CENTER FOR PARALLEL OPTIMIZATION

GLOBALLY CONVERGENT METHODS FOR NONLINEAR EQUATIONS

by
M. Ferris
S. Lucidi

Computer Sciences Technical Report #1030

July 1991

Globally Convergent Methods for Nonlinear Equations*

M. Ferris† S. Lucidi‡

July 1991

Abstract. We are concerned with enlarging the domain of convergence for solution methods of nonlinear equations. To this end, we produce a general framework in which to prove global convergence. Our framework relies on several notions: the use of a merit function, a generalization of a forcing function and conditions on the choice of direction. We also incorporate the idea of a nonmonotone stabilization procedure as a means of producing very good practical rates of convergence. The general theory is specialized to yield several well known results from the literature and is also used to generate three new algorithms for the solution of nonlinear equations. Numerical results for these algorithms applied to the nonlinear equations arising from nonlinear complementarity problems are given.

1 Introduction

In this paper we shall be concerned with proving global convergence of an algorithm to solve nonlinear equations. The algorithm we propose unifies many of the results given in the literature and also allows us to give new criteria under which a method can be expected to be globally convergent.

There are many known results on the global convergence of Newton type methods for nonlinear equations. Most of these methods are based on the use of a merit function. Several different types of merit function have been proposed in the literature (see for example [Bur80, SB80, Pol76, HPR89, HX90]). One of the main differences between many of the approaches is whether or not the underlying equations are smooth or nonsmooth. In the smooth case, global convergence is proven in many instances: we note in particular the results of Stoer[SB80]. However, many of the approaches for solving the nonlinear complementarity problem consider reformulations of the problem as a system of nonsmooth equations. Global convergence can also be established here, for example, some pertinent results are given in [HPR89, Ral91].

*This material is partially based on research supported by the Air Force Office of Scientific Research Grant AFOSR-89-0410, the National Science Foundation Grant CCR 9157632 and Istituto di Analisi dei Sistemi ed Informatica del CNR

†Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706

‡Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30, 00185 Roma, Italy

The emphasis of this paper will be on establishing many of the known results under very general conditions on a merit function via the use of a auxiliary function (a generalization of the familiar notion of a forcing function). While most of the paper deals with general conditions which a search direction must satisfy, the motivation for these conditions comes from considering Newton and Gauss–Newton methods for systems of n equations in n unknowns.

Several new ideas are encompassed in our framework. For instance, a nonmonotone stabilization procedure is used to overcome some of the difficulties associated with ill–conditioning and to enable a steps of length one to be taken much more frequently in Newton–type methods. Computationally, this has proven very effective, see for example, [GLL86] where the poor performance on the Rosenbrock problem has been overcome by nonmonotone linesearch (also [BS91]).

For different problems, several types of merit function have been proposed in the literature. By far the most popular is to use the Euclidean norm of the residual as the merit function. Other forms which can be considered are p–norm merit functions and ∞–norm merit functions. We will show that all of these forms of merit function can be treated in our framework.

While there is a wide literature on Gauss–Newton methods for the nonlinear least squares problem, only limited attention has been given to modifications of the search direction. In this paper, we propose several modifications of the search direction when the Newton direction cannot be found. Most of these find their basis in solving the equations in a least squares sense; see the survey paper of [Fra88] for several search direction modifications in least squares problems.

Much of this work was motivated by a desire to solve nonlinear complementarity problems. We are interested in using the smooth equations determined by Mangasarian[Man76] which are equivalent to the nonlinear complementarity problem. The essential difference is that while in least squares problems, the optimal value will not be zero, in this case it is, and so various new methods have been proposed. In particular, the method of Subramanian[Sub85] gives a direction which is consistent. The results in this paper show that our method can be used to establish global convergence in the case where the Jacobian is singular at the solution point (Dennis and Schnabel [DS83] give other results using the trust region approach).

The paper is organized as follows. The main theoretical results of the paper are given in Section 2. We describe the notion of a nonmonotone stabilization algorithm and give general conditions which are required for such a technique to give global convergence. These conditions are formulated in terms of a merit function, an auxiliary function (which resembles a forcing function) and the directions determined by the algorithm in question. We prove a general convergence result under these assumptions, without specifying the particular merit function, the auxiliary function or the direction, but only the conditions they must satisfy. We give several applications of this theory in Section 3 and describe three new algorithms which satisfy our conditions. Section 4 outlines several instances of work in the literature which can be formulated as special cases of our framework. In Section 5 of the paper we present some numerical results when the proposed algorithms are applied to nonlinear complementarity problems. Several standard examples are solved, including some equilibrium problems and the Karush–Kuhn–Tucker conditions for nonlinear programming.

2 Stabilization strategies for nonlinear equations

In this section we define general stabilization schemes to enable the solution of

$$H(x) = 0 \tag{1}$$

where $H: \mathbb{R}^n \to \mathbb{R}^n$ is a given function.

We use locally Lipschitzian merit functions θ with the property that $\theta(x) \geq 0$ for all $x \in \mathbb{R}^n$ and $\theta(x) = 0$ if and only if $H(x) = 0$ and we apply techniques from unconstrained optimization to effect the minimization of this merit function.

The algorithm we consider has the form

$$x^{k+1} = x^k + \alpha_k d^k, k = 0, 1, \ldots$$

where $x^0 \in \mathbb{R}^n$ is a given starting point, $d^k \neq 0$ is the search direction and α_k is the stepsize. Our formulation also relies on an auxiliary function, $\sigma: \mathbb{R}^{n+n+1} \to \mathbb{R}$, which is a generalization of the familiar notion of a forcing function [OR70]. The relationship between these constructs will be described in the sequel.

In order to to obtain a method for the solution of (1) we define a general stabilization scheme that includes different strategies for enforcing global convergence without requiring a monotonic reduction of the objective function. The proposed algorithm is similar to the one proposed by [GLL90].

NonMonotone Stabilization Algorithm (NMS)

Data: x^o, $\Delta_o > 0$, $\beta \in (0, 1)$, $\gamma \in (0, 1)$ and $N \geq 1$.

Step 1: Set $k = 0$, $\ell = 0$, $\Delta = \Delta_o$. Compute $\theta(x^o)$ and set $W = \theta(x^o)$.

Step 2: If $k = \ell + N$ compute $\theta(x^k)$. If $\theta(x^k) = 0$ stop; otherwise:

(a) if $\theta(x^k) \geq W$, replace x^k by x^ℓ, set $k = \ell$ and go to Step 4;

(b) if $\theta(x^k) < W$, set $\ell = k$, update W;
if $\|d^k\| \leq \Delta$, set $\alpha_k = 1$, $\Delta = \beta\Delta$ and go to Step 5; otherwise go to Step 4.

Step 3: If $k \neq \ell + N$ compute d^k. If $\|d^k\| = 0$ stop; otherwise:

(a) if $\|d^k\| \leq \Delta$, set set $\alpha_k = 1$, $\Delta = \beta\Delta$ and go to Step 5;

(b) if $\|d^k\| > \Delta$, compute $\theta(x^k)$; then:
if $\theta(x^k) \geq W$, replace x^k by x^ℓ, set $k = \ell$ and continue; otherwise set $\ell = k$, update W and continue.

Step 4: Find the smallest integer from $i = 0, 1, \ldots$ such that

$$\theta(x^k + 2^{-i}d^k) \leq W - \gamma 2^{-i}\sigma_k(x^k, d^k), \tag{2}$$

and set $\alpha_k = 2^{-i}$, $\ell = k+1$, update W and continue.

Step 5: Set $x^{k+1} = x^k + \alpha_k d^k$, $k = k+1$, and go to Step 2.

$\triangle$

In the description of the algorithm, ℓ denotes the index of the last accepted point where the objective function has been evaluated. For later reference we introduce a new index j which is set initially at $j = 0$ and incremented each time we define $\ell = k$. Then we indicate by $\{x^{\ell(j)}\}$ the sequence of points where the objective function is evaluated and by $\{W_j\}$ the sequence of reference values. Furthermore, we also need the index $q(k)$ defined by:

$$q(k) := \max\,[j : \ell(j) \leq k], \tag{3}$$

thus $\ell(q(k))$ is the largest iteration index not exceeding k where the merit function was evaluated.

In order to complete the description of the algorithm we must specify the the criterion employed for updating W_j. The reference value W_j for the merit function is initially set to $\theta(x^o)$. Whenever a point $x^{\ell(j)}$ is generated such that $\theta(x^{\ell(j)}) < W_j$, the reference value is updated by taking into account a prefixed number $m(j) \leq M$ of previous function values of the merit function. To be precise, we require that the updating rule for W_{j+1} satisfies the following condition.

Given $M \geq 0$, let $m(j+1)$ be such that

$$m(j+1) \leq \min\,[m(j)+1, M],$$

and let

$$F_{j+1} = \max_{0 \leq i \leq m(j+1)} \theta(x^{\ell(j+1-i)}), \tag{4}$$

choose the value W_{j+1} to satisfy

$$\theta(x^{\ell(j+1)}) \leq W_{j+1} \leq F_{j+1}. \tag{5}$$

These conditions on the reference values include several ways of determining the sequence $\{W_j\}$ in an implementation of the algorithm. For example, any of the following updating rules can be used:

$$W_{j+1} = F_{j+1} = \max_{0 \leq i \leq m(j+1)} \theta(x^{\ell(j+1-i)}), \tag{6}$$

$$W_{j+1} = \max\left[\theta(x^{\ell(j+1)}), \frac{1}{m(j+1)+1} \sum_{i=0}^{m(j+1)} \theta(x^{\ell(j+1-i)})\right], \tag{7}$$

$$W_{j+1} = \min\left[F_{j+1}, \frac{1}{2}\left(W_j + \theta(x^{\ell(j+1)})\right)\right]. \tag{8}$$

We note that (6) is the easiest to satisfy, while (7) and (8) define conditions which guarantee "mean descent".

We now describe the conditions which will ensure the global convergence of the aforementioned method. We will make frequent use of the following compactness assumption on the level set of the merit function

C: $\Omega_0 := \{x \mid \theta(x) \leq \theta(x^0)\}$ is bounded

The auxiliary function, σ, the merit function, θ and the search direction must satisfy the following properties:

A1: $\sigma_k(x^k, d^k) \to 0$ implies $\theta(x^k) \to 0$

A2: $0 \geq -\sigma_k(x^k, d^k) \geq \theta^D(x^k; d^k)$

A3: $F_{q(k)}^{p_3} \left\| d^k \right\|^{p_2} \leq L_2 \sigma_k(x^k, d^k)$, $p_2 \geq 1$, $p_3 > 0$, $L_2 > 0$ and $\left\| d^k \right\| \leq L_3$

where $\theta^D(x; v)$ is the Dini upper directional derivative of θ at x in the direction v, defined as

$$\theta^D(x; v) = \limsup_{\lambda \downarrow 0} \frac{\theta(x + \lambda v) - \theta(x)}{\lambda}$$

and $q(k)$ is defined in (3) and $F_{q(k)}$ in (4).

It is easy to show that assuming (C) and (A2), the following assumption implies (A3).

A3': $\left\| d^k \right\|^{p_2} \leq L_2 \sigma_k(x^k, d^k)$, $p_2 \geq 2$, $L_2 > 0$

In order to prove convergence of our model algorithm, we must first prove that the stepsize rule can be satisfied. The ensuing lemma establishes the existence of a step satisfying (2).

Lemma 1 *Let θ be locally Lipschitzian and $\gamma \in (0,1)$ be arbitrary. Suppose that assumptions (A1), (A2) and (A3) hold. Then, at Step 4 of Algorithm NMS, there exists a scalar $\bar{\alpha} > 0$ such that for all $\alpha \in [0, \bar{\alpha}]$*

$$\theta(x^k + \alpha d^k) \leq W_{q(k)} - \gamma \alpha \sigma_k(x^k, d^k)$$

Proof By the stopping criteria of Step 2 and Step 3, and by using assumptions (A1) and (A3) we have that $\sigma_k(x^k, d^k) \neq 0$. Assumption (A2) implies that $\theta^D(x^k; d^k) < 0$ and thus $d^k \neq 0$. Assume therefore that $d^k \neq 0$ but that the conclusion of the lemma is false. Then there exists a sequence $\{\alpha_l\}$ converging to zero such that

$$\theta(x^k + \alpha_l d^k) > W_{q(k)} - \gamma \alpha_l \sigma_k(x^k, d^k)$$

Using the definition of $W_{q(k)}$ it can be seen that

$$\theta(x^k + \alpha_l d^k) - \theta(x^k) > -\gamma \alpha_l \sigma_k(x^k, d^k)$$

Dividing both sides by α_l and passing to the limit we see

$$\theta^D(x^k; d^k) \geq -\gamma \sigma_k(x^k, d^k)$$

Assumption (A2) gives

$$-\sigma_k(x^k, d^k) \geq -\gamma \sigma_k(x^k, d^k)$$

which implies that $\sigma_k(x^k, d^k) = 0$, which is a contradiction. □

We shall need the following technical lemma in order to prove the convergence of the model algorithm.

Lemma 2 *Suppose θ is locally Lipschitzian and θ, σ and $\{d^k\}$ satisfy Assumption (A2), then:*

(a) *if $\{x^k\}$ converges to $\bar{x}$ and Assumption (A3) holds then $\{\sigma_k(x^k, d^k)\}$ is bounded;*

(b) *if $\{x^k\}$ is bounded and $\lim_{k\to\infty}\|d^k\| = 0$ then $\lim_{k\to\infty} \sigma_k(x^k, d^k) = 0$.*

Proof (a) Since θ is locally Lipschitzian and $\{x^k\}$ converges, it follows that there exists a constant $L > 0$ such that for all k

$$\left|\theta^D(x^k; d^k)\right| \leq L\left\|d^k\right\|$$

By assumption (A2), we see

$$L\left\|d^k\right\| \geq \left|\theta^D(x^k; d^k)\right| \geq \sigma_k(x^k, d^k)$$

The boundedness of $\{\sigma_k(x^k, d^k)\}$ now follows from (A3).

(b) If the conclusion of part (b) is false, then we have that $\limsup_{k\to\infty} \sigma_k(x^k, d^k) = \bar{\sigma} > 0$. Since $\{x^k\}$ is bounded, we can find a subsequence $k \in K$ such that

$$\lim_{k\in K} \sigma_k(x^k, d^k) = \bar{\sigma} > 0, \tag{9}$$

$$\lim_{k\in K} x^k = \bar{x}, \tag{10}$$

$$\lim_{k\in K} \|d^k\| = 0. \tag{11}$$

Then repeating the reasoning of part (a) we obtain:

$$L\left\|d^k\right\| \geq \left|\theta^D(x^k; d^k)\right| \geq \sigma_k(x^k, d^k), \qquad k \in K \tag{12}$$

and, by using (11) and (12), we have:

$$\lim_{k\in K} \sigma_k(x^k, d^k) = 0, \tag{13}$$

which contradicts (9). □

The next lemma shows some properties of the sequence $\{x^k\}$ produced by Algorithm NMS.

Lemma 3 *Assume that Assumption (C) holds and that Algorithm NMS produces an infinite sequence $\{x^k\}$; then:*

(a) $\{x^k\}$ *remains in a compact set;*

(b) *the sequence* $\{F_j\}$ *is non increasing and has a limit* $\hat{F}$*;*

(c) *let* $s(j)$ *be an index in the set* $\{\ell(j), \ell(j-1), \ldots, \ell(j-m(j))\}$ *such that:*

$$\theta(x^{s(j)}) = F_j = \max_{0 \le i \le m(j)} \theta(x^{\ell(j-i)}); \tag{14}$$

then, for any integer k*, there exist indices* h_k *and* j_k *such that:*

$$0 < h_k - k \le N(M+1), \qquad h_k = s(j_k),$$
$$F_{j_k} = \theta(x^{h_k}) < F_{q(k)}.$$

Proof The proof of lemma follows, with minor modification from the proofs of Lemma 1 and Lemma 2 of [GLL90]. □

The following result is central to our development. We show that the merit function converges to a limit and also the product of the step size and the auxiliary function tends to zero. Note that both of these conclusions are trivial in the case of a monotone line search procedure.

Lemma 4 *Let* $\{x^k\}$ *be a sequence produced by the algorithm. Suppose that Assumptions (A1), (A2), (A3) and (C) hold and that* θ *is locally Lipschitzian. Then:* $\lim_{k\to\infty} \theta(x^k)$ *exists and* $\lim_{k\to\infty} \alpha_k \sigma_k(x^k, d^k) = 0$.

Proof Let $\{x^k\}_K$ denote the set (possibly empty) of points satisfying the test at Step 2 (b) or at Step 3 (a), so that:

$$\|d^k\| \le \Delta_o \beta^t, \qquad \alpha_k = 1 \qquad \text{for } \ k \in K \tag{15}$$

where the integer t increases with $k \in K$. It follows from (15) that, if K is an infinite set,we have $\|d^k\| \to 0$, for $k \to \infty$ and by (b) of Lemma 2:

$$\lim_{\substack{k\to\infty \\ k \in K}} \alpha_k \sigma_k(x^k, d^k) = 0. \tag{16}$$

Now let $s(j)$ and $q(k)$ be the indices defined by (14) and (3). We prove by induction that, for any $i \ge 1$, we have:

$$\lim_{j\to\infty} \alpha_{s(j)-i} \sigma_{s(j)-i}(x^{s(j)-i}, d^{s(j)-i}) = 0, \tag{17}$$

$$\lim_{j\to\infty} \theta(x^{s(j)-i}) = \lim_{j\to\infty} \theta(x^{s(j)}) = \lim_{j\to\infty} F_j = \hat{F}. \tag{18}$$

Assume first that $i = 1$. If $s(j) - 1 \in K$, (17) holds with $k = s(j) - 1$. Otherwise, if $s(j) - 1 \notin K$, recalling the acceptability criterion of the nonmonotone line search, we can write:

$$\begin{aligned} F_j = \theta(x^{s(j)}) &= \theta(x^{s(j)-1} + \alpha_{s(j)-1} d^{s(j)-1}) \\ &\leq F_{q(s(j)-1)} + \gamma \alpha_{s(j)-1} \sigma_{s(j)-1}(x^{s(j)-1}), d^{s(j)-1}). \end{aligned}$$

It follows that:

$$F_{q(s(j)-1)} - F_j \geq \gamma \alpha_{s(j)-1} \sigma_{s(j)-1}(x^{s(j)-1}), d^{s(j)-1}) \tag{19}$$

Therefore, if $s(j) - 1 \notin K$ for an infinite subsequence, from (b) of Lemma 4 and (19) we get

$$\lim_{j \to \infty} \alpha_{s(j)-1} \sigma_{s(j)-1}(x^{s(j)-1}), d^{s(j)-1}) \to 0, \tag{20}$$

so that (17) hold for $i = 1$.

It follows from (A3) and (20) that

$$\lim_{j \to \infty} \alpha_{s(j)-1} F^{p_3}_{q(s(j)-1)} \left\| d^{s(j)-1} \right\|^{p_2} = 0$$

and since $\{\alpha_k\}$, $\{F_k\}$, $\{d^k\}$ are bounded above and $p_2 \geq 1$, $p_3 > 0$ that

$$\lim_{j \to \infty} \alpha_{s(j)-1} F_{q(s(j)-1)} \left\| d^{s(j)-1} \right\| = 0$$

We consider two cases. Suppose first that $\limsup_{j \to \infty} \alpha_{s(j)-1} \left\| d^{s(j)-1} \right\| > 0$. Then, since $\lim_{j \to \infty} F_j$ exists, it follows that $\lim_{j \to \infty} F_j = 0$. However, by recalling that, by the definition of F_j and the description of the algorithm, we have that $\theta(x^{s(j)-1}) \leq F_{q(s(j)-1)}$, hence it is immediate that $\lim_{j \to \infty} \theta(x^{s(j)-1}) = \lim_{j \to \infty} F_j = 0$. Then (18) clearly holds for $i = 1$. Otherwise, $\limsup_{j \to \infty} \alpha_{s(j)-1} \left\| d^{s(j)-1} \right\| = 0$, which implies that $\lim_{j \to \infty} \alpha_{s(j)-1} \left\| d^{s(j)-1} \right\| = 0$. This in turn shows that $\left\| x^{s(j)} - x^{s(j)-1} \right\| \to 0$, so that (18) holds for $i = 1$ by the uniform continuity of θ on the the compact set containing $\{x^k\}$ (see (a) of Lemma 3).

Assume now that (17) and (18) hold for a given i and consider the point $x^{s(j)-(i+1)}$. Reasoning as before, we can again distinguish the case $s(j) - (i + 1) \in K$, when (16) holds with $k = s(j) - (i + 1)$, and the case $s(j) - (i + 1) \notin K$, in which we have:

$$\theta(x^{s(j)-i}) \leq F_{q(s(j)-(i+1))} + \gamma \alpha_{s(j)-(i+1)} \sigma_{su}(x^{s(j)-(i+1)}, d^{s(j)-(i+1)})$$

and hence:

$$F_{q(s(j)-(i+1))} - \theta(x^{s(j)-i}) \geq \gamma \alpha_{s(j)-(i+1)} \sigma_{s(j)-(i+1)}(x^{s(j)-(i+1)}, d^{s(j)-(i+1)}). \tag{21}$$

Then, using (16), (18), (21) and we can assert that equation (17) holds with i replaced by $i + 1$.

Invoking (A3) and using a similar argument to that above we see that

$$\lim_{j\to\infty} \alpha_{s(j)-(i+1)} F_{q(s(j)-(i+1))} \left\| d^{s(j)-(i+1)} \right\| = 0$$

Again, we must consider two cases. Suppose first that $\limsup_{j\to\infty} \alpha_{s(j)-(i+1)} \left\| d^{s(j)-(i+1)} \right\| > 0$. Then since $\lim_{j\to\infty} F_j$ exists, it follows that $\lim_{j\to\infty} F_j = 0$, and using, again, that $\theta(x^{s(j)-(i+1)}) \leq F_{q(s(j)-(i+1)}$ we have $\lim_{j\to\infty} \theta(x^{s(j)-(i+1)}) = \lim_{j\to\infty} F_j = 0$. Thus, in this case, (18) holds for $j+1$. In the other case, $\limsup_{j\to\infty} \alpha_{s(j)-(i+1)} \left\| d^{s(j)-(i+1))} \right\| = 0$, which implies that $\lim_{j\to\infty} \alpha_{s(j)-(i+1)} \left\| d^{s(j)-(i+1)} \right\| = 0$. This implies, moreover, that $\left\| x^{s(j)-i} - x^{s(j)-(i+1)} \right\| \to 0$, so that by (18) and the uniform continuity of θ on the compact set containing $\{x^k\}$:

$$\lim_{j\to\infty} \theta(x^{s(j)-(i+1)}) = \lim_{j\to\infty} \theta(x^{s(j)-i}) = \lim_{j\to\infty} F_j,$$

so that (18) is satisfied with i replaced by $i+1$, which completes the induction.

Now let x^k be any given point produced by the algorithm. Then by (c) of Lemma 3 there is a point $x^{h_k} \in \{x^{s(j)}\}$ such that

$$0 < h_k - k \leq (M+1)N. \tag{22}$$

Then, we can write:

$$x^k = x^{h_k} - \sum_{i=1}^{h_k-k} \alpha_{h_k-i} d^{h_k-i}$$

and this implies, by (17) and (22), that:

$$\lim_{k\to\infty} \|x^k - x^{h_k}\| = 0. \tag{23}$$

From the uniform continuity of θ, it follows that

$$\lim_{k\to\infty} \theta(x^k) = \lim_{k\to\infty} \theta(x^{h_k}) = \lim_{j\to\infty} F_j, \tag{24}$$

so that we have proved that $\lim_{k\to\infty} \theta(x^k)$ exists.

If $k \notin K$, we obtain $\theta(x^{k+1}) \leq F_{q(k)} - \gamma\alpha_k\sigma_k(x^k, d^k)$ and hence we have that:

$$F_{q(k)} - \theta(x^{k+1}) \geq \gamma\alpha_k\sigma_k(x^k, d^k). \tag{25}$$

Therefore by (16), (24), (25), we can conclude that:

$$\lim_{k\to\infty} \alpha_k\sigma_k(x^k, d^k) = 0.$$

□

We shall need the following assumption to complete our convergence proof.

A4 for every sequence $\{x^k\}$ converging to $\bar{x}$, every convergent sequence $\{d^k\}$ and every sequence $\{\lambda_k\}$ of positive scalars converging to zero

$$\lim_{k\to\infty} -\sigma_k(x^k, d^k) \geq \limsup_{k\to\infty} \frac{\theta(x^k + \lambda_k d^k) - \theta(x^k)}{\lambda_k}$$

whenever the limit in the left hand side exists.

This assumption is a strengthening of (A2). In fact, we note that if θ is subdifferentially regular [Cla83] then both (A2) and (A4) are equivalent to

$$0 \geq -\sigma_k(x^k, d^k) \geq \theta'(x^k; d^k)$$

We are now able to prove our convergence result.

Theorem 5 *Let θ be a locally Lipschizian merit function and suppose that (A1), (A2), (A3) and (C) hold. Then*

1. *If $\limsup_{k\to\infty} \alpha_k > 0$, then $\lim_{k\to\infty} \theta(x^k) = 0$.*
2. *If $\limsup_{k\to\infty} \alpha_k = 0$ and if $\bar{x}$ is an accumulation point of $\{x^k\}$ where (A4) holds, then $\theta(\bar{x}) = 0$.*

Proof Suppose that $\limsup_{k\to\infty} \alpha_k = \bar{\alpha} > 0$. Since $\{x^k\}$ is bounded can find a subsequence $k \in K$ such that $\lim_{k\in K} \alpha_k = \bar{\alpha}$ and $\lim_{k\in K} x^k = \bar{x}$. By Lemma 2, it follows that $\left\{\sigma_k(x^k, d^k) \mid k \in K\right\}$ is bounded. By taking further subsequences if necessary we may assume that $\lim_{k\in K} \alpha_k = \bar{\alpha}$ and $\lim_{k\in K} \sigma_k(x^k, d^k)$ exists. However, from Lemma 4 we have $\lim_{k\in K} \alpha_k \sigma_k(x^k, d^k) = 0$. Since $\bar{\alpha} > 0$, it follows that $\lim_{k\in K} \sigma_k(x^k, d^k) = 0$. The assumption (A1) gives $\lim_{k\in K} \theta(x^k) = 0$. Hence $\lim_{k\to\infty} \theta(x^k) = 0$ since from Lemma 4 the sequence $\{\theta(x^k)\}$ converges.

Otherwise, $\limsup_{k\to\infty} \alpha_k = 0$ implying that $\lim_{k\to\infty} \alpha_k = 0$. Let $\bar{x}$ be an accumulation point of $\{x^k\}$ where (A4) holds and let $\left\{x^k \mid k \in K\right\}$ converge to $\bar{x}$. Using Lemma 2, we may assume that $\left\{x^k \mid k \in K\right\}$, $\left\{d^k \mid k \in K\right\}$, $\{\alpha_k \mid k \in K\}$ and $\left\{\sigma_k(x^k, d^k) \mid k \in K\right\}$ converge for some subsequence $k \in K$. Now, for sufficiently large values of k and $k \in K$, we have that $\alpha_k < 1$ and, hence, that α_k, $k \in K$, is eventually produced at Step 4. Then by the properties of the linesearch (2)

$$\theta(x^k + (\alpha_k/\nu)d^k) - W_{q(k)} > -\gamma(\alpha_k/\nu)\sigma_k(x^k, d^k)$$

and by the definition of $W_{q(k)}$ we have

$$\theta(x^k + (\alpha_k/\nu)d^k) - \theta(x^k) > -\gamma(\alpha_k/\nu)\sigma_k(x^k, d^k)$$

Using assumption (A4) we have

$$\begin{aligned} -\lim_{k\in K}\sigma_k(x^k,d^k) &\geq \limsup_{k\in K}\left[\frac{\theta(x^k+(\alpha_k/\nu)d^k)-\theta(x^k)}{\alpha_k/\nu}\right] \\ &\geq -\gamma\lim_{k\in K}\sigma_k(x^k,d^k) \end{aligned}$$

which implies that $\lim_{k\in K}\sigma_k(x^k,d^k)=0$. It is now easy to show using (A1) that $\theta(\bar{x})=0$. □

3 Computation of Search Direction

In this section we describe several methods for the computation of a search direction satisfying the assumptions of our model algorithm.

First of all, we specialize to the following situation. We assume the merit function θ is given by

$$\theta(x):=\frac{1}{2}\|H(x)\|^2$$

and that H is continuously differentiable. Under this assumption, we note that (A4) is equivalent to (A2).

We shall make the following assumption in order to guarantee that stationary points of this merit function solve the equations:

$$\nabla\theta(x)=0 \text{ implies } \theta(x)=0 \tag{26}$$

There are several possibilities for the direction choice. Essentially, we have a system of smooth equations and therefore it is possible to consider standard search directions for systems of equations. A first choice is to let $d^k=-\nabla\theta(x^k)$. If we define $\sigma_k(x^k,d^k):=-\nabla\theta(x^k)d^k$ then assuming (26), the assumptions (A1), (A2), (A3) and (A4) hold. In practice, it is well known that this gives poor convergence, so a more reasonable choice is the Newton direction for equations, that is, d^k solves

$$H(x^k)+\nabla H(x^k)d=0$$

If we assume (26) and the Jacobian of H is nonsingular, then by defining $\sigma_k(x^k,d^k):=-\nabla\theta(x^k)d^k=2\theta(x^k)$ we see that the assumptions (A1),(A2),(A3) and (A4) from Section 2 are satisfied if (C) holds.

In the remainder of this section we propose several techniques for dealing with the case when $\nabla H(x)$ is singular or badly conditioned.

We consider three techniques for modifying the Newton direction. The first possibility is to consider a modified Gauss–Newton direction, along the lines of the work of [Sub85].

Algorithm 1 *Given x^k, calculate the search direction as a solution of the following system*

$$(\nabla H(x^k)^T\nabla H(x^k)+\lambda_k I)d=-\nabla H(x^k)^T H(x^k) \tag{27}$$

Determine x^{k+1} by using Algorithm NMS.

In an implementation of this algorithm we need to specify λ_k. In order that the results from Section 2 might be applied we propose to define $\lambda_k := c\max_{0\le i\le M(N+1)} \left\|H(x^{k-i})\right\|$ (where N and M are the constants used in Algorithm NMS). Note that the perturbation to the Gauss–Newton direction gets smaller as we approach the solution.

Proposition 6 *If we set $\sigma_k(x^k,d^k) := \left\|\nabla H(x^k)d^k\right\|^2 + \lambda_k \left\|d^k\right\|^2$ then assuming (C) and (26), any accumulation point $\bar{x}$ of the sequence of points $\{x^k\}$ produced by Algorithm 1 is a solution of (1).*

Moreover, if there exists an accumulation point $\bar{x}$ of $\{x^k\}$ where the Jacobian matrix $\nabla H(\bar{x})$ is nonsingular then the sequence $\{x^k\}$ converges superlinearly to $\bar{x}$, the stepsize $\alpha_k = 1$ is accepted for sufficiently large k and condition $\|d^k\| \le \Delta_o \beta^k$ holds eventually for any $\Delta_o > 0$ and $\beta > 0$.

Proof We show that assumptions (A1), (A2), (A3) and (A4) are satisfied. First of all, we note that $\sigma_k(x^k,d^k) = -\nabla\theta(x^k)d^k$ and so assumptions (A2) and (A4) are satisfied. Furthermore, if $\sigma_k(x^k,d^k) \to 0$, then either $\lambda_k \to 0$ or $\|d\| \to 0$. In the first case, $\theta(x^k) \to 0$ is clear, and in the second, the same conclusion follows from the definition of d^k and (26). In order to obtain the first part of (A3) we note that, by definition of $F_{q(k)}$, we have:

$$F_{q(k)} \le \max_{0\le i\le M(N+1)} \theta(x^{k-i}),$$

and then we can take $p_2 = 2$ and $p_3 = 1/2$. For the second part of (A3) we have

$$\begin{aligned} \lambda_k \left\|d^k\right\|^2 &\le \sigma_k(x^k,d^k) \le \left|\nabla\theta(x^k)d^k\right| \\ &\le \left\|\nabla\theta(x^k)\right\| \left\|d^k\right\| \le \left\|H(x^k)\right\| \left\|\nabla H(x^k)\right\| \left\|d^k\right\| \end{aligned}$$

and, hence, the boundedness of $\{d^k\}$ follows from (C) and the fact that $\lambda_k \ge c\left\|H(x^k)\right\|$.

As regards the superlinear convergence rate, first we observe that the Hessian matrix $\nabla^2\theta(\bar{x}) = \nabla H(\bar{x})^T\nabla H(\bar{x})$ is positive definite. Since $\bar{x}$ is an accumulation point of $\{x^k\}$, there exists an index $\bar{k}$ such that the point $x^{\bar{k}}$ is sufficiently close to $\bar{x}$ so as to have:

$$\gamma_{\min}\left(\nabla H(x^{\bar{k}})^T\nabla H(x^{\bar{k}}) + \lambda_{\bar{k}} I\right) \ge \frac{1}{2}\gamma_{\min}\left(\nabla^2\theta(\bar{x})\right) > 0, \tag{28}$$

where we have indicated by $\gamma_{\min}(B)$ the minimum eigenvalue of a matrix B. By (27) and (28) we have

$$\|d^{\bar{k}}\| \le \frac{2}{\gamma_{\min}(\nabla^2\theta(\bar{x}))}\|\nabla\theta(x^{\bar{k}})\|,$$

and by repeating, with minor modifications, the same reasonings of the proof of Proposition 1.12 in [Ber82] and by recalling that Theorem 5 implies that $\theta(x^k) \to 0$, we have that $\lim_{k\to\infty} x^k = \bar{x}$. Finally we can observe that the matrix $\nabla H(x^k)^T\nabla H(x^k) + \lambda_k I$, with $\lambda_k = c\max_{0\le i\le M(N+1)} \left\|H(x^{k-i}\right\|$, satisfies the Dennis–More condition which ensures both that the unit stepsize is eventually accepted by the linesearch technique and that the sequence $\{x^k\}$ converges superlinearly (see, for example, [Ber82, Proposition 1.15]). Finally, the last statement of the proposition follows from the superlinear convergence rate of $\{x^k\}$. $\square$

In the second case we take as our direction a solution of a modified system of equations, where the Jacobian of H has been modified by a diagonal matrix. Unfortunately, the new direction might not even be a descent direction for our merit function, so in certain cases we revert to taking the direction of steepest descent for the merit function. A full description of the method is given below.

Algorithm 2 *Let $\lambda > 0$, $L_4 > 0$, $L_5 > 0$ and $L_6 > 0$. Given x^k, if $H(x^k) = 0$ stop, else compute the smallest nonnegative integer m such that the matrix $\nabla H(x^k) + m\lambda I$ has condition number less than $L_6/\|H(x^k)\|$.*

Calculate the search direction, d^k, as a solution of the following system

$$(\nabla H(x^k) + m\lambda I)d = -H(x^k)$$

Evaluate $\sigma_k(x^k, d^k) := -\nabla\theta(x^k)d^k$. If $\sigma_k(x^k, d^k) < 0$ let $d^k = -d^k$.

If $\sigma_k(x^k, d^k) < L_4 \left\|\nabla\theta(x^k)\right\|^3$ or $\sigma_k(x^k, d^k) < L_5 \left\|d^k\right\|^3$ then let $d^k = -\nabla\theta(x^k)$. Calculate x^{k+1} by using Algorithm NMS.

Proposition 7 *Assuming (C) and (26), any accumulation point $\bar{x}$ of the sequence of points $\{x^k\}$ produced by Algorithm 2 is a solution of (1).*

Moreover, if there exists a accumulation point $\bar{x}$ of $\{x^k\}$ where the Jacobian matrix $\nabla H(\bar{x})$ is nonsingular then the sequence $\{x^k\}$ converges superlinearly to $\bar{x}$, the stepsize $\alpha_k = 1$ is accepted for sufficiently large k and condition $\|d^k\| \leq \Delta_o\beta^k$ holds eventually for any $\Delta_o > 0$ and $\beta > 0$.

Proof The global convergence property of the algorithm follows easily by noting that the conditions of the algorithm are determined precisely to ensure the satisfaction of assumptions A1, A2 and A3'.

By using the assumption that the matrix $\nabla H(\bar{x})$ is nonsingular and recalling that the first part of the proposition ensures that $\nabla\theta(\bar{x}) = 0$, we can find a neighborhood Ω of $\bar{x}$ such that for any $x \in \Omega$ and $x \neq \bar{x}$ we have:

$$\kappa(\nabla H(x)) \leq \frac{L_6}{\|H(x)\|}, \tag{29}$$

$$\gamma_{\min}\left((\nabla H(x)^T \nabla H(x)\right) \geq \frac{1}{2}\gamma_{\min}\left((\nabla H(\bar{x})^T \nabla H(\bar{x})\right) > 0, \tag{30}$$

$$\|\nabla\theta(x)\| \leq \frac{1}{\gamma_{\max}(\nabla H(x)^T \nabla H(x))} \min\left[\frac{1}{L_4}, \frac{\gamma_{\min}\left((\nabla H(x)^T \nabla H(x)\right)^3}{L_5}\right], \tag{31}$$

where we have indicated by $\kappa(B)$, $\gamma_{\max}(B)$ and $\gamma_{\min}(B)$ the condition number, the maximum eigenvalue and the minimum eigenvalue of a matrix B.

Since $\bar{x}$ is an accumulation point of the sequence $\{x^k\}$ there an index $\bar{k}$ such that $x^{\bar{k}} \in \Omega$. Therefore by (29) we have that the direction $d^{\bar{k}}$ is computed by solving the system:

$$\nabla H(x^{\bar{k}})d^{\bar{k}} = -H(x^{\bar{k}}),$$

or, equivalently, by solving the following system:

$$\nabla H(x^{\bar{k}})^T \nabla H(x^{\bar{k}}) d^{\bar{k}} = -\nabla\theta(x^{\bar{k}}). \tag{32}$$

This direction $d^{\bar{k}}$ satisfies all the tests of Algorithm 2. In fact we have:

$$\begin{aligned}
\sigma_{\bar{k}}(x^{\bar{k}}, d^{\bar{k}}) &\geq \nabla\theta(x^{\bar{k}})^T \left(\nabla H(x^{\bar{k}})^T \nabla H(x^{\bar{k}})\right)^{-1} \nabla\theta(x^{\bar{k}}) \\
&\geq \frac{\|\nabla\theta(x^{\bar{k}})\|^2}{\gamma_{\max}\left(\nabla H(x^{\bar{k}})^T \nabla H(x^{\bar{k}})\right)} > 0.
\end{aligned}$$

Then by the preceding relation, (31) and (32) we obtain:

$$\sigma_{\bar{k}}(x^{\bar{k}}, d^{\bar{k}}) \geq L_4 \|\nabla\theta(x^{\bar{k}})\|^3,$$

$$\sigma_{\bar{k}}(x^{\bar{k}}, d^{\bar{k}}) \geq \frac{L_5 \|\nabla\theta(x^{\bar{k}})\|^3}{\gamma_{\min}\left(\nabla H(x^{\bar{k}})^T \nabla H(x^{\bar{k}})\right)^3} \geq L_5 \|d^{\bar{k}}\|^3.$$

Now the superlinear convergence property follows by repeating the same steps of the second part of the proof of Proposition 6 after having noted that by (32) and (30) we have that

$$\|d^{\bar{k}}\| \leq \frac{2}{\gamma_{\min}\left(\nabla H(\bar{x})^T \nabla H(\bar{x})\right)} \|\nabla\theta(x^{\bar{k}})\|.$$

□

In practice, the constants in the algorithm have to be chosen appropriately. Furthermore, the choice of λ can be estimated from a calculation of the condition number of the matrix $\nabla H(x) + \lambda I$ at the previous iteration. Normally this information is available from a factorization routine.

In the third case we try to modify the above technique in such a way as to guarantee at least that for appropriate choice of λ we obtain a descent direction for the merit function.

Algorithm 3 *Let $\lambda > 0$, $L_4 \in (0,1)$ and $L_5 \in (0,1)$. Given x^k, if $H(x^k) = 0$ stop, else compute the smallest nonnegative integer m such that the following system*

$$\left(\nabla H(x^k) + m\lambda I\right) d = -\left(m\lambda \nabla H(x^k)^T + I\right) H(x^k) \tag{33}$$

admits a solution d^k which satisfies the conditions:

$$\sigma_k(x^k, d^k) \geq L_4 \min\left[\|\nabla\theta(x^k)\|^3, \|\nabla\theta(x^k)\|^2\right] \tag{34}$$

$$\sigma_k(x^k, d^k) \geq L_5 \min\left[\|d^k\|^3, \|d^k\|^2\right] \tag{35}$$

where $\sigma_k(x^k, d^k) := -\nabla\theta(x^k) d^k$.

Evaluate x^{k+1} by using Algorithm NMS.

Proposition 8 *Assuming (C) and (26), any accumulation point $\bar{x}$ of the sequence of points $\{x^k\}$ produced by Algorithm 3 is a solution of (1).*

Moreover, if there exists a accumulation point $\bar{x}$ of $\{x^k\}$ where the Jacobian matrix $\nabla H(\bar{x})$ is nonsingular then the sequence $\{x^k\}$ converges superlinearly to $\bar{x}$, the stepsize $\alpha_k = 1$ is accepted for sufficiently large k and condition $\|d^k\| \leq \Delta_o \beta^k$ holds eventually for any $\Delta_o > 0$ and $\beta > 0$.

Proof First we must show that condition (34) and (35) are satisfied for sufficiently large λ to show the algorithm is well defined. Let $C_1 := \max_{x\in\Omega_0} \|H(x)\|$ and $C_2 := \max_{x\in\Omega_0} \|\nabla\theta(x)\|$. By (33) we have:

$$\left(1 - \frac{\|\nabla H(x^k)\|}{m\lambda}\right)\|d\| \leq \|H(x^k)\| + \frac{\|\nabla\theta(x^k)\|}{m\lambda}.$$

Therefore for $m \geq m_1$ with $m_1 := \max\left[2C_1, 1\right]/\lambda$ we obtain:

$$\|d^k\| \leq 2C_1 + 2C_2 := C_3. \tag{36}$$

Now, again from (33) we obtain:

$$d^k = -\nabla\theta(x^k) - \frac{1}{m\lambda}\left(\nabla H(x^k)d^k + H(x^k)\right) \tag{37}$$

If we premultiply (37) by $-\nabla\theta(x^k)^T$ and we take into account (36)then we have:

$$\sigma_k(x^k, d^k) \geq \|\nabla\theta(x^k)\|^2 - \frac{1}{m\lambda}\|\nabla\theta(x^k)\|C_4, \tag{38}$$

where $C_4 := C_3 \max_{x\in\Omega_0} \|\nabla H(x^k)\| + C_1$. By introducing the scalar

$$m_2 := \max\left[m_1, C_4/((1-L_4)\lambda\|\nabla\theta(x^k)\|)\right]$$

(condition (26) yields $\|\nabla\theta(x^k)\| \neq 0$) we can see that for all $m \geq m_2$ (38) implies (34).

Now, if we premultiply (37) by $(d^k)^T$ and we use again (36), we have:

$$\|d^k\|^2 \leq \sigma_k(x^k, d^k) + \frac{1}{m\lambda}C_4C_3 \tag{39}$$

Recalling that conditions (26) and (34) yield that $\sigma_k(x^k, d^k) \neq 0$ we can define

$$m_3 := \max\left[m_2, C_3C_4L_5/((1-L_5)\lambda\sigma_k(x^k, d^k))\right]$$

and we can observe that, for all $m \geq m_3$, (39) implies (35).

Now the global and superlinear convergence properties of Algorithm 3 follow directly by repeating, with minor modifications, the same arguments of proof of Proposition 7 $\square$

4 Examples of the Method

In this section we will give examples of the application of the method to several problem instances. In fact, some well known results from the literature can be cast in our framework.

In [HPR89], the following method is described. Let $\theta(x) = \frac{1}{2}\|H(x)\|^2$ and choose the search direction to satisfy

$$H(x^k) + G(x^k, d^k) = 0$$

at each iteration. Here, $G(x^k, d)$ is an appropriate approximation of the directional derivative of H in the direction d at x^k. The assumptions made in [HPR89] are essentially equivalent to the ones we make in Section 2. This can be seen by defining $\sigma_k(x^k, d^k) := 2\theta(x^k)$.

In the same paper, a Gauss–Newton method is also proposed. The same merit function is used. In this case, the direction is calculated by solving

$$\min_{d \in \mathbb{R}^n} H(x^k)^T G(x^k, d) + \frac{1}{2}\psi_k(d)$$

The assumptions made to prove convergence are essentially equivalent to (A1), (A2), (A3') and (C). Particular instances of functions ψ_k which are considered are

- $\psi_k(d) = d^T B_k d$ where B_k is a symmetric positive definite $n \times n$ matrix
- $\psi_k(d) = \left\|G(x^k, d)\right\|^2 + \epsilon_k \|d\|^2$ where ϵ_k is a nonnegative scalar.

In [HPR89] the conditions on ψ_k require in the first case that the sequence $\{B_k\}$ should have eigenvalues which are bounded away from zero and in the second case that $\{\epsilon_k\}$ should be bounded away from zero. In the model we propose, we can relax these conditions by essentially using the following forms

- $\psi_k(d) = d^T B_k d + c_1 F_k \|d\|^2$ where B_k is a symmetric positive semidefinite $n \times n$ matrix
- $\psi_k(d) = \left\|G(x^k, d)\right\|^2 + c_1 F_k \|d\|^2$

The motivation for this type of approach comes from the work of [Sub85] and can be thought of a a generalization of Algorithm 1 from the previous section.

The work of Burdakov [Bur80] can also be seen to be a special case of the method given in Section 2. In this case, the merit function is given by $\theta(x) := \|H(x)\|^r$ for some $r > 0$.

An interesting extension to the work of Pshenichny and Danilin [PD78] on minimax problems can be seen by using our formulation. In this case the merit function is given by

$$\theta(x) = \max_{1 \le i \le n} |H_i(x)|$$

where H_i are assumed to be continuously differentiable functions whose gradients satisfy a Lipschitz condition. By defining the set

$$\mathcal{I}_\delta := \{i \mid |H_i(x)| \ge \theta(x) - \delta\}$$

the linearization method of Pshenichny and Danilin has been shown to converge under the following assumptions:

1. $\exists \delta > 0$ such that for all x with $\theta(x) > 0$, $\theta(x) \le \theta(x^0)$ the linearized system

$$\nabla H_i(x^k)d + H_i(x^k) = 0, i \in \mathcal{I}_\delta \tag{40}$$

is solvable.

2. Let $d(x)$ denote the minimum norm solution of (40). Then, $\exists c > 0$ such that for all x with $\theta(x) > 0$ we have

$$\|d(x)\| \le c\theta(x)$$

3. $\{x \mid \theta(x) \le \theta(x^0)\}$ is bounded.

It can be shown that these assumptions imply the assumptions of the previous section. Let $\sigma_k(x^k, d^k) = \epsilon\theta(x^k)$, for some $\epsilon \in (0,1)$. Assumption (A1) is then immediate. Assumption (A2) follows from [PD78, Theorem 6.1] since it is shown that there exists an $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}_k)$

$$\theta(x^k + \alpha d^k) - \theta(x^k) \le -\epsilon\alpha\theta(x^k)$$

for any $\epsilon < 1$. Assumption (A3) follows immediately from 2. For assumption (A4), it is proven in [Kiw85] that θ is subdifferentially regular and hence (A4) is equivalent to (A2).

5 Application to Nonlinear Complementarity Problems

We consider applying the algorithms described in Section 3 to solve the nonlinear complementarity problem, NCP(F), namely to find $x \in \mathbb{R}^n$ such that

$$z \ge 0, F(z) \ge 0, \langle z, F(z)\rangle = 0 \tag{41}$$

for a given function $F\colon \mathbb{R}^n \to \mathbb{R}^n$. We will use the following equivalent set of equations

$$H_i(x) = (F_i(x) - x_i)^2 - F_i(x)\,|F_i(x)| - x_i\,|x_i|\,, i = 1, \ldots, n \tag{42}$$

The solution of these equations was shown to be equivalent to the nonlinear complementarity problem. In particular, Mangasarian[Man76] proved the following theorem.

Theorem 9 *x solves NCP(F) if and only if $H(x) = 0$, where H is defined in (42).*

In particular, note that H is differentiable.

It can be shown that our regularity condition (26) is equivalent to

$$\begin{aligned}\textstyle\sum_{i=1}^n &[F_i(x)(F_i(x) - |F_i(x)| - x_i)^2 + x_i(x_i - |x_i| - F_i(x))(F_i(x) - |F_i(x)| - x_i)]\,\nabla F_i(x) \\ &+ [F_i(x)(F_i(x) - |F_i(x)| - x_i)(x_i - |x_i| - F_i(x)) + x_i(x_i - |x_i| - F_i(x))^2]\,e_i = 0\end{aligned}$$

implying that x is a solution of NCP(F).

Other algorithmic approaches for solving NCP(F) have been tried. Recently, much emphasis has been placed on approaches using nonsmooth equations[Rob88, Pan89, HPR89, HX90]. We mention that there are many ways of formulating this problem as a system of nonsmooth equations. For example, the following ways are well–known.

1. The min operator technique, where

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

   or equivalently

$$H(z) := z - (z - F(z))_+$$

   where the "plus" operator signifies projection onto the positive orthant, that is $(x_+)_i := \max\{x_i, 0\}$. It is easy to see that $H(z) = 0$ if and only if z solves (41). In the affine case, the above H has been termed the "natural residual".

2. The Minty equations

$$H(x) := F(x_+) + x - x_+$$

   It is easy to show that $H(x) = 0$ implies that $z := x_+$ solves (41) and if z solves (41) then $x := z - F(z)$ satisfies $H(x) = 0$.

We favor using the differentiable equations (42) since the direction finding subproblem consists of solving a set of linear equations, rather than a mixed linear complementarity problem. Furthermore, we believe that the trade–off for this easier subproblem manifests itself in the form of ill–conditioning for which the nonmonotone line search technique has proven effective[GLL86].

The condition (26) appears to be a mild assumption on the problem in this case. Certainly it is not implied by the conditions which are assumed to guarantee convergence for the nonsmooth algorithms mentioned above. In particular, the following example satisfies (26) but is not regular in the sense of [HPR89].

Example 10 *Let* $F(x) = \begin{bmatrix} 1 - x_2 \\ x_1 \end{bmatrix}$. *The solution set of this linear complementarity problem is* $\{(0, \delta) \mid 0 \leq \delta \leq 1\}$. *It can be shown that* $\nabla\theta(x) = 0$ *implies that* x *solves the linear complementarity problem and thus (26) is satisfied. However, all the solution points are not regular in the sense of [HPR89, page 15].*

In order to justify performing a standard Newton method, we need the Jacobian of H to be nonsingular. Mangasarian[Man76] gives the following sufficient conditions to guarantee this nonsingularity.

Proposition 11 *Let* x *solve NCP(F) and satisfy* $x + F(x) > 0$. *If* $\nabla F(x)$ *has nonsingular principal minors, then* $\nabla H(x)$ *is nonsingular.*

We present some results on standard nonlinear complementarity problems found in the literature. A fuller description of the problems can be found in [HX90]. Note that an entry of F in the table signifies that the algorithm failed to converge.

In all the above examples we used the starting points suggested in [HX90]. In some cases, we added extra starting points when the points chosen in the aforementioned paper

Starting Point	Algorithm 1 Gradient (Function) Evaluations	Algorithm 2 Gradient (Function) Evaluations	Algorithm 3 Gradient (Function) Evaluations
(0,...,0)	F	30(30)	35(14)
(1,...,1)	9(3)	9(3)	8(3)
(100,...,100)	19(8)	19(8)	18(34)
(1,0,1,0)	8(3)	9(3)	8(8)
(1,0,0,0)	8(3)	9(3)	8(3)
(0,1,1,0)	20(4)	12(3)	8(5)

Table 1: Josephy's example[Jos79]

LCP Dimension (Started at Origin)	Algorithm 1 Gradient (Function) Evaluations	Algorithm 2 Gradient (Function) Evaluations	Algorithm 3 Gradient (Function) Evaluations
8	23(4)	18(30)	8(8)
16	20(3)	16(28)	8(9)
32	17(3)	15(4)	8(11)
64	42(12)	18(4)	9(12)
128	21(3)	18(4)	9(14)

Table 2: Murty's exponential linear complementarity problem[Mur88]

Starting Point	Algorithm 1 Gradient (Function) Evaluations	Algorithm 2 Gradient (Function) Evaluations	Algorithm 3 Gradient (Function) Evaluations
(2.5,5.5,1.5,4.5)	6(2)	21(4)	19(11)
(3.5,4.5,0.5,4.0)	7(3)	20(4)	22(16)
(2.5,1.5,1.5,3.5)	12(3)	21(4)	21(16)
(3.5,6.5,0.5,5.5)	8(3)	20(4)	22(15)
(1.0,1.0,1.0,1.0)	F	20(4)	20(17)
(10.0,10.0,10.0,10.0)	15(4)	25(8)	23(11)

Table 3: Mathiesen's Walrasian equilibrium model[Mat87]

Starting Point	Algorithm 1 Gradient (Function) Evaluations	Algorithm 2 Gradient (Function) Evaluations	Algorithm 3 Gradient (Function) Evaluations
x_1	19(3)	15(3)	18(22)
x_2	23(4)	28(4)	F

$x_1 = (0.2, 0.2, 0.2, 0.1, 0.1, 0.2, 0.5, 0.0, 4.0, 0.0, 0.0, 0.0, 0.4, 0.0)$
$x_2 = (0.1, 0.1, 0.1, 0.1, 0.1, 0.2, 0.5, 0.0, 4.0, 0.0, 0.0, 0.0, 0.4, 0.0)$

Table 4: Scarf's economic equilibrium model[Sca73]

Starting Point	Algorithm 1 Gradient (Function) Evaluations	Algorithm 2 Gradient (Function) Evaluations	Algorithm 3 Gradient (Function) Evaluations
(1,...,1)	15(3)	15(3)	13(18)
(10,...,10)	11(2)	11(2)	9(15)

Table 5: Nash noncooperative game example[Har88]

Starting Point	Algorithm 1 Gradient (Function) Evaluations	Algorithm 2 Gradient (Function) Evaluations	Algorithm 3 Gradient (Function) Evaluations
(0,...,0)	28(23)	28(40)	42(101)
(1,...,1)	33(18)	29(32)	27(72)
x_1	34(20)	25(40)	24(62)

$x_1 = (1.0, 1.1, \ldots, 1.9, 1.0, 1.1, \ldots, 1.9, 1.0, 1.1, \ldots, 1.9, 1.0.1.1, \ldots, 1.9, 1.0, 1.1)$

Table 6: Tobin's spatial price equilibrium model[Tob88]

Starting Point	Algorithm 1 Gradient (Function) Evaluations	Algorithm 2 Gradient (Function) Evaluations	Algorithm 3 Gradient (Function) Evaluations
x_1	53(7)	28(40)	F
x_2	17(3)	29(32)	30(52)
x_3	24(4)	28(40)	54(79)
x_4	19(3)	29(32)	39(31)

$$x_1 = (0,2,2,0,0,2,0,0,1,1,0,0,0,0,0)$$
$$x_2 = (0,3,3,0,0,3,0,0,1,1,0,0,0,0,0)$$
$$x_3 = (0,10,10,0,0,10,0,0,1,1,0,0,0,0,0)$$
$$x_4 = (0,5,5,0,0,5,0,0,1,1,0,0,0,0,0)$$

Table 7: Powell's nonlinear programming problem[Pow69]

Starting Point	Algorithm 1 Gradient (Function) Evaluations	Algorithm 2 Gradient (Function) Evaluations	Algorithm 3 Gradient (Function) Evaluations
(0,...,0)	F	33(24)	35(34)
(10,...,10)	22(5)	18(5)	17(4)

Table 8: Spatial competition example[Har86]

were considered too close to an optimal point. Furthermore, the functions and gradients were coded without any problem specific knowledge. In particular, the use of fixing a numeraire in the economic equilibrium problems was not used to force the Jacobian to be nonsingular at the solution. Needless to say, this would improve our results, but this removes the essential difficulty of the problems. While it is true that some of the iteration counts given above are relatively large, we wish to emphasize that all these results were obtained with one setting of the algorithm parameters and without particular fixes for specific problems.

References

[Ber82] D.P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, New York, 1982.

[BS91] R.S. Bain and W.E. Stewart. Application of robust nonmonotonic descent criteria to the solution of nonlinear algebraic equation systems. *Computers in Chemical Engineering*, 15:203–208, 1991.

[Bur80] O.P. Burdakov. Some globally convergent modifications of Newtons's method for solving systems of nonlinear equations. *Soviet Math Doklady*, 22(2):376–379, 1980.

[Cla83] F.H. Clarke. *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York, 1983.

[DS83] J.E. Dennis and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, New Jersey, 1983.

[Fra88] C. Fraley. Algorithms for nonlinear least squares. Technical Report SOL 88.16, Stanford University, July 1988.

[GLL86] L. Grippo, F. Lampariello, and S. Lucidi. A nonmonotone line search technique for Newton's method. *SIAM Journal of Numerical Analysis*, 23:707–716, 1986.

[GLL90] L. Grippo, F. Lampariello, and S. Lucidi. A class of nonmonotone stabilization methods in unconstrained optimization. Technical Report IASI-CNR R.290, 1990 (to appear in *Numerische Mathematik*).

[Har86] P.T. Harker. Alternative models of spatial competition. *Operations Research*, 34:410–425, 1986.

[Har88] P.T. Harker. Accelerating the convergence of the diagonalization and projection algorithms for finite–dimensional variational inequalities. *Mathematical Programming*, 41:29–50, 1988.

[HPR89] S.-P. Han, J.-S. Pang, and N. Rangaraj. Globally convergent Newton methods for nonsmooth equations. Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218, 1989.

[HX90] P.T. Harker and B. Xiao. Newton's method for the nonlinear complementarity problem: A B–differentiable equation approach. *Mathematical Programming*, 48:339–358, 1990.

[Jos79] N.H. Josephy. Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin, 1979.

[Kiw85] K.C. Kiwiel. *Methods of Descent for Nondifferentiable Optimization.* Springer–Verlag, 1985.

[Man76] O.L. Mangasarian. Equivalence of the complementarity problem to a system of nonlinear equations. *SIAM Journal of Applied Mathematics*, 31:89–92, 1976.

[Mat87] L. Mathiesen. An algorithm based on a sequence of linear complementarity problems applied to a Walrasian equilibrium model: An example. *Mathematical Programming*, 37:1–18, 1987.

[Mur88] K.G. Murty. *Linear Complementarity, Linear and Nonlinear Programming.* Helderman–Verlag, Berlin, 1988.

[OR70] J.M. Ortega and W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, 1970.

[Pan89] J.-S. Pang. A B–differentiable equation based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems. Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218, 1989.

[PD78] B.N. Pshenichny and Yu. M. Danilin. *Numerical Methods in Extremal Problems.* MIR Publishers, Moscow, 1978.

[Pol76] E. Polak. On the global stabilization of locally convergent algorithms. *Automatica*, 12:337–349, 1976.

[Pow69] M.J.D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, pages 283–298. Academic Press, London, 1969.

[Ral91] D. Ralph. Global convergence of damped Newton's method for nonsmooth equations, via the path search. Technical Report 90–1181, Computer Science Department, Cornell University, Ithaca, NY 14853, 1991.

[Rob88] S.M. Robinson. Newton's method for a class of nonsmooth functions. Working paper, Department of Industrial Engineering, University of Wisconsin, Madison, Wisconsin, 1988.

[SB80] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis.* Springer–Verlag, New York, 1980.

[Sca73] H.E. Scarf. *The Computation of Economic Equilibria.* Yale University Press, New Haven, Conneticut, 1973.

[Sub85] P.K. Subramanian. *Iterative Methods of Solution for Complementarity Problems.* PhD thesis, University of Wisconsin, Madison, Wisconsin, 1985.

[Tob88] R.L. Tobin. A variable dimension solution approach for the general spatial equilibrium problem. *Mathematical Programming*, 40:33–51, 1988.