *Generalizations of P_0- and P-properties; extended vertical and horizontal LCPs.* Linear Algebra and its Applications 223/224, 1995, pp. 695–715.
20. N.H. Josephy: *Newton's method for generalized equations.* Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, WI 1979.
21. C. Kanzow: *An inexact QP-based method for nonlinear complementarity problems.* Numerische Mathematik, to appear.
22. C. Kanzow and H. Kleinmichel: *A new class of semismooth Newton-type methods for nonlinear complementarity problems.* Computational Optimization and Applications, to appear.
23. C. Kanzow and H.-D. Qi: *A QP-free constrained Newton-type method for variational inequality problems.* Mathematical Programming, to appear.
24. J. Liu: *Strong stability in variational inequalities.* SIAM Journal on Control and Optimization 33, 1995, pp. 725–749.
25. Z.–Q. Luo and P. Tseng: *Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem.* SIAM Journal on Optimization 2, 1992, pp. 43–54.
26. L. Mathiesen: *An algorithm based on a sequence of linear complementarity problems applied to a Walrasian equilibrium model: an example.* Mathematical Programming 37, 1987, pp. 1–18.
27. J.J. Moré and D.C. Sorensen: *Computing a trust region step.* SIAM Journal on Scientific and Statistical Computing 4, 1983, pp. 553–572.
28. J.-S. Pang and S.A. Gabriel: *NE/SQP: A robust algorithm for the nonlinear complementarity problem.* Mathematical Programming 60, 1993, pp. 295–337.
29. L. Qi: *Regular pseudo-smooth NCP and BVIP functions and globally and quadratically convergent generalized Newton methods for complementarity and variational inequality problems.* Applied Mathematics Report AMR 97/14, School of Mathematics, University of New South Wales, Sydney, Australia, July 1997 (revised September 1997).
30. D. Ralph: *Global convergence of Newton's method for nonsmooth equations via the path search.* Mathematics of Operations Research 19, 1994, pp. 352–389.
31. S.M. Robinson: *Strongly regular generalized equations.* Mathematics of Operations Research 5, 1980, pp. 43–62.

32. S.M. Robinson: *Normal maps induced by linear transformations.* Mathematics of Operations Research 17, 1992, pp. 691–714.
33. T.F. Rutherford: *Extensions of GAMS for complementarity problems arising in applied economic analysis.* Journal of Economic Dynamics and Control 19, 1995, pp. 1299–1324.
34. T.F. Rutherford: *Applied general equilibrium modeling with MPSGE as a GAMS subsystem: An overview of the modeling framework and syntax.* Computational Economics, to appear.
35. D. Sun and R.S. Womersley: *A new unconstrained differentiable merit function for box constrained variational inequality problems and a damped Gauss-Newton method.* SIAM Journal on Optimization, to appear.