

A TRULY GLOBALLY CONVERGENT NEWTON-TYPE METHOD FOR THE MONOTONE NONLINEAR COMPLEMENTARITY PROBLEM*

M. V. SOLODOV[†] AND B. F. SVAITER[†]

Abstract. The Josephy–Newton method for solving a nonlinear complementarity problem consists of solving, possibly inexactly, a sequence of linear complementarity problems. Under appropriate regularity assumptions, this method is known to be locally (superlinearly) convergent. To enlarge the domain of convergence of the Newton method, some globalization strategy based on a chosen merit function is typically used. However, to ensure global convergence to a solution, some additional restrictive assumptions are needed. These assumptions imply boundedness of level sets of the merit function and often even (global) uniqueness of the solution. We present a new globalization strategy for monotone problems which is not based on any merit function. Our linesearch procedure utilizes the regularized Newton direction and the monotonicity structure of the problem to force global convergence by means of a (computationally explicit) projection step which reduces the distance to the solution set of the problem. The resulting algorithm is truly globally convergent in the sense that the subproblems are always solvable, and the whole sequence of iterates converges to a solution of the problem without any regularity assumptions. In fact, the solution set can even be unbounded. Each iteration of the new method has the same order of computational cost as an iteration of the damped Newton method. Under natural assumptions, the local superlinear rate of convergence is also achieved.

Key words. nonlinear complementarity problem, Newton method, proximal point method, projection method, global convergence, superlinear convergence

AMS subject classifications. 90C30, 90C33

PII. S1052623498337546

1. Introduction. The classical nonlinear complementarity problem [28, 7], $\text{NCP}(F)$, is to find a point $x \in \mathfrak{R}^n$ such that

$$(1.1) \quad x \geq 0, \quad F(x) \geq 0, \quad \langle x, F(x) \rangle = 0,$$

where $F : \mathfrak{R}^n \rightarrow \mathfrak{R}^n$ and $\langle \cdot, \cdot \rangle$ denotes the usual inner product in \mathfrak{R}^n . Throughout this paper, we shall assume that $F(\cdot)$ is continuous and monotone, i.e.,

$$\langle F(x) - F(y), x - y \rangle \geq 0 \quad \text{for all } x, y \in \mathfrak{R}^n.$$

Note that under this assumption, the solution set of (1.1) is convex.

While there exists a wide range of approaches to solving $\text{NCP}(F)$ (see [14, 28, 7]), some of the most successful and widely used are Newton-type algorithms based on solving successive linearizations of the problem (see, for example, the more detailed discussion in [8]). Given a point x^k , the (Josephy–) Newton method [17, 30, 26, 14, 1] generates the next iterate x^{k+1} by solving the linear complementarity problem

$$(1.2) \quad x \geq 0, \quad F_k(x) \geq 0, \quad \langle x, F_k(x) \rangle = 0,$$

*Received by the editors April 16, 1998; accepted for publication (in revised form) June 1, 1999; published electronically February 24, 2000.

<http://www.siam.org/journals/siopt/10-2/33754.html>

[†]Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil (solodov@impa.br, benar@impa.br). The research of the first author was supported by CNPq grant 300734/95-6 and by PRONEX–Optimization. The research of the second author was supported by CNPq grant 301200/93-9(RN) and by PRONEX–Optimization.

where $F_k(\cdot)$ is the first-order approximation of $F(\cdot)$ at x^k ,

$$(1.3) \quad F_k(x) := F(x^k) + \nabla F(x^k)(x - x^k),$$

assuming $F(\cdot)$ is differentiable. If the starting point is sufficiently close to some *regular* [35] solution \bar{x} of (1.1), the sequence generated by the Newton method is well defined and converges to \bar{x} superlinearly or quadratically, depending on further assumptions. We note, in passing, that monotonicity of F is not needed for such local analysis.

There are two key difficulties with using the Newton method given by (1.2), (1.3) for solving $\text{NCP}(F)$. First, in the absence of regularity even local convergence cannot be ensured. Second, even if regularity holds at a solution, there are serious problems with ensuring global convergence. In particular, far from a regular solution of the problem, there is no guarantee that the linearization subproblems are solvable. And even if the subproblem solution exists, there is no guarantee that it actually constitutes some progress toward solving the original problem, $\text{NCP}(F)$. In this paper, we shall successfully address each of these difficulties in the context of monotone problems.

To enlarge the domain of convergence of the Newton method, some globalization strategy has to be used. However, as pointed out in [14, p. 182], “the global convergence remains a rare property for most of the modified methods.” As mentioned in [14, p. 185], “the trouble with general variational inequalities and nonlinear complementarity problems is that valid merit functions which are relatively easy to compute are very difficult if not impossible to find.” Although considerable progress has been made in the theory and numerical use of merit functions in recent years [11, 23, 10, 19, 6, 44, 22, 24, 45, 46] (see [12] for a survey of merit functions for variational inequality and complementarity problems), most of the known merit functions do not appear to be useful for the specific task of globalizing the Newton method given by (1.2), (1.3). At this point, it is worth emphasizing that the Newton method under consideration should not be confused with other Newton-like methods which are not of the Josephy type—for example, Newton methods for minimizing a particular merit function or Newton methods for equation-based NCP reformulations [18, 34, 21, 4, 6, 15, 49, 53, 16, 20, 13, 48].

Perhaps the most natural globalization strategy for (1.2), (1.3) is a linesearch procedure in the obtained Newton direction (if it exists!) aimed at decreasing the value of some valid merit function. In this regard, we note the following. First, many of the known NCP merit functions are not differentiable, e.g., those based on the natural residual [26] and the normal map [36]. This makes linesearch difficult, although an alternative pathsearch approach is possible [34, 3]. Second, for some differentiable merit functions, e.g., (the square of) the Fischer–Burmeister function [9, 19, 6], it appears that the Newton direction need not be a direction of descent, particularly far from a solution or in the absence of regularity. Thus the use of this function results in a rather indirect globalization of the Newton method [8]. (More on this later.)

To our knowledge, the only merit functions which have been used to globalize the Newton method of the form (1.2), (1.3) are the gap function [25], the regularized gap function [51, 50], the D-gap function [31, 32], and (the square of) the Fischer–Burmeister function [8]. (We note that some of these methods were developed for the more general mixed complementarity or variational inequality setting.) However, each of these globalizations has certain drawbacks. Using the gap function [25] requires exact minimization along the line to compute the stepsize, which is not implementable. Moreover, the gap function itself is not easy to compute in general.

In addition, compactness of the feasible set of the variational inequality problem is required, which precludes an application to complementarity problems. Using the regularized gap function [51, 50] admits inexact Armijo-type linesearch but requires strong monotonicity of F for global convergence (see also [52]). In addition, methods of [25, 51, 50] also need the (restrictive) strict complementarity assumption to establish superlinear/quadratic rate of convergence. Note that the subproblems in [25, 51, 50] are solvable due to the compactness of the feasible set and the strong monotonicity of F , respectively.

The methods based on the D-gap function [31, 32] and the Fischer–Burmeister function [8] globalize the Newton method in a rather indirect way. In [31, 32, 8] there is no guarantee that the subproblems are solvable. If the Newton direction does not exist (which cannot be checked a priori, so some wasteful computing will inevitably be done), the method resorts to an “escape” mechanism of taking a gradient descent step for the merit function. Even if the Newton direction exists, it still may happen that it does not satisfy conditions needed to obtain sufficient descent of the merit function by means of a linesearch procedure. In this case, the Newton point will be discarded altogether, and again a gradient descent step will be taken. In a sense, this is a rather indirect globalization of the Newton method, because such a strategy does not correspond to the damped Newton methodology. As for the convergence results in [31, 32, 8], every accumulation point of the generated sequence of iterates is a stationary point of the employed merit function (even without the monotonicity assumption on F). However, the existence of such accumulation points, and the equivalence of stationary points of the merit functions to solutions of $\text{NCP}(F)$, cannot be guaranteed without further assumptions. For example, in [31, 32], F is assumed to be a uniform P -function, which implies that $\text{NCP}(F)$ has a (globally) unique solution.

Finally, we briefly comment on the interesting regularization approach proposed in [5], which converges globally when the solution set of the NCP is compact and F is a P_0 -function. First, it is important to note that the subproblems in the method of [5] are *nonlinear* complementarity problems, which are structurally just as difficult to solve as the original problem itself (although they are better behaved due to regularization). Therefore, this method is not of the Newton type. Second, the global convergence result of [5] states that the sequence of iterates remains bounded, and its every accumulation point solves the NCP. This is weaker than the full convergence of the whole sequence which we shall establish for our method. On the other hand, the P_0 assumption on F is, of course, weaker than our assumption of monotonicity. It should also be mentioned that it is often unknown whether the solution set is compact. When F is monotone, this holds if the NCP is strictly feasible, i.e., there exists an x such that $x > 0$ and $F(x) > 0$. However, this condition is not easy to verify in general. In any case, the key conceptual difference between our method and that of [5] is in the linear versus nonlinear structure of subproblems solved at each iteration. It is worth noting that the latter is also the important difference between our method and the proximal point algorithm [37], which does converge globally under the monotonicity assumption only.

We emphasize that for each of the cited Josephy–Newton algorithms, to ensure global convergence of the whole sequence of iterates to a solution of the problem, one needs assumptions which, among other things, imply that the solution is unique. In fact, relatively restrictive assumptions are required even to prove boundedness of the iterates and convergence to zero of the distance to the solution set.

In this paper, we present a new globalization strategy for monotone problems which overcomes the above mentioned drawbacks. In particular, our algorithm is

truly globally convergent in the sense that from any starting point the whole sequence of iterates converges to a solution of $\text{NCP}(F)$ under no assumptions other than monotonicity and continuity of F (and of course, the existence of a solution); see Theorem 3.2. In particular, no regularity-type assumptions are needed. In fact, the solution set may even be unbounded. In addition, the linear complementarity subproblems are always solvable, and our algorithm allows for their inexact solution similar to the setting of [26]. This feature of approximate subproblem solutions is of particular importance for large-scale problems. (We note that the effect of an inexact subproblem solution has not been analyzed in globalizations proposed in [51, 50, 31, 32, 8].) Under the assumptions of positive definiteness of the Jacobian of F at the solution, and its Hölder continuity in some neighborhood of it, the local superlinear rate of convergence is also established (see Theorem 4.3). Each iteration of our algorithm (see Algorithm 2.2) consists of an approximate solution of a (strongly monotone) linear complementarity problem, followed by a linesearch procedure and a (computationally trivial) projection step. Thus computational cost of each iteration is of the same order as that of the damped Newton method.

2. The algorithm. We start with some equivalent formulations of $\text{NCP}(F)$, each of which will be useful in the subsequent analysis. In particular, the following five statements are equivalent:

1. \bar{x} solves $\text{NCP}(F)$.
2. \bar{x} is a solution of the variational inequality problem over the nonnegative orthant \mathfrak{R}_+^n :

$$\bar{x} \in \mathfrak{R}_+^n, \quad \langle F(\bar{x}), x - \bar{x} \rangle \geq 0 \quad \text{for all } x \in \mathfrak{R}_+^n.$$

3. \bar{x} yields a zero of the (maximal monotone) operator $F + N$:

$$0 \in F(\bar{x}) + N(\bar{x}),$$

where $N(x)$ is the normal cone to \mathfrak{R}_+^n at the point x .

4. \bar{x} is a zero of the natural (projection) residual:

$$0 = r(\bar{x}) := \min\{\bar{x}; F(\bar{x})\} = \bar{x} - [\bar{x} - F(\bar{x})]^+,$$

where the minimum is taken componentwise and $[\cdot]^+$ stands for the orthogonal projection map onto \mathfrak{R}_+^n .

The approach presented in this paper is in some ways motivated by the hybrid projection–proximal point method of [43] and the projection method of [47], which already proved to be useful for developing globally (and locally superlinearly) convergent Newton methods for systems of monotone equations [40]. In fact, in a sufficiently small neighborhood of a regular solution of $\text{NCP}(F)$ our Newton-type Algorithm 2.2 takes steps which can be viewed, in a certain sense, as iterations of the hybrid projection–proximal point method. This will prove to be the key to the local superlinear rate of convergence. We therefore first state the algorithm of [43] in the more general context of finding zeros of set-valued maximal monotone operators in a Hilbert space. Let T be a maximal monotone operator on a real Hilbert space \mathcal{H} . And consider, for a moment, the problem of finding an $x \in \mathcal{H}$ such that $0 \in T(x)$. Note that $\text{NCP}(F)$ considered here is a particular instance of this problem with $T(x) = (F + N)(x)$ and $\mathcal{H} = \mathfrak{R}^n$.

ALGORITHM 2.1 (hybrid projection–proximal point method [43]). Choose any $x^0 \in \mathcal{H}$ and $\sigma \in [0, 1)$; set $k := 0$.

Inexact Proximal Step. Choose $\mu_k > 0$ and find $y^k \in \mathcal{H}$ and $v^k \in T(y^k)$ such that

$$(2.1) \quad 0 = v^k + \mu_k(y^k - x^k) + \varepsilon^k,$$

where

$$(2.2) \quad \|\varepsilon^k\| \leq \sigma \max\{\|v^k\|, \mu_k\|y^k - x^k\|\}.$$

Stop if $v^k = 0$ or $y^k = x^k$. Otherwise,

Projection Step. Compute

$$x^{k+1} = x^k - \frac{\langle v^k, x^k - y^k \rangle}{\|v^k\|^2} v^k.$$

Set $k := k + 1$; and repeat.

If problem $0 \in T(x)$ has a solution and the sequence $\{\mu_k\}$ is bounded above, then the generated sequence $\{x^k\}$ either is finite and terminates at a solution or is infinite and converges (weakly) to a solution. The linear/superlinear rate of convergence is also achieved under standard assumptions. For complete properties of the method, see [43]. The idea of Algorithm 2.1 is to use an approximate proximal iteration to construct a hyperplane

$$(2.3) \quad H_k := \{x \mid \langle v^k, x - y^k \rangle = 0\},$$

which separates the current iterate x^k from the solutions of $0 \in T(x)$. Let us make this more precise. If $0 \in T(\bar{x})$, then, by monotonicity of T ,

$$\langle v^k, \bar{x} - y^k \rangle \leq 0$$

for any y^k and any $v^k \in T(y^k)$. Here we will consider a condition somewhat stronger than (2.2), namely,

$$(2.4) \quad \|\varepsilon^k\| \leq \sigma \mu_k \|y^k - x^k\|,$$

because this will be the only condition used in our Algorithm 2.2. Using (2.1) and the Cauchy–Schwarz inequality, we obtain

$$(2.5) \quad \begin{aligned} \langle v^k, x^k - y^k \rangle &= \mu_k \|x^k - y^k\|^2 - \langle \varepsilon^k, x^k - y^k \rangle \\ &\geq \mu_k \|x^k - y^k\|^2 - \|\varepsilon^k\| \|x^k - y^k\| \\ &\geq \mu_k (1 - \sigma) \|x^k - y^k\|^2 \\ &> 0, \end{aligned}$$

where the last inequality follows from (2.4). Thus whenever (2.4) holds, we have (2.5), and so H_k given by (2.3) indeed separates x^k from zeros of T . The last step of Algorithm 2.1 is equivalent to projecting x^k onto this hyperplane. Separation arguments show that the distance to the solution set for a sequence thus constructed monotonically decreases, which essentially ensures global convergence of the algorithm. Algorithm 2.1 has certain advantages over the classical proximal point method [37] in the sense of less restrictive and more constructive tolerance requirements imposed on approximate solutions of proximal subproblems. See [43] for a more detailed discussion. Other related works are [39, 41, 38, 42].

Of course, in the context of this paper straightforward application of Algorithm 2.1 to solving $\text{NCP}(F)$ is not practical, since this would involve solving a sequence of *nonlinear* subproblems which are in general just as difficult as the original NCP, even if they are better behaved due to the regularization. In this sense, the situation is similar to the method of [5] discussed in the introduction. However, the regularization and projection methodology of Algorithm 2.1 would prove to be crucial for devising a globally convergent method.

Given a current iterate x^k and a regularization parameter $\mu_k > 0$, consider the regularized linear complementarity problem $\text{LCP}(\varphi_k)$

$$(2.6) \quad x \geq 0, \quad \varphi_k(x) \geq 0, \quad \langle x, \varphi_k(x) \rangle = 0,$$

where

$$(2.7) \quad \varphi_k(x) := F(x^k) + G_k(x - x^k) + \mu_k(x - x^k)$$

with G_k being a positive semidefinite matrix (presumably, the Jacobian of F or its approximation, if F is differentiable at x^k). Suppose $z^k \geq 0$ is some approximate solution of this problem with e^k being the associated natural residual [2, 26]:

$$(2.8) \quad \min\{z^k; \varphi_k(z^k)\} = e^k,$$

where the minimum is taken componentwise (see (2.9), (2.10) in Algorithm 2.2 for conditions imposed on the error tolerance e^k). Note that since the matrix $G_k + \mu_k I$ is positive definite, $\text{LCP}(\varphi_k)$ always has (unique) solution \hat{z}^k [2] for which the residual is zero:

$$0 = \min\{\hat{z}^k; \varphi_k(\hat{z}^k)\}.$$

Hence, $\text{LCP}(\varphi_k)$ always has inexact solutions such that $\|e^k\| \leq \delta_k$, whichever $\delta_k \geq 0$ we choose. Therefore this inexact Newton step is well defined.

The next step of our algorithm is checking whether the inexact Newton point obtained by solving $\text{LCP}(\varphi_k)$ provides an acceptable (in the sense of Algorithm 2.1) approximate solution for the proximal point subproblem

$$0 \in (F + N)(x) + \mu_k(x - x^k).$$

If this is the case, the previously described separation property holds, and our algorithm proceeds to make the projection step prescribed by Algorithm 2.1.

To make this more precise, we provide the following considerations. Since

$$e^k = \min\{z^k; \varphi_k(z^k)\} = z^k - [z^k - \varphi_k(z^k)]^+,$$

we have that

$$z^k - e^k = [z^k - \varphi_k(z^k)]^+ \in \mathfrak{R}_+^n.$$

By properties of the projection operator [33, p. 121],

$$\langle z^k - \varphi_k(z^k) - (z^k - e^k), x - (z^k - e^k) \rangle \leq 0 \quad \text{for all } x \in \mathfrak{R}_+^n.$$

Therefore

$$\langle -\varphi_k(z^k) + e^k, x - (z^k - e^k) \rangle \leq 0 \quad \text{for all } x \in \mathfrak{R}_+^n,$$

which implies that

$$h^k := -\varphi_k(z^k) + e^k \in N(z^k - e^k),$$

so that we have available an element in the normal cone N at the point $z^k - e^k \in \mathfrak{R}_+^n$. Consider now the pair

$$y^k = z^k - e^k$$

and

$$v^k = F(z^k - e^k) - \varphi_k(z^k) + e^k \in (F + N)(z^k - e^k) = (F + N)(y^k).$$

Let us analyze y^k, v^k as an approximate solution of the proximal subproblem (2.1) in Algorithm 2.1 with $T = F + N$. We have to check whether (2.4), the stronger version of condition (2.2), is satisfied with

$$\varepsilon^k = -v^k - \mu_k(y^k - x^k)$$

and y^k, v^k given above. If (2.4) is satisfied, then (2.5) holds and the hyperplane H_k given by (2.3) separates the current iterate x^k from zeros of $T = F + N$ or, equivalently, from solutions of $\text{NCP}(F)$. Therefore we can make progress toward a solution of $\text{NCP}(F)$ by making the projection step of Algorithm 2.1 with y^k and v^k defined above (followed by projection onto \mathfrak{R}_+^n to preserve feasibility). As we shall see, the test (2.4) and the resulting step will be crucial for obtaining fast local convergence when the regularity assumption holds.

However, far from the solution or if regularity does not hold, an approximate (and even exact) Newton point obtained from solving the *linear* model may not satisfy the tolerance requirements (2.4) for the *nonlinear* proximal subproblem. In this case, the preceding separation arguments are not valid, and the Newton point has to be refined. For this task, we employ a linesearch procedure in the approximate Newton direction $z^k - x^k$ (see (2.12) in Algorithm 2.2) which computes a point $y^k = x^k + \alpha_k(z^k - x^k)$ such that

$$0 < \langle F(y^k), x^k - y^k \rangle.$$

A similar linesearch technique was used in [47, 40]. Note that for any \bar{x} which solves $\text{NCP}(F)$ we have

$$0 \geq \langle F(\bar{x}), \bar{x} - y^k \rangle,$$

where $y^k \in \mathfrak{R}_+^n$. Hence, by monotonicity of F ,

$$0 \geq \langle F(y^k), \bar{x} - y^k \rangle.$$

Therefore in this case another hyperplane, namely,

$$H_k := \{x \in \mathfrak{R}^n \mid \langle F(y^k), x - y^k \rangle = 0\},$$

strictly separates the current iterate x^k from solutions of the problem. Once the separating hyperplane H_k is obtained, the next iterate x^{k+1} is computed by projecting x^k onto H_k and then onto the nonnegative orthant \mathfrak{R}_+^n .

We now formally state the algorithm.

ALGORITHM 2.2. Choose any $x^0 \in \mathfrak{R}^n$, $\sigma \in (0, 1)$, $\beta \in (0, 1)$, and $\lambda \in (0, 1)$. Set $k := 0$.

Inexact Newton Step. Choose a positive semidefinite matrix G_k and $\mu_k > 0$. Choose $\rho_k \in [0, 1)$ and compute $z^k \in \mathfrak{R}_+^n$, an inexact solution of $LCP(\varphi_k)$ given by (2.6)–(2.8), such that

$$(2.9) \quad \|e^k\| \leq \rho_k \mu_k \|z^k - x^k\|$$

and

$$(2.10) \quad \langle e^k, \varphi_k(z^k) + z^k - x^k \rangle \leq \rho_k \mu_k \|z^k - x^k\|^2.$$

Stop if $z^k = x^k$. Otherwise,

Linesearch Step. Set

$$v^k := F(z^k - e^k) - \varphi_k(z^k) + e^k \quad \text{and} \quad y^k := z^k - e^k.$$

Let

$$\varepsilon^k = -v^k - \mu_k(y^k - x^k).$$

If

$$(2.11) \quad \|\varepsilon^k\| \leq \sigma \mu_k \|y^k - x^k\|,$$

then go to the Projection Step.

Otherwise, find $y^k = x^k + \alpha_k(z^k - x^k)$, where $\alpha_k = \beta^{m_k}$ with m_k being the smallest nonnegative integer m such that

$$(2.12) \quad \langle F(x^k + \beta^m(z^k - x^k)), x^k - z^k \rangle \geq \lambda(1 - \rho_k)\mu_k \|z^k - x^k\|^2.$$

Set $v^k := F(y^k)$ and go to the Projection Step.

Projection Step. Compute

$$(2.13) \quad x^{k+1} = \max \left\{ 0; x^k - \frac{\langle v^k, x^k - y^k \rangle}{\|v^k\|^2} v^k \right\}.$$

Set $k := k + 1$; and repeat.

To compute z^k , an approximate solution of $LCP(\varphi_k)$ satisfying (2.6)–(2.10), one can employ any appropriate algorithm known to converge for the strongly monotone linear complementarity problem. There are many algorithms which would generate a sequence converging to the unique solution of $LCP(\varphi_k)$ with a (global) quadratic rate. This guarantees that after finitely many iterations (just a few, one hopes), the $LCP(\varphi_k)$ residual e^k would be small enough, so that (2.9)–(2.10) holds.

Note that Algorithm 2.2 has computational cost per iteration of the same order as any other damped Newton method: solving, possibly inexact, a linear complementarity problem followed by a simple linesearch procedure (if (2.11) is not satisfied). The projection step is explicit, and therefore its computational cost is negligible. The advantage of Algorithm 2.2 over other Newton-type methods is that it is truly globally convergent under minimal assumptions.

3. Global convergence analysis. We start with the global convergence analysis. Throughout we assume that the solution set of the problem is nonempty. First note that if Algorithm 2.2 terminates with $z^k = x^k$, then $e^k = 0$ by (2.9), and we have that z^k is the exact solution of $\text{LCP}(\varphi_k)$. Therefore $\langle \varphi_k(z^k), x - z^k \rangle \geq 0$ for all $x \in \mathfrak{R}_+^n$. Because $z^k = x^k$ implies that $\varphi_k(z^k) = F(z^k)$, it follows from the latter inequality that z^k solves $\text{NCP}(F)$. From now on, we assume that $z^k \neq x^k$ for all k , and an infinite sequence $\{x^k\}$ is generated.

We first state a preliminary result [43] whose simple proof we include for completeness.

LEMMA 3.1. *Let x, y, v, \bar{x} be any elements of \mathfrak{R}^n such that*

$$\langle v, x - y \rangle > 0 \quad \text{and} \quad \langle v, \bar{x} - y \rangle \leq 0.$$

Let

$$\hat{x} = x - \frac{\langle v, x - y \rangle}{\|v\|^2} v.$$

Then

$$\|\hat{x} - \bar{x}\|^2 \leq \|x - \bar{x}\|^2 - \|\hat{x} - x\|^2.$$

Proof. It follows from the hypothesis that the hyperplane $H = \{s \mid \langle v, s - y \rangle = 0\}$ separates x from \bar{x} . Moreover, \hat{x} is the projection of x onto the half-space $\{s \mid \langle v, s - y \rangle \leq 0\}$. Since \bar{x} belongs to this half-space, it follows from the basic properties of the projection operator (see [33, p. 121]) that $\langle x - \hat{x}, \hat{x} - \bar{x} \rangle \geq 0$. Therefore

$$\begin{aligned} \|x - \bar{x}\|^2 &= \|x - \hat{x}\|^2 + \|\hat{x} - \bar{x}\|^2 + 2\langle x - \hat{x}, \hat{x} - \bar{x} \rangle \\ &\geq \|x - \hat{x}\|^2 + \|\hat{x} - \bar{x}\|^2, \end{aligned}$$

which completes the proof. \square

We are now ready to prove our main global convergence result. The remarkable property established in Theorem 3.2 is that the *whole* sequence of iterates always converges to some solution of $\text{NCP}(F)$ under the assumptions of merely monotonicity and continuity of F . $\text{NCP}(F)$ need not have a unique solution; in fact, the solution set may even be unbounded. In the latter case the level sets of any merit function for $\text{NCP}(F)$ are also unbounded, so that for typical Newton methods with globalization strategies based on merit functions, even boundedness of iterates cannot be established.

THEOREM 3.2. *Suppose that F is continuous and monotone. Then any sequence $\{x^k\}$ generated by Algorithm 2.2 is bounded.*

Suppose further that there exist constants $C_1, C_2, C_3 > 0$ and $t > 0$ such that $\|G_k\| \leq C_1$ for all k and, starting with some index k_0 ,

$$C_2 \geq \mu_k \geq C_3 \|r(x^k)\|^t.$$

Suppose that

$$\min\{1; 1/C_2\} > \limsup_{k \rightarrow \infty} \rho_k.$$

Then $\{x^k\}$ converges to some \bar{x} , which is a solution of $\text{NCP}(F)$.

Proof. First note that because

$$z^k - e^k = [z^k - \varphi_k(z^k)]^+$$

and $x^k \in \mathfrak{R}_+^n$, by properties of the projection operator [33, p. 121] it follows that

$$\langle z^k - \varphi_k(z^k) - (z^k - e^k), x^k - (z^k - e^k) \rangle \leq 0.$$

Therefore

$$\langle -\varphi_k(z^k) + e^k, x^k - z^k + e^k \rangle \leq 0.$$

Making use of the latter inequality, we further obtain

$$\begin{aligned} \langle F(x^k), x^k - z^k \rangle &\geq \langle F(x^k) - \varphi_k(z^k) + e^k, x^k - z^k + e^k \rangle - \langle F(x^k), e^k \rangle \\ &= \langle (G_k + \mu_k I)(x^k - z^k), x^k - z^k + e^k \rangle + \langle e^k, x^k - z^k - F(x^k) \rangle + \|e^k\|^2 \\ &\geq \mu_k \|z^k - x^k\|^2 - \langle e^k, (G_k + \mu_k I)(z^k - x^k) + F(x^k) + z^k - x^k \rangle \\ &= \mu_k \|z^k - x^k\|^2 - \langle e^k, \varphi_k(z^k) + z^k - x^k \rangle \\ (3.1) \quad &\geq \mu_k(1 - \rho_k) \|z^k - x^k\|^2, \end{aligned}$$

where the last inequality follows from (2.10) and the next to last follows from positive semidefiniteness of G_k .

We next show that the linesearch procedure (2.12), if activated, always terminates with a positive stepsize α_k . Suppose that for some iteration index k this is not the case. That is, for all integers m we have

$$\langle F(x^k + \beta^m(z^k - x^k)), x^k - z^k \rangle < \lambda(1 - \rho_k)\mu_k \|z^k - x^k\|^2.$$

Since F is continuous, passing onto the limit as $m \rightarrow \infty$, we obtain

$$(3.2) \quad \langle F(x^k), x^k - z^k \rangle \leq \lambda(1 - \rho_k)\mu_k \|z^k - x^k\|^2.$$

Now since $\lambda \in (0, 1)$, $\rho_k \in [0, 1)$, and $z^k \neq x^k$, (3.2) contradicts (3.1). Therefore the linesearch step is well defined.

Denote

$$(3.3) \quad \hat{x}^k := x^k - \frac{\langle v^k, x^k - y^k \rangle}{\|v^k\|^2} v^k.$$

Observe that this is well defined because $v^k \neq 0$. Indeed, if $v^k = 0$ in the case when the linesearch is not activated, then (2.11) must hold. Furthermore, in that case $v^k = 0$ implies $\varepsilon^k = -\mu_k(y^k - x^k)$, which together with (2.11) implies that $\varepsilon^k = y^k - x^k = 0$. Therefore, using the definition of y^k , we conclude that $x^k = z^k - e^k$. If $\mu_k \leq 1$, then $\rho_k \mu_k < 1$, so (2.9) implies that $z^k = x^k$, in contradiction with the stopping test. Suppose now that $\mu_k > 1$. Note that $0 = v^k = F(x^k) - \varphi_k(z^k) + e^k = (-G_k + (1 - \mu_k)I)e^k$. It then follows that $0 = \langle (-G_k + (1 - \mu_k)I)e^k, e^k \rangle \leq (1 - \mu_k)\|e^k\|^2$. Since $1 < \mu_k$, we again have that $e^k = 0$, i.e., $z^k = x^k$, which contradicts the stopping rule. If the linesearch is used to compute y^k , then $v^k = F(y^k)$, which again cannot be zero by (2.12). The observation that $v^k \neq 0$ concludes the proof that the whole algorithm is well defined.

With the notation (3.3), we have that $x^{k+1} = [\hat{x}^k]^+$. Let \bar{x} be any solution of $\text{NCP}(F)$. Since $\bar{x} \in \mathfrak{R}_+^n$, it is easy to see that

$$(3.4) \quad \|x^{k+1} - \bar{x}\| \leq \|\hat{x}^k - \bar{x}\|.$$

Note that if (2.11) is satisfied, then $\langle v^k, x^k - y^k \rangle > 0$ (see (2.5)). If (2.11) is not satisfied, then $\langle v^k, x^k - y^k \rangle = \alpha_k \langle F(y^k), x^k - z^k \rangle > 0$ still holds by (2.12). In this respect, the only difference is the choice of y^k and v^k . As discussed before, in either case $\langle v^k, \bar{x} - y^k \rangle \leq 0$, so that by Lemma 3.1, it follows that

$$\|\hat{x}^k - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|\hat{x}^k - x^k\|^2.$$

Combining the latter relation with (3.4), we obtain

$$(3.5) \quad \|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|\hat{x}^k - x^k\|^2.$$

It immediately follows that the sequence $\{\|x^k - \bar{x}\|\}$ is monotone, so it converges. Therefore, $\{x^k\}$ is bounded.

We next consider the two possible cases:

$$(3.6) \quad 0 = \liminf_{k \rightarrow \infty} \|r(x^k)\|$$

and

$$(3.7) \quad 0 < \liminf_{k \rightarrow \infty} \|r(x^k)\|.$$

In the first case, by continuity of $r(\cdot)$ and boundedness of $\{x^k\}$, there exists \tilde{x} , an accumulation point of $\{x^k\}$, such that $r(\tilde{x}) = 0$. Therefore \tilde{x} is a solution of $\text{NCP}(F)$. Since \bar{x} was an arbitrary solution, we can now choose $\bar{x} = \tilde{x}$ in (3.5). Because the sequence $\{\|x^k - \tilde{x}\|\}$ converges and \tilde{x} is an accumulation point of $\{x^k\}$, it must be the case that $\{x^k\}$ converges to \tilde{x} , which is a solution of $\text{NCP}(F)$.

Consider now the second case. By (3.5), it follows that

$$0 = \lim_{k \rightarrow \infty} \|\hat{x}^k - x^k\|$$

or, equivalently,

$$(3.8) \quad 0 = \lim_{k \rightarrow \infty} \frac{\langle v^k, x^k - y^k \rangle}{\|v^k\|}.$$

By the choice of μ_k and (3.7), it then follows that $\mu_k \geq C_3 \|r(x^k)\|^t \geq C_4 > 0$ for all k . By (3.1) and the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} \|F(x^k)\| \|x^k - z^k\| &\geq \langle F(x^k), x^k - z^k \rangle \\ &\geq \mu_k (1 - \rho_k) \|x^k - z^k\|^2. \end{aligned}$$

Hence,

$$\|F(x^k)\| \geq C_4 (1 - \rho_k) \|x^k - z^k\|.$$

Taking into account boundedness of $\{x^k\}$ and continuity of F , and the fact that $1 > \limsup_{k \rightarrow \infty} \rho_k$, we conclude that the sequence $\{x^k - z^k\}$ is bounded. It now easily follows that the sequences $\{z^k\}$, $\{e^k\}$, and $\{y^k\}$ are all bounded.

By the triangle and Cauchy–Schwarz inequalities, and the nonexpansiveness of the projection operator, we have

$$\begin{aligned} \|x^k - z^k\| &\geq \|x^k - (z^k - e^k)\| - \|e^k\| \\ &= \|x^k - [z^k - \varphi_k(z^k)]^+ \| - \|e^k\| \\ &\geq \|x^k - [x^k - F(x^k)]^+ \| - \|[x^k - F(x^k)]^+ - [z^k - \varphi_k(z^k)]^+ \| - \|e^k\| \\ &\geq \|r(x^k)\| - \|x^k - z^k - F(x^k) + \varphi_k(z^k)\| - \|e^k\| \\ &\geq \|r(x^k)\| - (1 + \rho_k \mu_k) \|x^k - z^k\| - \|(G_k + \mu_k I)(x^k - z^k)\| \\ &\geq \|r(x^k)\| - (1 + C_1 + 2C_2) \|x^k - z^k\|. \end{aligned}$$

Hence,

$$(2 + C_1 + 2C_2)\|x^k - z^k\| \geq \|r(x^k)\|.$$

It follows from (3.7) that

$$(3.9) \quad 0 < \liminf_{k \rightarrow \infty} \|x^k - z^k\|.$$

Suppose that condition (2.11) in Algorithm 2.2 holds an infinite number of times. For such iterations k , by (2.5), we have (recall also that $v^k \neq 0$)

$$(3.10) \quad \begin{aligned} \frac{\langle v^k, x^k - y^k \rangle}{\|v^k\|} &\geq \frac{\mu_k(1 - \sigma)\|x^k - y^k\|^2}{\|v^k\|} \\ &\geq \frac{C_4(1 - \sigma)\|x^k - z^k + e^k\|^2}{\|F(z^k - e^k) - \varphi_k(z^k) + e^k\|}. \end{aligned}$$

Since $\|G_k\| \leq C_1$, $\mu_k \leq C_2$ and $\{z^k\}$, $\{e^k\}$ are bounded, it follows that $\{F(z^k - e^k)\}$ and $\{\varphi_k(z^k)\}$ are bounded. Therefore, for some $C_5 > 0$,

$$\frac{C_4(1 - \sigma)}{\|F(z^k - e^k) - \varphi_k(z^k) + e^k\|} \geq C_5.$$

Passing onto the limit in (3.10) (along the indices k for which (2.11) holds) and taking into account (3.8), we obtain that

$$0 = \liminf_{k \rightarrow \infty} \|x^k - z^k + e^k\|.$$

By the triangle inequality and (2.9), we have

$$\begin{aligned} \|x^k - z^k + e^k\| &\geq \|x^k - z^k\| - \|e^k\| \\ &\geq (1 - \rho_k \mu_k)\|x^k - z^k\| \\ &\geq (1 - \rho_k C_2)\|x^k - z^k\|. \end{aligned}$$

Because $1/C_2 > \limsup_{k \rightarrow \infty} \rho_k$, we further conclude that

$$0 = \liminf_{k \rightarrow \infty} \|x^k - z^k\|$$

which contradicts (3.9). We conclude that if $0 < \liminf_{k \rightarrow \infty} \|r(x^k)\|$, then condition (2.11) in Algorithm 2.2 may hold no more than a finite number of times.

Hence, we can assume that for all k sufficiently large, y^k and v^k are obtained through the linesearch step (2.12), in which case

$$\begin{aligned} \frac{\langle v^k, x^k - y^k \rangle}{\|v^k\|} &= \frac{\alpha_k \langle F(y^k), x^k - z^k \rangle}{\|F(y^k)\|} \\ &\geq \frac{\alpha_k \lambda (1 - \rho_k) \mu_k \|x^k - z^k\|^2}{\|F(y^k)\|}. \end{aligned}$$

Using (3.8) and taking into account the boundedness of $\{F(y^k)\}$ and the fact that $\mu_k \geq C_4$ and $1 > \limsup_{k \rightarrow \infty} \rho_k$, we obtain that

$$0 = \lim_{k \rightarrow \infty} \alpha_k \|x^k - z^k\|.$$

Now, because of (3.9), we conclude that it must be the case that

$$0 = \lim_{k \rightarrow \infty} \alpha_k.$$

The latter is equivalent to saying that $m_k \rightarrow \infty$. It follows that for every sufficiently large k , the stepsize is decreased at least twice, i.e., $m_k \geq 2$. Hence, the stepsize rule (2.12) is not satisfied for the value of β^{m_k-1} , i.e.,

$$\langle F(x^k + \beta^{m_k-1}(z^k - x^k)), x^k - z^k \rangle < \lambda(1 - \rho_k)\mu_k \|z^k - x^k\|^2.$$

Taking into account boundedness of the sequences $\{x^k\}$, $\{\mu_k\}$, $\{\rho_k\}$, and $\{z^k\}$, and passing onto a subsequence if necessary, as $k \rightarrow \infty$ we obtain

$$\langle F(\hat{x}), \hat{x} - \hat{z} \rangle \leq \lambda(1 - \hat{\rho})\hat{\mu} \|\hat{z} - \hat{x}\|^2,$$

where \hat{x} , $\hat{\mu}$, $\hat{\rho}$, and \hat{z} are limits of corresponding subsequences. On the other hand, passing onto the limit in (3.1), we have that

$$\langle F(\hat{x}), \hat{x} - \hat{z} \rangle \geq (1 - \hat{\rho})\hat{\mu} \|\hat{z} - \hat{x}\|^2.$$

Taking into account that $\hat{\mu} > 0$ and $\|\hat{z} - \hat{x}\| > 0$ (by (3.9)), and $\hat{\rho} \leq \limsup_{k \rightarrow \infty} \rho_k < 1$, the last two relations are a contradiction because $\lambda \in (0, 1)$. Hence the case $0 < \liminf_{k \rightarrow \infty} \|r(x^k)\|$ is not possible.

This completes the proof. \square

Remark. Since boundedness of $\{x^k\}$ was established without any boundedness assumptions on $\{G_k\}$, for the special choice of $G_k = \nabla F(x^k)$ the condition $\|G_k\| \leq C_1$ can be removed, due to the continuity of $\nabla F(\cdot)$.

4. Local convergence analysis. The following *error bound* [29] result will be crucial for establishing the superlinear rate of convergence of our algorithm. Note that this error bound actually holds under the more general assumption of \bar{x} being a *regular* solution [35, 26, 14]. Positive definiteness of $\nabla F(\bar{x})$ is a simple sufficient condition for \bar{x} to be regular [31]. Here we state only the simplified result.

LEMMA 4.1 ([14, Proposition 4.4]). *If \bar{x} is a solution of NCP(F) where $\nabla F(\bar{x})$ is positive definite, then there exist a constant $\theta > 0$ and a neighborhood B of \bar{x} such that*

$$\|x - \bar{x}\| \leq \theta \|r(x)\|$$

for all $x \in B$.

We will also need the following error bound result for strongly monotone linear complementarity problems [27]. This result is actually related to Lemma 4.1, but note that the constant θ can be estimated explicitly and the error bound holds globally.

LEMMA 4.2. *Let \hat{z} be the solution of the linear complementarity problem*

$$z \geq 0, \quad Mz + q \geq 0, \quad \langle Mz + q, z \rangle = 0,$$

where M is a positive definite matrix and q is an arbitrary vector. For all $z \in \Re^n$ it holds that

$$\|z - \hat{z}\| \leq \frac{1 + \|M\|}{C(M)} \|\min\{z; Mz + q\}\|,$$

where $C(M) > 0$ is the smallest eigenvalue of $(M + M^T)/2$.

We are now ready to establish the superlinear convergence of our algorithm for solving the monotone nonlinear complementarity problems under the assumptions of positive definiteness of ∇F at the solution and its local Hölder continuity. The proof relies on the fact that in a sufficiently small neighborhood of such a solution, the approximate Newton point computed by Algorithm 2.2 satisfies the error tolerance requirements of Algorithm 2.1 for solving the corresponding nonlinear proximal point subproblem. Therefore the corresponding projection step in Algorithm 2.2 is taken immediately, and the linesearch procedure is not used.

THEOREM 4.3. *Let F be monotone and continuous on \mathfrak{R}^n . Let \bar{x} be the (unique) solution of $NCP(F)$ at which F is differentiable with $\nabla F(\bar{x})$ positive definite. Let ∇F be locally Hölder continuous around \bar{x} with degree $p \in (0, 1]$. Suppose that the assumptions of Theorem 3.2 are satisfied. In addition, suppose that*

$$\mu_k = \|r(x^k)\|^t, \quad t \in (0, p),$$

$$0 = \lim_{k \rightarrow \infty} \rho_k,$$

and starting with some index k_0 , $G_k = \nabla F(x^k)$.

Then the sequence $\{x^k\}$ converges to \bar{x} Q -superlinearly.

Proof. By Theorem 3.2, we already know that the sequence $\{x^k\}$ converges to \bar{x} . (Note that monotonicity of F and nonsingularity of $\nabla F(\bar{x})$ imply that \bar{x} is the unique solution.) Note that by the choice of μ_k , we have that $0 = \lim_{k \rightarrow \infty} \mu_k$.

By Hölder continuity of ∇F in the neighborhood of \bar{x} , it follows that there exists some $L > 0$ such that for all $u \in \mathfrak{R}^n$ sufficiently small and all indices k sufficiently large

$$\|\nabla F(x^k + u) - \nabla F(x^k)\| \leq L\|u\|^p, \quad p \in (0, 1].$$

Therefore

$$\begin{aligned} F(x^k + u) - F(x^k) &= \int_0^1 \nabla F(x^k + su)u \, ds \\ &= \nabla F(x^k)u + \int_0^1 (\nabla F(x^k + su) - \nabla F(x^k))u \, ds \\ &= \nabla F(x^k)u + R^k(u), \end{aligned}$$

where

$$\begin{aligned} \|R^k(u)\| &\leq \int_0^1 \|\nabla F(x^k + su) - \nabla F(x^k)\| \|u\| \, ds \\ &\leq L \int_0^1 s^p \|u\|^{1+p} \, ds \\ &= \frac{L}{1+p} \|u\|^{1+p}. \end{aligned}$$

Hence, defining $C_6 := L/(1+p)$, we have that

$$(4.1) \quad \begin{aligned} F(x^k + u) - F(x^k) - \nabla F(x^k)u &= R^k(u), \\ \|R^k(u)\| &\leq C_6 \|u\|^{1+p}. \end{aligned}$$

We next show that (2.11) is always satisfied for all k sufficiently large, so that the linesearch step is never used. Thus let $y^k = z^k - e^k$ and $v^k = F(z^k - e^k) - \varphi_k(z^k) + e^k$. Assume that k is sufficiently large so that (4.1) holds and $y^k - x^k$ is sufficiently small (later we shall verify that $\{y^k - x^k\} \rightarrow 0$). Then we have that

$$\begin{aligned}
 -\varepsilon^k &= v^k + \mu_k(y^k - x^k) \\
 &= F(z^k - e^k) - \varphi_k(z^k) + e^k + \mu_k(z^k - e^k - x^k) \\
 &= F(z^k - e^k) - F(x^k) - \nabla F(x^k)(z^k - e^k - x^k) - \nabla F(x^k)e^k + (1 - \mu_k)e^k \\
 (4.2) \quad &= R^k(z^k - e^k - x^k) - \nabla F(x^k)e^k + (1 - \mu_k)e^k.
 \end{aligned}$$

By (4.1) and the Cauchy–Schwarz and triangle inequalities, it follows that

$$\begin{aligned}
 \|\varepsilon^k\| &\leq C_6\|y^k - x^k\|^{1+p} + (C_1 + 1)\|e^k\| \\
 (4.3) \quad &\leq C_6\|y^k - x^k\|^{1+p} + (C_1 + 1)\rho_k\mu_k\|z^k - x^k\|.
 \end{aligned}$$

Furthermore,

$$\begin{aligned}
 \|z^k - x^k\| &\leq \|y^k - x^k\| + \|e^k\| \\
 &\leq \|y^k - x^k\| + \rho_k\mu_k\|z^k - x^k\|.
 \end{aligned}$$

Hence,

$$(4.4) \quad \|z^k - x^k\| \leq C_7\|y^k - x^k\|,$$

where $C_7 \geq 1/(1 - \rho_k\mu_k)$ (recall that $\rho_k\mu_k \rightarrow 0$). Defining $C_8 := C_7(C_1 + 1)$, and combining (4.4) with (4.3), we obtain

$$\begin{aligned}
 \|\varepsilon^k\| &\leq (C_6\|y^k - x^k\|^p + C_8\rho_k\mu_k)\|y^k - x^k\| \\
 &= \left(C_6 \frac{\|y^k - x^k\|^p}{\|r(x^k)\|^t} + C_8\rho_k \right) \mu_k\|y^k - x^k\|.
 \end{aligned}$$

Clearly, condition (2.11) is satisfied whenever

$$C_6 \frac{\|y^k - x^k\|^p}{\|r(x^k)\|^t} + C_8\rho_k \leq \sigma.$$

Since

$$0 = \lim_{k \rightarrow \infty} \rho_k,$$

the latter relation is satisfied for all indices k sufficiently large if it holds that

$$(4.5) \quad 0 = \lim_{k \rightarrow \infty} \frac{\|y^k - x^k\|^p}{\|r(x^k)\|^t}.$$

Thus for establishing that the linesearch procedure is never used for indices k sufficiently large, it is left to prove (4.5). (This would also imply that $\{y^k - x^k\} \rightarrow 0$.)

Let \hat{z}^k be the exact solution of LCP(φ_k). It holds that

$$\langle F(\bar{x}) - \varphi_k(\hat{z}^k), \bar{x} - \hat{z}^k \rangle = -\langle \hat{z}^k, F(\bar{x}) \rangle - \langle \bar{x}, \varphi_k(\hat{z}^k) \rangle \leq 0,$$

because $\bar{x}, F(\bar{x}), \hat{z}^k, \varphi_k(\hat{z}^k)$ are all nonnegative and $\langle \bar{x}, F(\bar{x}) \rangle = \langle \hat{z}^k, \varphi_k(\hat{z}^k) \rangle = 0$. Furthermore, by (4.1) we obtain

$$\begin{aligned} F(\bar{x}) - \varphi_k(\hat{z}^k) &= F(\bar{x}) - F(x^k) - (\nabla F(x^k) + \mu_k I)(\hat{z}^k - x^k) \\ &= F(\bar{x}) - F(x^k) - \nabla F(x^k)(\bar{x} - x^k) - \nabla F(x^k)(\hat{z}^k - \bar{x}) - \mu_k(\hat{z}^k - x^k) \\ &= R^k(\bar{x} - x^k) - \nabla F(x^k)(\hat{z}^k - \bar{x}) - \mu_k(\hat{z}^k - x^k). \end{aligned}$$

Hence,

$$(4.6) \quad \begin{aligned} 0 &\geq \langle F(\bar{x}) - \varphi_k(\hat{z}^k), \bar{x} - \hat{z}^k \rangle \\ &= \langle R^k(\bar{x} - x^k), \bar{x} - \hat{z}^k \rangle - \langle \nabla F(x^k)(\hat{z}^k - \bar{x}), \bar{x} - \hat{z}^k \rangle - \mu_k \langle \hat{z}^k - x^k, \bar{x} - \hat{z}^k \rangle. \end{aligned}$$

By positive definiteness of $\nabla F(\bar{x})$, it follows that for some $C_9 > 0$ and all indices k sufficiently large

$$(4.7) \quad \langle \nabla F(x^k)d, d \rangle \geq C_9 \|d\|^2 \quad \text{for all } d \in \mathfrak{R}^n.$$

By (4.6) and the latter relation we obtain

$$\begin{aligned} C_9 \|\hat{z}^k - \bar{x}\|^2 &\leq \langle \nabla F(x^k)(\hat{z}^k - \bar{x}), \hat{z}^k - \bar{x} \rangle \\ &\leq \langle R^k(\bar{x} - x^k), \hat{z}^k - \bar{x} \rangle + \mu_k \langle \hat{z}^k - x^k, \bar{x} - \hat{z}^k \rangle \\ &\leq C_6 \|x^k - \bar{x}\|^{1+p} \|\hat{z}^k - \bar{x}\| + \mu_k \|\hat{z}^k - x^k\| \|\hat{z}^k - \bar{x}\|, \end{aligned}$$

where the last inequality follows from (4.1) and the Cauchy–Schwarz inequality. Therefore,

$$(4.8) \quad C_9 \|\hat{z}^k - \bar{x}\| \leq C_6 \|x^k - \bar{x}\|^{1+p} + \mu_k \|\hat{z}^k - x^k\|.$$

By the triangle inequality,

$$\|\hat{z}^k - x^k\| \leq \|\hat{z}^k - \bar{x}\| + \|x^k - \bar{x}\|.$$

Combining the latter inequality with (4.8), we obtain

$$\|\hat{z}^k - x^k\| \leq C_9^{-1} (C_6 \|x^k - \bar{x}\|^{1+p} + \mu_k \|\hat{z}^k - x^k\|) + \|x^k - \bar{x}\|.$$

Hence,

$$(1 - \mu_k/C_9) \|\hat{z}^k - x^k\| \leq C_6/C_9 \|x^k - \bar{x}\|^{1+p} + \|x^k - \bar{x}\|.$$

And for $C_{11} \geq 1/(1 - \mu_k/C_9)$ (recall that $\mu_k \rightarrow 0$), we obtain

$$(4.9) \quad \|\hat{z}^k - x^k\| \leq C_{11} (1 + C_6/C_9 \|x^k - \bar{x}\|^p) \|x^k - \bar{x}\|.$$

By the triangle inequality,

$$(4.10) \quad \|z^k - x^k\| \leq \|\hat{z}^k - z^k\| + \|\hat{z}^k - x^k\|.$$

Furthermore by Lemma 4.2, we have that

$$\|z^k - \hat{z}^k\| \leq \frac{1 + \|\nabla F(x^k) + \mu_k I\|}{c_k} \|e^k\|,$$

where c_k is the smallest eigenvalue of $(\nabla F(x^k) + \nabla F(x^k)^\top)/2 + \mu_k I$. In view of (4.7), clearly $c_k \geq C_9$. Therefore

$$\|z^k - \hat{z}^k\| \leq \frac{2 + C_1}{C_9} \|e^k\|.$$

Furthermore, defining $C_{10} := (2 + C_1)/C_9$, we obtain

$$(4.11) \quad \|z^k - \hat{z}^k\| \leq C_{10} \rho_k \mu_k \|z^k - x^k\|.$$

Combining the latter relation with (4.10), we have that

$$\|z^k - x^k\| \leq C_{10} \rho_k \mu_k \|z^k - x^k\| + \|\hat{z}^k - x^k\|.$$

Since $\rho_k \mu_k \rightarrow 0$, taking into account (4.9), it is now clear that there exists $C_{12} > 0$ such that

$$(4.12) \quad \|z^k - x^k\| \leq C_{12} \|x^k - \bar{x}\|.$$

Note that

$$(4.13) \quad \begin{aligned} \|y^k - x^k\| &\leq \|z^k - x^k\| + \|e^k\| \\ &\leq (1 + \rho_k \mu_k) \|z^k - x^k\| \\ &\leq 2 \|z^k - x^k\| \\ &\leq 2C_{12} \|x^k - \bar{x}\|, \end{aligned}$$

where the last inequality follows from (4.12). Hence, by Lemma 4.1,

$$(4.14) \quad \|y^k - x^k\| \leq 2C_{12} \theta \|r(x^k)\|.$$

Therefore

$$(4.15) \quad \frac{\|y^k - x^k\|^p}{\|r(x^k)\|^t} \leq 2C_{12} \theta \|r(x^k)\|^{p-t},$$

which establishes (4.5) by the choice of $t \in (0, p)$.

Hence, from now on, we can assume that the linesearch is never used, so that $y^k = z^k - e^k$ and $v^k = F(z^k - e^k) - \varphi_k(z^k) + e^k$. Recalling the definition

$$\hat{x}^k := x^k - \frac{\langle v^k, x^k - y^k \rangle}{\|v^k\|^2} v^k,$$

it is easy to see that

$$\|x^{k+1} - \bar{x}\| \leq \|\hat{x}^k - \bar{x}\|$$

because $\bar{x} \in \mathfrak{R}_+^n$ and x^{k+1} is the orthogonal projection of \hat{x}^k onto \mathfrak{R}_+^n . Applying further the triangle inequality, we obtain

$$(4.16) \quad \|x^{k+1} - \bar{x}\| \leq \|\hat{x}^k - y^k\| + \|y^k - z^k\| + \|z^k - \hat{z}^k\| + \|\hat{z}^k - \bar{x}\|.$$

We proceed to analyze the four terms in the right-hand side of (4.16).

For the second term in (4.16) we have

$$\|y^k - z^k\| = \|e^k\| \leq \rho_k \mu_k \|z^k - x^k\|,$$

and for the third term (see (4.11)) we have

$$\|z^k - \hat{z}^k\| \leq C_{10}\|e^k\| \leq C_{10}\rho_k\mu_k\|z^k - x^k\|.$$

Using (4.12) we obtain

$$(4.17) \quad \begin{aligned} \|y^k - z^k\| + \|z^k - \hat{z}^k\| &\leq (1 + C_{10})\rho_k\mu_k\|z^k - x^k\| \\ &\leq (1 + C_{10})C_{12}\rho_k\|r(x^k)\|^t\|x^k - \bar{x}\|. \end{aligned}$$

Next, we consider the last term in (4.16). By (4.8), we have

$$\begin{aligned} C_9\|\hat{z}^k - \bar{x}\| &\leq C_6\|x^k - \bar{x}\|^{1+p} + \mu_k\|\hat{z}^k - x^k\| \\ &\leq C_6\|x^k - \bar{x}\|^{1+p} + \mu_k(\|\hat{z}^k - \bar{x}\| + \|x^k - \bar{x}\|), \end{aligned}$$

where the last inequality follows from the triangle inequality. Therefore,

$$(C_9 - \mu_k)\|\hat{z}^k - \bar{x}\| \leq C_6\|x^k - \bar{x}\|^{1+p} + \mu_k\|x^k - \bar{x}\|.$$

Since $\mu_k \rightarrow 0$, for $C_{13} \geq \max\{1; C_6\}/(C_9 - \mu_k)$, we have

$$(4.18) \quad \|\hat{z}^k - \bar{x}\| \leq C_{13}(\|x^k - \bar{x}\|^p + \|r(x^k)\|^t)\|x^k - \bar{x}\|.$$

Finally, we consider the first term in the right-hand side of (4.16). Since the point \hat{x}^k is the projection of x^k onto the hyperplane $H_k = \{x \mid \langle v^k, x - y^k \rangle = 0\}$, and $y^k \in H_k$, the vectors $\hat{x}^k - x^k$ and $\hat{x}^k - y^k$ are orthogonal. Hence,

$$(4.19) \quad \|\hat{x}^k - y^k\| = \|y^k - x^k\| \sin \xi_k,$$

where ξ_k is the angle between $\hat{x}^k - x^k$ and $y^k - x^k$. Because $\hat{x}^k - x^k = -s_k v^k$ for a certain $s_k > 0$, the angle between the vectors v^k and $-\mu_k(y^k - x^k)$ is also ξ_k . Observe from (4.2) that

$$v^k = -\mu_k(y^k - x^k) + R^k(y^k - x^k) - \nabla F(x^k)e^k + (1 - \mu_k)e^k.$$

Given the above relation, \sin of the angle between v^k and $-\mu_k(y^k - x^k)$ can be easily bounded:

$$\begin{aligned} \sin \xi_k &\leq \frac{\|R^k(y^k - x^k) + ((1 - \mu_k)I - \nabla F(x^k))e^k\|}{\mu_k\|y^k - x^k\|} \\ &\leq \frac{C_6\|y^k - x^k\|^{1+p} + (1 + C_1)\|e^k\|}{\mu_k\|y^k - x^k\|} \\ &\leq C_6 \frac{\|y^k - x^k\|^p}{\|r(x^k)\|^t} + (1 + C_1) \frac{\rho_k\|z^k - x^k\|}{\|y^k - x^k\|} \\ &\leq 2C_6C_{12}\theta\|r(x^k)\|^{p-t} + C_7(1 + C_1)\rho_k, \end{aligned}$$

where the second inequality follows from (4.1) and the last inequality follows from (4.15) and (4.4). Hence, for some $C_{14} > 0$,

$$(4.20) \quad \sin \xi_k \leq C_{14}(\|r(x^k)\|^{p-t} + \rho_k).$$

By (4.19), using (4.13) and (4.20), we obtain

$$(4.21) \quad \begin{aligned} \|\hat{x}^k - y^k\| &= \|y^k - x^k\| \sin \xi_k \\ &\leq 2C_{12}C_{14}(\|r(x^k)\|^{p-t} + \rho_k)\|x^k - \bar{x}\|. \end{aligned}$$

By (4.16), combining (4.21), (4.17), and (4.18), we conclude that there exists $C_{15} > 0$ such that

$$(4.22) \quad \begin{aligned} \|x^{k+1} - \bar{x}\| &\leq C_{15} (\|r(x^k)\|^{p-t} + \rho_k + (1 + \rho_k)\|r(x^k)\|^t + \|x^k - \bar{x}\|^p) \\ &\quad \times \|x^k - \bar{x}\|, \end{aligned}$$

which means that $\{x^k\}$ converges to \bar{x} at least superlinearly. \square

Remark. Finally, we note that

$$\begin{aligned} \|r(x)\| &= \|x - [x - F(x)]^+ - \bar{x} + [\bar{x} - F(\bar{x})]^+\| \\ &\leq \|x - \bar{x}\| + \|[x - F(x)]^+ - [\bar{x} - F(\bar{x})]^+\| \\ &\leq (2 + C_{16})\|x - \bar{x}\|, \end{aligned}$$

where C_{16} is the constant of local Lipschitz continuity of F . Therefore if one further chooses

$$\rho_k = \|r(x^k)\|^p,$$

relation (4.22) implies that the order of superlinear convergence is at least $1 + p - t$, where $t \in (0, p)$.

5. Concluding remarks. We presented a new globalization strategy for the Newton method applied to the monotone nonlinear complementarity problem. Our strategy is based on the projection–proximal point methodology and makes full use of the monotonicity structure of the problem. The resulting hybrid algorithm is truly globally convergent to a solution without any additional assumptions, even if the solution set is unbounded. This is an important property which is not possessed by globalization approaches based on merit functions. Under natural assumptions, locally superlinear rate of convergence was also established.

Acknowledgments. We thank the two anonymous referees and the associate editor for constructive suggestions which helped us to improve the paper.

REFERENCES

- [1] J. F. BONNANS, *Local analysis of Newton-type methods for variational inequalities and nonlinear programming*, Appl. Math. Optim., 29 (1994), pp. 161–186.
- [2] R. W. COTTLE, J.-S. PANG, AND R. E. STONE, *The Linear Complementarity Problem*, Academic Press, New York, 1992.
- [3] S. P. DIRKSE AND M. C. FERRIS, *The PATH solver: A nonmonotone stabilization scheme for mixed complementarity problems*, Optim. Methods Softw., 5 (1995), pp. 123–156.
- [4] F. FACCHINEI, A. FISCHER, AND C. KANZOW, *Inexact Newton methods for semismooth equations with applications to variational inequality problems*, in Nonlinear Optimization and Applications, G. D. Pillo and F. Giannessi, eds., Plenum Press, New York, 1996, pp. 125–139.
- [5] F. FACCHINEI AND C. KANZOW, *Beyond monotonicity in regularization methods for nonlinear complementarity problems*, SIAM J. Control Optim., 37 (1999), pp. 150–1161.
- [6] F. FACCHINEI AND J. SOARES, *A new merit function for nonlinear complementarity problems and a related algorithm*, SIAM J. Optim., 7 (1997), pp. 225–247.
- [7] M. C. FERRIS AND J.-S. PANG, EDS., *Complementarity and Variational Problems: State of the Art*, SIAM, Philadelphia, PA, 1997.
- [8] M. C. FERRIS, C. KANZOW, AND T. S. MUNSON, *Feasible descent algorithms for mixed complementarity problems*, Math. Programming, 86 (1999), pp. 475–497.
- [9] A. FISCHER, *A special Newton-type optimization method*, Optimization, 24 (1992), pp. 269–284.

- [10] A. FISCHER, *An NCP function and its use for the solution of complementarity problems*, in Recent Advances in Nonsmooth Optimization, D.-Z. Du, L. Qi, and R. Womersley, eds., World Scientific Publishers, Singapore, 1995, pp. 88–105.
- [11] M. FUKUSHIMA, *Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems*, Math. Programming, 53 (1992), pp. 99–110.
- [12] M. FUKUSHIMA, *Merit functions for variational inequality and complementarity problems*, in Nonlinear Optimization and Applications, G. D. Pillo and F. Giannessi, eds., Plenum Press, New York, 1996, pp. 155–170.
- [13] J. HAN AND D. SUN, *Newton-Type Methods for Variational Inequalities*, manuscript, 1997.
- [14] P. HARKER AND J.-S. PANG, *Finite-dimensional variational inequality problems: A survey of theory, algorithms and applications*, Math. Programming, 48 (1990), pp. 161–220.
- [15] H. JIANG AND L. QI, *A new nonsmooth equations approach to nonlinear complementarity problems*, SIAM J. Control Optim., 35 (1997), pp. 178–193.
- [16] H. JIANG AND D. RALPH, *Global and local superlinear convergence analysis of Newton-type methods for semismooth equations with smooth least squares*, in Reformulation—Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Norwell, MA, 1998, pp. 181–210.
- [17] N. JOSEPHY, *Newton's Method for Generalized Equations*, Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, WI, 1979.
- [18] C. KANZOW, *Some equation-based methods for the nonlinear complementarity problem*, Optim. Methods Softw., 3 (1994), pp. 327–340.
- [19] C. KANZOW, *Nonlinear complementarity as unconstrained optimization*, J. Optim. Theory Appl., 88 (1996), pp. 139–155.
- [20] C. KANZOW AND M. FUKUSHIMA, *Theoretical and numerical investigation of the D-gap function for box constrained variational inequalities*, Math. Programming, 83 (1998), pp. 55–87.
- [21] T. D. LUCA, F. FACCHINEI, AND C. KANZOW, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Programming, 75 (1996), pp. 407–439.
- [22] Z.-Q. LUO AND P. TSENG, *A new class of merit functions for the nonlinear complementarity problem*, in Complementarity and Variational Problems: State of the art, M. C. Ferris and J.-S. Pang, eds., SIAM, Philadelphia, PA, 1997, pp. 204–225.
- [23] O. L. MANGASARIAN AND M. V. SOLODOV, *Nonlinear complementarity as unconstrained and constrained minimization*, Math. Programming, 62 (1993), pp. 277–297.
- [24] O. L. MANGASARIAN AND M. V. SOLODOV, *A linearly convergent derivative-free descent method for strongly monotone complementarity problems*, Comput. Optim. Appl., 14 (1999), pp. 5–16.
- [25] P. MARCOTTE AND J.-P. DUSSAULT, *A note on a globally convergent Newton method for solving monotone variational inequalities*, Oper. Res. Lett., 6 (1987), pp. 35–42.
- [26] J.-S. PANG, *Inexact Newton methods for the nonlinear complementarity problem*, Math. Programming, 36 (1986), pp. 54–71.
- [27] J.-S. PANG, *A posteriori error bounds for the linearly-constrained variational inequality problem*, Math. Oper. Res., 12 (1987), pp. 474–484.
- [28] J.-S. PANG, *Complementarity problems*, in Handbook of Global Optimization, R. Horst and P. Pardalos, eds., Kluwer Academic Publishers, Boston, MA, 1995, pp. 271–338.
- [29] J.-S. PANG, *Error bounds in mathematical programming*, Math. Programming, 79 (1997), pp. 299–332.
- [30] J.-S. PANG AND D. CHAN, *Iterative methods for variational and complementarity problems*, Math. Programming, 24 (1982), pp. 284–313.
- [31] J.-M. PENG AND M. FUKUSHIMA, *A hybrid Newton method for solving the variational inequality problem via the D-gap function*, Math. Programming, 86 (1999), pp. 367–386.
- [32] J.-M. PENG, C. KANZOW, AND M. FUKUSHIMA, *A hybrid Josephy-Newton method for solving box constrained variational inequality problem via the D-gap function*, Optim. Methods Softw., 10 (1999), pp. 687–710.
- [33] B. T. POLYAK, *Introduction to Optimization*, Optimization Software, Inc., Publications Division, New York, 1987.
- [34] D. RALPH, *Global convergence of damped Newton's method for nonsmooth equations via the path search*, Math. Oper. Res., 19 (1994), pp. 352–389.
- [35] S. M. ROBINSON, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43–62.
- [36] S. M. ROBINSON, *Normal maps induced by linear transformations*, Math. Oper. Res., 17 (1992), pp. 691–714.
- [37] R. T. ROCKAFELLAR, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.
- [38] M. SOLODOV AND B. SVAITER, *A comparison of rates of convergence of two inexact proximal point algorithms*, in Nonlinear Optimization and Related Topics, G. D. Pillo and F. Gian-

- nessi, eds., Kluwer Academic Publishers, Norwell, MA, to appear.
- [39] M. SOLODOV AND B. SVAITER, *Forcing strong convergence of proximal point iterations in a Hilbert space*, Math. Programming, to appear.
- [40] M. SOLODOV AND B. SVAITER, *A globally convergent inexact Newton method for systems of monotone equations*, in Reformulation—Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Norwell, MA, 1998, pp. 355–369.
- [41] M. SOLODOV AND B. SVAITER, *A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator*, Set-Valued Anal., 7 (1999), pp. 323–345.
- [42] M. SOLODOV AND B. SVAITER, *Error bounds for proximal point subproblems and associated inexact proximal point algorithms*, Math. Programming, submitted.
- [43] M. SOLODOV AND B. SVAITER, *A hybrid projection–proximal point algorithm*, J. Convex Anal., 6 (1999), pp. 59–70.
- [44] M. V. SOLODOV, *Stationary points of bound constrained reformulations of complementarity problems*, J. Optim. Theory Appl., 94 (1997), pp. 449–467.
- [45] M. V. SOLODOV, *Implicit Lagrangian*, in Encyclopedia of Optimization, C. Floudas and P. Pardalos, eds., Kluwer Academic Publishers, Norwell, MA, 2000.
- [46] M. V. SOLODOV, *Some optimization reformulations of the extended linear complementarity problem*, Comput. Optim. Appl., 12 (1999), pp. 187–200.
- [47] M. V. SOLODOV AND B. F. SVAITER, *A new projection method for variational inequality problems*, SIAM J. Control Optim., 37 (1999), pp. 765–776.
- [48] D. SUN, *A regularization Newton method for solving nonlinear complementarity problems*, Appl. Math. Optim., 40 (1999), pp. 315–339.
- [49] D. SUN, M. FUKUSHIMA, AND L. QI, *A computable generalized Hessian of the D-gap function and Newton-type methods for variational inequality problems*, in Complementarity and Variational Problems: State of the Art, M. Ferris and J.-S. Pang, eds., SIAM, Philadelphia, PA, 1997, pp. 452–473.
- [50] K. TAJI AND M. FUKUSHIMA, *Optimization based globally convergent methods for the nonlinear complementarity problems*, J. Oper. Res. Soc. Japan, 37 (1994), pp. 310–331.
- [51] K. TAJI, M. FUKUSHIMA, AND T. IBARAKI, *A globally convergent Newton method for solving strongly monotone variational inequalities*, Math. Programming, 58 (1993), pp. 369–383.
- [52] J. WU, M. FLORIAN, AND P. MARCOTTE, *A general descent framework for the monotone variational inequality problem*, Math. Programming, 61 (1993), pp. 281–300.
- [53] N. YAMASHITA AND M. FUKUSHIMA, *Modified Newton methods for solving a semismooth reformulation of monotone complementarity problems*, Math. Programming, 76 (1997), pp. 469–491.