JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 81, No. 1, APRIL 1994

Nonmonotone Stabilization Methods for Nonlinear Equations[1]

M. C. FERRIS[2] AND S. LUCIDI[3]

Communicated by O. L. Mangasarian

Abstract. We are concerned with defining new globalization criteria for solution methods of nonlinear equations. The current criteria used in these methods require a sufficient decrease of a particular merit function at each iteration of the algorithm. As was observed in the field of smooth unconstrained optimization, this descent requirement can considerably slow the rate of convergence of the sequence of points produced and, in some cases, can heavily deteriorate the performance of algorithms. The aim of this paper is to show that the global convergence of most methods proposed in the literature for solving systems of nonlinear equations can be obtained using less restrictive criteria that do not enforce a monotonic decrease of the chosen merit function. In particular, we show that a general stabilization scheme, recently proposed for the unconstrained minimization of continuously differentiable functions, can be extended to methods for the solution of nonlinear (nonsmooth) equations. This scheme includes different kinds of relaxation of the descent requirement and opens up the possibility of describing new classes of algorithms where the old monotone linesearch techniques are replaced with more flexible nonmonotone stabilization procedures. As in the case of smooth unconstrained optimization, this should be the basis for defining more efficient algorithms with very good practical rates of convergence.

Key Words. Nonlinear equations, stabilization techniques, global convergence.

---

[1]This material is partially based on research supported by the Air Force Office of Scientific Research Grant AFOSR-89-0410, National Science Foundation Grant CCR-91-57632, and Istituto di Analisi dei Sistemi ed Informatica del CNR.

[2]Professor, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin.
[3]Professor, Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza," Roma, Italy.

1. Introduction

In this paper, we shall be concerned with enforcing global convergence of Newton-type algorithms to solve systems of nonlinear equations where the underlying equations are not necessarily smooth. Since much of this work was motivated by a desire to solve nonlinear complementarity problems, we are particularly interested in the case of nonsmooth equations. In fact, many of the approaches for solving the nonlinear complementarity problem consider a reformulation of the problem as a system of nonsmooth equations.

Many of the globally convergent Newton-type methods for nonlinear equations proposed in the literature (see, for example, Refs. 1–6) present the following form:

$$x_{k+1} = x_k + \alpha_k d_k,$$

where d_k is the search direction and α_k is the stepsize along this direction.

Usually, the search direction is computed by considering the particular structure of the original problem. Typically, the direction solves a linear approximation of the system of nonlinear equations. This endows the search direction with good theoretical properties. In fact, if the unit stepsize is used, these algorithms produce sequences of points which are locally superlinearly convergent.

As for the stepsize α_k, it is chosen to satisfy stabilization criteria that guarantee the global convergence of the algorithm. The criteria are based on the requirement of sufficient decrease of a particular merit function at each step of the algorithm. These merit functions are scalar functions with the property that their global minima correspond to solutions of the given system of equations. The original problem of solving the system of equations is considered equivalent to the unconstrained minimization of the merit function and, hence, it appears natural to use classical globalization techniques from unconstrained optimization to determine the stepsize α_k. The search direction (obtained by solving a linearization of the system of equations) may need to be modified to ensure descent for the merit function, and the steplength may have to be reduced in order to enforce sufficient decrease of the merit function. These modifications may destroy some of the good properties of the original Newton-type directions. This is well known in the field of smooth unconstrained optimization (see, for example, Refs. 7–8), where it has been observed that imposing strong descent requirements considerably slows the rate of convergence and, in certain cases, heavily deteriorates the performance of Newton-type algorithms. In our opinion, the descent requirements could be even more detrimental to solution methods for systems of nonlinear equations as there is no longer a strong connection between the merit function and the original problem (in particular, there is no general relation between the critical points of the merit function and the solutions of the system of equations). Since it appears very reasonable to use the structure of the original problem as much as possible (see also Ref. 9), the ideal algorithm should use the unit stepsize along the original Newton-type direction as frequently as possible without losing global convergence.

Recently in smooth unconstrained optimization, some results (see, for example, Refs. 7–8) have been proposed based on the preceding considerations. In fact, new more tolerant stabilization criteria have been proposed for the unconstrained minimization of continuously differentiable functions. These criteria ensure the global convergence without imposing, at each step, a sufficient reduction of the objective function. A first attempt at using these new results in the field of nonlinear equations has been made in Ref. 6, where the nonmonotone linesearch technique of Ref. 7 was proposed to minimize locally Lipschitzian functions.

In this paper, we draw our inspiration from the stabilization scheme described in Ref. 8. This scheme includes different strategies for enforcing global convergence without requiring a monotonic reduction of the merit function. The numerical results reported in Ref. 8 show that computationally this scheme can be very effective especially in the minimization of ill-conditioned functions. In particular, numerical examples were given to show that the scheme avoids problems associated with the Maratos effect. This behavior is very attractive if we want to solve a system of nonlinear equations. In fact, all the merit functions used to globalize the Newton-type methods proposed in the field of nonlinear equations are often either ill-conditioned smooth functions or nonsmooth functions that may be considered similar, from a numerical point of view, to ill-conditioned functions (see Refs. 10–11). Therefore, extending the stabilization scheme of Ref. 8 to nonlinear equations and combining it with an efficient Newton-type method should be the basis of efficient algorithms for solving systems of nonlinear equations. Unfortunately, this extension is not immediate because all the results described in Ref. 8 are based strongly on the following points:

(i) the merit function is continuously differentiable;
(ii) in some circumstances, the search direction must be uniformly related (see Ref. 12) to the gradient of the merit function;
(iii) in other circumstances, the search direction must be a forcing function of the gradient of the merit function;
(iv) the linesearch procedure requires exact knowledge (in addition to continuity) of the directional derivative of the merit function along the search direction.

Certainly, the preceding points are not satisfied by the solution methods for systems of nonsmooth equations; frequently, they are not even satisfied by the solution methods for nonlinear smooth equations.

The aim of this paper is to define a stabilization scheme similar to the one proposed in Ref. 8 that can be combined with the wide class of Newton-type methods proposed in the literature to solve systems of nonlinear (smooth or nonsmooth) equations. In order to do this, we describe our scheme in a very general framework. This framework is characterized by a merit function, an auxiliary function (which resembles a forcing function), and the search directions determined by the algorithm in question. We state minimal conditions to impose on the merit function, the auxiliary function, and the search directions in order to guarantee the global convergence of the stabilization scheme. Most of the efficient Newton-type methods proposed for nonlinear equations satisfy these conditions. Therefore, the results reported open the possibility of describing new classes of algorithms where the old monotone linesearch techniques are substituted by more flexible stabilization procedures. However, the definition of a particular algorithm is beyond the scope of this paper and will be the subject of future work.

2. Stabilization Strategies for Nonlinear Equations

In this section, we define general stabilization schemes for the solution of

$$\text{(NE)} \quad H(x) = 0,$$

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

We assume that there exists a locally Lipschitzian merit function M with the property that

$$M(x) \geq 0, \qquad \text{for all } x \in \mathbb{R}^n,$$

$$M(x) = 0, \qquad \text{if and only if } H(x) = 0.$$

We apply techniques from unconstrained optimization to effect the minimization of this merit function.

The algorithm that we consider has the form

$$x_{k+1} = x_k + \alpha_k d_k, \qquad k = 0, 1, \ldots,$$

where $x_0 \in \mathbb{R}^n$ is a given starting point, $d_k \neq 0$ is a search direction, and α_k is a stepsize. Our formulation also relies on an auxiliary function, $A: \mathbb{R}^{n+n+1} \to \mathbb{R}$, which is a generalization of the familiar notion of a forcing function (Ref. 13). The relationship between these constructs will be described in the sequel.

In order to obtain a method for the solution of (NE), we define a general stabilization scheme that includes different strategies for enforcing global convergence without requiring a monotonic reduction of the merit function. This scheme is based on the following observations. As we have said in the introduction, a Newton-type direction conveys a lot of information about the system of equations under consideration (certainly much more than the merit function). In particular, if the sequence $\{\|d_k\|\}$ goes to zero, then the corresponding sequence of points $\{x_k\}$ converges to a solution of the system of equations. Therefore, an effective criterion to control if we are in a region where the unit stepsize produces a superlinearly convergent sequence is to check whether the norm of the Newton-type direction is decreasing. Thus the normal step of the algorithm is to check whether the norm of the direction has sufficiently decreased. If it has, the algorithm accepts the unit stepsize without computing the merit function. Otherwise, after a check of the merit function and a possible backtrack, the algorithm again tries to accept the unit stepsize by using a nonmonotone Armijo-type linesearch procedure (see Ref. 7). In the description of the algorithm that follows, l denotes the iteration index where the merit function was evaluated and the corresponding iterate was last accepted to modify the reference value. The actual linesearch procedure is given below.

Linesearch. Find the smallest integer from $i = 0, 1, \ldots$ such that

$$M(x_k + 2^{-i}d_k) \leq \mathscr{R} - \gamma 2^{-i} A_k(x_k, d_k);$$

then, set $\alpha_k = 2^{-i}$, $l = k + 1$, and update $\mathscr{R}$.

In order to prevent the sequence of points leaving the region of interest (with possible occurrence of overflows) the merit function is computed and its value is compared with an adjustable reference value at least every $\mathscr{N}$th iteration. If the value of the merit function is smaller than the reference value, the algorithm proceeds with a normal step, as above. Otherwise, the algorithm backtracks by restoring the vector of variables to the last point where the reference value test had been satisfied and performs a nonmonotone linesearch from that point. The precise form of the backtracking procedure is given below.

Backtrack. Replace x_k by x_l, set $k = l$, and recalculate d_k.

Formally, our complete algorithm model is the following.

Algorithm NMS. Nonmonotone Stabilization Algorithm.

Data. Choose x_0, $\delta_0 > 0$, $\beta \in (0, 1)$, $\gamma \in (0, 1)$, and $\mathcal{N} \geq 1$.

Initialization. Set $k = 0$, $l = 0$, $\delta = \delta_0$.
Compute $M(x_0)$ and set $\mathcal{R} = M(x_0)$.

Iteration. If $k \neq l + \mathcal{N}$, perform a d-step to calculate α_k; otherwise, perform an m-step to calculate α_k.
Set $x_{k+1} = x_k + \alpha_k d_k$, $k = k + 1$, and repeat the iteration.

d-Step. Compute d_k and stop if $\|d_k\| = 0$.
If $\|d_k\| \leq \delta$, perform (a); otherwise, perform (b) below.

(a) Set $\alpha_k = 1$, $\delta = \beta\delta$.

(b) Compute $M(x_k)$. If $M(x_k) \geq \mathcal{R}$, perform a backtrack and linesearch; otherwise, set $l = k$, update $\mathcal{R}$, and linesearch.

m-Step. Compute $M(x_k)$ and stop if $M(x_k) = 0$.
If $M(x_k) \geq \mathcal{R}$, perform (c); otherwise, perform (d) below.

(c) Perform a backtrack and a linesearch.

(d) Set $l = k$ and update $\mathcal{R}$. If $\|d_k\| \leq \delta$, set $\alpha_k = 1$, $\delta = \beta\delta$; otherwise, perform a linesearch.

For later reference, we introduce a new index j which is set initially at $j = 0$ and incremented each time we define $l = k$. Then, we indicate by $\{x_{l(j)}\}$ the sequence of points where the merit function is evaluated and by $\{\mathcal{R}_j\}$ the sequence of reference values. Furthermore, we also need the index $q(k)$ defined by

$$q(k) := \max[j \mid l(j) \leq k]. \tag{1}$$

Thus, $l(q(k))$ is the largest iteration index not exceeding k where the merit function was evaluated. For example,

$k =$	0	1	2	$\cdots$		10	11	$\cdots$		57	$\cdots$	$\cdots$
$j =$	0					1				2		$\cdots$
$l(j) =$	0					10				57		$\cdots$
$q(k) =$	0			$\cdots$	0	1		$\cdots$	1	2	2	$\cdots$

In order to complete the description of the algorithm, we must specify the criterion employed for updating $\mathcal{R}_j$, the reference value for the merit function. This is initially set to $M(x_0)$. Whenever a point $x_{l(j)}$ is generated such that $M(x_{l(j)}) < \mathcal{R}_j$, the reference value is updated by taking into account the memory [that is, a fixed number $m(j) \leq \bar{m}$ of previous values] of the merit function. To be precise, we require the updating rule for $\mathscr{R}_{j+1}$ to satisfy the following condition.

Reference Updating Rule. Given $\bar{m} \geq 0$, let $m(j+1)$ be such that

$$m(j+1) \leq \min[m(j)+1, \bar{m}].$$

Let

$$\mathscr{M}_{j+1} := \max_{0 \leq i \leq m(j+1)} M(x_{l(j+1-i)}), \tag{2}$$

and choose the value $\mathscr{R}_{j+1}$ to satisfy

$$M(x_{l(j+1)}) \leq \mathscr{R}_{j+1} \leq \mathscr{M}_{j+1}. \tag{3}$$

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

$$\mathscr{R}_{j+1} = \mathfrak{M}_{j+1} = \max_{0 \leq i \leq m(j+1)} M(x_{l(j+1-i)}), \tag{4}$$

$$\mathscr{R}_{j+1} = \max\left[M(x_{l(j+1)}), (1/[m(j+1)+1]) \sum_{i=0}^{m(j+1)} M(x_{l(j+1-i)})\right], \tag{5}$$

$$\mathscr{R}_{j+1} = \min[\mathscr{M}_{j+1}, (1/2)(\mathscr{R}_j + M(x_{l(j+1)}))]. \tag{6}$$

We note that (4) is the easiest to satisfy, while (5) and (6) define conditions which guarantee mean descent. See Fig. 1 for an example.

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) $\Omega_0 := \{x \mid M(x) \leq M(x_0)\}$ is bounded.

The auxiliary function A, the merit function M, and the search direction d must satisfy the following properties:

(A1) $A_k(x_k, d_k) \to 0$ implies $M(x_k) \to 0$;

(A2) $0 \geq -A_k(x_k, d_k) \geq M^D(x_k; d_k)$;

(A3) $[\mathscr{M}_{q(k)}]^\sigma \|d_k\|^\tau \leq \lambda A_k(x_k, d_k)$, $\tau \geq 1, \sigma > 0, \lambda > 0$, $\|d_k\| \leq \mu$.

Here, $M^D(x; v)$ is the Dini upper directional derivative of M at x in the direction v, defined as

$$M^D(x; v) = \limsup_{\lambda \downarrow 0} [M(x + \lambda v) - M(x)]/\lambda,$$

and $q(k)$ and $\mathscr{M}_{q(k)}$ are defined in (1) and (2), respectively.

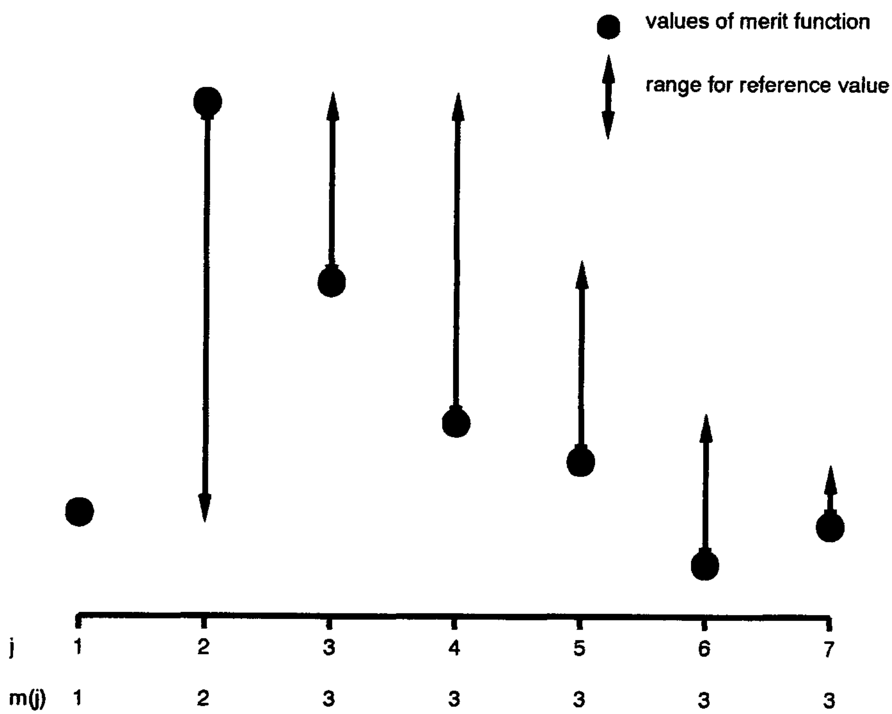

Fig. 1. Reference value ($\bar{m} = 3$).

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

(A4) $\|d_k\|^{\tau} \leq \lambda A_k(x_k, d_k), \qquad \tau \geq 2, \lambda > 0.$

Assumptions (A1)–(A3) appear to be minimal assumptions on the auxiliary function, the merit function, and the search direction. In particular, roughly speaking, Assumption (A1) implies that forcing the auxiliary function A to zero forces the current point to a solution of the system of equations. Assumption (A2) is needed to ensure, as we will see, that there exist values of the stepsize α which satisfy the sufficient decrease test of the linesearch procedure. Assumption (A3) requires that the sequence of search directions be bounded and that the sequence of values of the auxiliary function can go to zero only if either the norms of the directions go to zero or there exist, again roughly speaking, a subsequence of points that converges to a solution of the problem.

A further technical assumption is required to hold when the algorithm produces a sequence of stepsizes $\{\alpha_k\}$ which converge to zero. This is as follows:

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,

$$\text{(A5)} \quad \lim_{k\to\infty} -A_k(x_k, d_k) \geq \limsup_{k\to\infty} [M(x_k + \lambda_k d_k) - M(x_k)]/\lambda_k,$$

whenever the limit in the left-hand side exists.

This assumption is a strengthening of (A2). In fact, we note that, if M is subdifferentially regular (Ref. 14), then both (A2) and (A5) are equivalent to

$$0 \geq -A_k(x_k, d_k) \geq M'(x_k; d_k).$$

In order to prove the 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 the linesearch criterion of Algorithm NMS.

Lemma 2.1. Let M be locally Lipschitzian, and let $\gamma \in (0, 1)$ be arbitrary. Suppose that Assumptions (A1), (A2), (A3) hold. Then, there exists a scalar $\bar{\alpha} > 0$ such that, for all $\alpha \in [0, \bar{\alpha}]$,

$$M(x_k + \alpha d_k) \leq \mathscr{R}_{q(k)} - \gamma\alpha A_k(x_k, d_k);$$

that is, the linesearch criterion of Algorithm NMS can be satisfied.

Proof. If the linesearch procedure is carried out, either $M(x_k) > 0$ or $\|d_k\| > \delta > 0$. In the first case, Assumption (A1) guarantees that $A_k(x_k, d_k) \neq 0$; hence, Assumption (A2) implies that $M^D(x_k; d_k) < 0$, resulting in $d_k \neq 0$. 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

$$M(x_k + \alpha_l d_k) > \mathscr{R}_{q(k)} - \gamma\alpha_l A_k(x_k, d_k).$$

Using the definition of $\mathscr{R}_{q(k)}$, it can be seen that

$$M(x_k + \alpha_l d_k) - M(x_k) > -\gamma\alpha_l A_k(x_k, d_k).$$

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

$$M^D(x_k; d_k) \geq -\gamma A_k(x_k, d_k).$$

Assumption (A2) gives

$$-A_k(x_k, d_k) \geq -\gamma A_k(x_k, d_k),$$

which implies that $A_k(x_k, d_k) = 0$, which is a contradiction to either (A1) or (A3). □

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

Lemma 2.2. Suppose that M is locally Lipschitzian and that M, A, d_k satisfy Assumption (A2) for all k. Then:

(a) if $\{x_k\}$ converges to $\bar{x}$ and Assumption (A3) holds, then $\{A_k(x_k, d_k)\}$ is bounded;
(b) if $\{x_k\}$ is bounded and $\lim_{k\to\infty} \|d_k\| = 0$, then $\lim_{k\to\infty} A_k(x_k, d_k) = 0$.

Proof.
(a) Since M is locally Lipschitzian and $\{x_k\}$ converges, it follows that there exists a constant $\mu > 0$ such that, for all k,

$$|M^D(x_k; d_k)| \leq \mu \|d_k\|.$$

By Assumption (A2), we see that

$$\mu \|d_k\| \geq |M^D(x_k; d_k)| \geq A_k(x_k, d_k).$$

The boundedness of $\{A_k(x_k, d_k)\}$ now follows from (A3).
(b) If the conclusion of (b) is false, then

$$\limsup_{k\to\infty} A_k(x_k, d_k) = \bar{A} > 0.$$

Since $\{x_k\}$ is bounded, we can find a subsequence $k \in K$ such that

$$\lim_{k\in K} A_k(x_k, d_k) = \bar{A} > 0, \tag{7}$$

$$\lim_{k\in K} x_k = \bar{x}, \qquad \lim_{k\in K} \|d_k\| = 0. \tag{8}$$

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

$$\mu \|d_k\| \geq |M^D(x_k; d_k)| \geq A_k(x_k, d_k), \qquad k \in K, \tag{9}$$

and by using (8) and (9), we have

$$\lim_{k\in K} A_k(x_k, d_k) = 0,$$

which contradicts (7). □

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

Lemma 2.3. Suppose the 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 $\{\mathcal{M}_j\}$ is nonincreasing and has a limit $\mathcal{M}^*$;
(c) let $s(j)$ be an index in the set $\{l(j), l(j-1), \ldots, l(j-m(j))\}$ such that

$$M(x_{s(j)}) = \mathcal{M}_j = \max_{0 \le i \le m(j)} M(x_{l(j-i)}); \tag{10}$$

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

$$0 < h_k - k \le \mathcal{N}(\bar{m}+1), \qquad h_k = s(j_k),$$

$$\mathcal{M}_{j_k} = M(x_{h_k}) < \mathcal{M}_{q(k)}.$$

Proof. The proof of the lemma follows with minor modification from the proofs of Lemma 1 and Lemma 2 of Ref. 8. □

The following result is central to our development. We show that the merit function converges to a limit and also the product of the stepsize and the auxiliary function tends to zero. Note that these conclusions are trivial in the case of a monotone linesearch procedure. Unfortunately, in our case, the proof is more involved.

Proposition 2.1. Let $\{x_k\}$ be a sequence produced by the algorithm. Suppose that Assumptions (A1), (A2), (A3), and (C) hold and that M is locally Lipschitzian. Then, $\lim_{k\to\infty} M(x_k)$ exists and $\lim_{k\to\infty} \alpha_k A_k(x_k, d_k) = 0$.

Proof. We can split the iteration sequence $\{k\}$ into two parts, $\mathcal{L}$ and $\mathcal{U}$, namely those iterations where a linesearch of the merit function is carried out and those where the unit stepsize is accepted without performing a linesearch. Let $\{x_k\}_{k\in\mathcal{U}}$ denote the set (possibly empty) of points where the unit stepsize is accepted without a linesearch. Then,

$$\|d_k\| \le \delta_0 \beta^t, \qquad \alpha_k = 1, \qquad k \in \mathcal{U}, \tag{11}$$

where the integer t increases with $k\in\mathcal{U}$. It follows from (11) that, if $\mathcal{U}$ is an infinite set, $\|d_k\| \to 0$, for $k \to \infty$, $k\in\mathcal{U}$. Also in this case, by Lemma 2.2(b) and Lemma 2.3(a), we have

$$\lim_{\substack{k\to\infty \\ k\in\mathcal{U}}} \alpha_k A_k(x_k, d_k) = 0. \tag{12}$$

□

Claim 2.1. For any $i \geq 1$, we have

$$\lim_{j \to \infty} \alpha_{s(j)-i} A_{s(j)-i}(x_{s(j)-i}, d_{s(j)-i}) = 0, \tag{13}$$

$$\lim_{j \to \infty} M(x_{s(j)-i}) = \lim_{j \to \infty} M(x_{s(j)}) = \lim_{j \to \infty} \mathscr{M}_j = \mathscr{M}^*, \tag{14}$$

where $s(j)$ is defined by (10) and (1).

Proof. We proceed by induction. Assume first that $i = 1$. If $s(j) - 1 \in \mathscr{U}$, (13) is evident from (12) with $k = s(j) - 1$. Otherwise, if $s(j) - 1 \in \mathscr{L}$, recalling the acceptance criterion of the nonmonotone line search, we can write

$$\begin{aligned} \mathscr{M}_j = M(x_{s(j)}) &= M(x_{s(j)-1} + \alpha_{s(j)-1} d_{s(j)-1}) \\ &\leq \mathscr{M}_{q(s(j)-1)} - \gamma \alpha_{s(j)-1} A_{s(j)-1}(x_{s(j)-1}, d_{s(j)-1}). \end{aligned}$$

It follows that

$$\mathscr{M}_{q(s(j)-1)} - \mathscr{M}_j \geq \gamma \alpha_{s(j)-1} A_{s(j)-1}(x_{s(j)-1}, d_{s(j)-1}). \tag{15}$$

Therefore, if $s(j) - 1 \in \mathscr{L}$ for an infinite subsequence, from Lemma 2.3(b) and (15) we get

$$\lim_{j \to \infty} \alpha_{s(j)-1} A_{s(j)-1}(x_{s(j)-1}, d_{s(j)-1}) \to 0, \tag{16}$$

so that (13) holds for $i = 1$.

If follows from (A3) and (16) that

$$\lim_{j \to \infty} \alpha_{s(j)-1} \mathscr{M}^{\sigma}_{q(s(j)-1)} \|d_{s(j)-1}\|^{\tau} = 0,$$

and since $\{\alpha_k\}$, $\{\mathscr{M}_k\}$, $\{d_k\}$ are bounded from above and $\tau \geq 1$, $\sigma > 0$, it follows that

$$\lim_{j \to \infty} \alpha_{s(j)-1} \mathscr{M}_{q(s(j)-1)} \|d_{s(j)-1}\| = 0.$$

We consider two cases. Suppose first that

$$\limsup_{j \to \infty} \alpha_{s(j)-1} \|d_{s(j)-1}\| > 0.$$

Then, since $\lim_{j \to \infty} \mathscr{M}_j$ exists, it follows that $\lim_{j \to \infty} \mathscr{M}_j = 0$. However, by recalling that, by the definition of $\mathscr{M}_j$ and the description of the algorithm, we have that

$$M(x_{s(j)-1}) \leq \mathscr{M}_{q(s(j)-1)},$$

hence it is immediate that

$$\lim_{j \to \infty} M(x_{s(j)-1}) = \lim_{j \to \infty} \mathscr{M}_j = 0.$$

Then, (14) clearly holds for $i = 1$. Otherwise,

$$\limsup_{j \to \infty} \alpha_{s(j)-1} \|d_{s(j)-1}\| = 0,$$

which implies that

$$\lim_{j \to \infty} \alpha_{s(j)-1} \|d_{s(j)-1}\| = 0.$$

This in turn shows that

$$\|x_{s(j)} - x_{s(j)-1}\| \to 0,$$

so that (14) holds for $i = 1$ by the uniform continuity of M on the compact set containing $\{x_k\}$; see Lemma 2.3(a).

Assume now that (13) and (14) hold for a given i and consider the point $x_{s(j)-(i+1)}$. Reasoning as before, we can again distinguish the case $s(j) - (i+1) \in \mathscr{U}$, when (12) holds with $k = s(j) - (i+1)$, and the case $s(j) - (i+1) < \mathscr{L}$, in which we have

$$M(x_{s(j)-i}) \leq \mathscr{M}_{q(s(j)-(i+1))} - \gamma \alpha_{s(j)-(i+1)} A_{s(j)-(i+1)}(x_{s(j)-(i+1)}, d_{s(j)-(i+1)}),$$

and hence,

$$\mathscr{M}_{q(s(j)-(i+1))} - M(x_{s(j)-i}) \geq \gamma \alpha_{s(j)-(i+1)} A_{s(j)-(i+1)}(x_{s(j)-(i+1)}, d_{s(j)-(i+1)}). \tag{17}$$

Then, using (12), (14), (17), we can assert the Eq. (13) holds with i replaced by $i + 1$.

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

$$\lim_{j \to \infty} \alpha_{s(j)-(i+1)} \mathscr{M}_{q(s(j)-(i+1))} \|d_{s(j)-(i+1)}\| = 0.$$

Again, we must consider two cases. Suppose first that

$$\limsup_{j \to \infty} \alpha_{s(j)-(i+1)} \|d_{s(j)-(i+1)}\| > 0.$$

Then, since $\lim_{j \to \infty} \mathscr{M}_j$ exists, it follows that $\lim_{j \to \infty} \mathscr{M}_j = 0$, and using again that

$$M(x_{s(j)-(i+1)}) \leq \mathscr{M}_{q(s(j)-(i+1))},$$

we have

$$\lim_{j \to \infty} M(x_{s(j)-(i+1)}) = \lim_{j \to \infty} \mathscr{M}_j = 0.$$

Thus, in this case, (14) holds for $j + 1$. In the other case,

$$\limsup_{j \to \infty} \alpha_{s(j)-(i+1)} \|d_{s(j)-(i+1)}\| = 0,$$

which implies that

$$\lim_{j\to\infty} \sigma_{s(j)-(i+1)} \|d_{s(j)-(i+1)}\| = 0.$$

This implies, moreover, that

$$\|x_{s(j)-i} - x_{s(j)-(i+1)}\| \to 0,$$

so that by (14) and the uniform continuity of M on the compact set containing $\{x_k\}$,

$$\lim_{j\to\infty} M(x_{s(j)-(i+1)}) = \lim_{j\to\infty} M(x_{s(j)-i}) = \lim_{j\to\infty} \mathcal{M}_j.$$

Thus, (14) is satisfied with i replaced by $i+1$, which completes the induction. □

We now complete the proof of Proposition 2.1. We first show the $\lim_{k\to\infty} M(x_k)$ exists. Note that (13) and Assumption (A3) imply that, for every $i \geq 1$,

$$\lim_{j\to\infty} \alpha_{s(j)-i} \mathcal{M}_{q(s(j)-i)} \|d_{s(j)-i}\| = 0.$$

Again, there are two cases. The first one is that there exists an index $\bar{i}$ such that

$$\limsup_{j\to\infty} \alpha_{s(j)-\bar{i}} \|d_{s(j)-\bar{i}}\| > 0.$$

Then, since $\lim_{j\to\infty} \mathcal{M}_j$ exists, it follows that $\lim_{j\to\infty} \mathcal{M}_j = 0$; and since $0 \leq M(x_k) \leq \mathcal{M}_{q(k)}$, we have $\lim_{j\to\infty} M(x_k) = 0$. Now consider the case where, for every i, we have

$$\lim_{j\to\infty} \alpha_{s(j)-i} \|d_{s(j)-i}\| = 0, \tag{18}$$

and let x_k be any given point produced by the algorithm. By Lemma 2.3(c), there is a point x_{h_k} such that

$$x_{h_k} \in \{x_{s(j)}\} \quad \text{and} \quad 0 < h_k - k \leq \mathcal{N}(\bar{m}+1). \tag{19}$$

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 (18) [with $h_k = s(j)$] and (19), that

$$\lim_{k\to\infty} \|x_k - x_{h_k}\| = 0. \tag{20}$$

From the uniform continuity of M, it follows that

$$\lim_{k\to\infty} M(x_k) = \lim_{k\to\infty} M(x_{h_k}) = \lim_{j\to\infty} \mathcal{M}_j, \tag{21}$$

proving that $\lim_{k\to\infty} M(x_k)$ exists.

If $k \in \mathscr{L}$, we obtain

$$M(x_{k+1}) \leq \mathscr{M}_{q(k)} - \gamma \alpha_k A_k(x_k, d_k),$$

and hence,

$$\mathscr{M}_{q(k)} - M(x_{k+1}) \geq \gamma \alpha_k A_k(x_k, d_k). \tag{22}$$

Therefore, by (12), (21), (22), we conclude that

$$\lim_{k \to \infty} \alpha_k A_k(x_k, d_k) = 0,$$

as required. □

We are now able to prove our convergence result. Note that Assumption (A5) is only needed when $\limsup_{k \to \infty} \alpha_k = 0$.

Theorem 2.1. Let M be a locally Lipschitzian merit function, and suppose that (A1), (A2), (A3), and (C) hold. Then:

(a) if $\limsup_{k \to \infty} \alpha_k > 0$, then $\lim_{k \to \infty} M(x_k) = 0$;
(b) if $\limsup_{k \to \infty} \alpha_k = 0$ and if $\bar{x}$ is an accumulation point of $\{x_k\}$, where (A5) holds, then $M(\bar{x}) = 0$.

Proof. Suppose that

$$\limsup_{k \to \infty} \alpha_k = \bar{\alpha} > 0.$$

Since $\{x_k\}$ is bounded, we can find a subsequence $k \in K$ such that

$$\lim_{k \in K} \alpha_k = \bar{\alpha} \quad \text{and} \quad \lim_{k \in K} x_k = \bar{x}.$$

By Lemma 2.2, it follows that $\{A_k(x_k, d_k) \mid k \in K\}$ is bounded. By taking further subsequences if necessary, we may assume that $\lim_{k \in K} \alpha_k = \bar{\alpha}$ and $\lim_{k \in K} A_k(x_k, d_k)$ exists. However, from Proposition 2.1 we have

$$\lim_{k \in K} \alpha_k A_k(x_k, d_k) = 0.$$

Since $\bar{\alpha} > 0$, it follows that

$$\lim_{k \in K} A_k(x_k, d_k) = 0.$$

Assumption (A1) gives $\lim_{k \in K} M(x_k) = 0$, hence $\lim_{k \to \infty} M(x_k) = 0$, since from Proposition 2.1 the sequence $\{M(x_k)\}$ 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 (A5) holds, and let $\{x_k \mid k \in K\}$ converge to $\bar{x}$. Using Lemma 2.2, we may assume that $\{x_k \mid k \in K\}$, $\{d_k \mid k \in K\}$, $\{\alpha_k \mid k \in K\}$, and $\{A_k(x_k, d_k) \mid k \in K\}$ 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 $\alpha_k, k \in K$, is eventually produced by procedure linesearch. The definition of the linesearch in Algorithm NMS gives

$$M(x_k + (\alpha_k/v)d_k) - \mathscr{R}_{q(k)} > -\gamma(\alpha_k/v)A_k(x_k, d_k),$$

and the definition of $\mathscr{R}_{q(k)}$ implies

$$M(x_k + (\alpha_k/v)d_k) - M(x_k) > -\gamma(\alpha_k/v)A_k(x_k, d_k).$$

Using Assumption (A5) we have

$$\begin{aligned} -\lim_{k \in K} A_k(x_k, d_k) &\geq \limsup_{k \in K} [[M(x_k + (\alpha_k/v)d_k) - M(x_k)]/(\alpha_k/v)] \\ &\geq -\gamma \lim_{k \in K} A_k(x_k, d_k), \end{aligned}$$

which shows that $\lim_{k \in K} A_k(x_k, d_k) = 0$. It follows from (A1) that $M(\bar{x}) = 0$. □

3. Examples of Applications of the New Stabilization Strategies

In the preceding section, we have described a nonmonotone stabilization algorithm and we have given 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, and the directions determined by a particular algorithm. We prove the general convergence result under these assumptions, without specifying the particular merit function, the auxiliary function, or the direction, but only the conditions which they must satisfy. The described conditions are so weak that almost all the merit functions, auxiliary functions, and search directions used by the Newton-type methods proposed in the literature satisfy our conditions. Therefore, it is possible to combine our stabilization technique with these algorithms. In this section, we report on some examples of these methods.

As a first example, we consider the approach proposed in the work of Burdakov (Ref. 1). In this work, a system of smooth equations is considered, and it is assumed that the Jacobian of the system of equations is invertible everywhere and its inverse is uniformly bounded. The proposed merit function is given by

$$M(x) := \|H(x)\|,$$

the auxiliary function is given by

$$A_k(x_k, d_k) = M(x_k),$$

and the search direction is the Newton direction, namely,

$$d_k = -\nabla H(x_k)^{-1} H(x_k).$$

Finally Assumption (C) is required. It is very easy to show that Assumptions (A1)-(A5) hold. In fact, Assumption (A1) is immediate by the definition of A. Assumptions (A2) and (A5) follow from the fact that the author of Ref. 1 shows that

$$M^D(x; d_k) = -M(x).$$

Finally, Assumption (C) and the assumption that the inverse of the Jacobian is uniformly bounded imply Assumption (A3).

An interesting extension to the work of Pshenichny and Danilin (Ref. 15) on minimax problems can be seen by using our formulation. In this case, the merit function is given by

$$M(x) = \max_{1 \leq i < n} |H_i(x)|,$$

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

$$\mathcal{A}_\delta := \{i \mid |H_i(x)| \geq M(x) - \delta\},$$

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

(B1) $\exists \delta > 0$ such that, for all x with $M(x) > 0$, $M(x) \leq M(x_0)$, the linearized system

$$\nabla H_i(x_k)d + H_i(x_k) = 0, \qquad i \in \mathcal{A}_\delta \tag{23}$$

is solvable

(B2) Let $d(x)$ denote the minimum norm solution of (23). Then, $\exists c > 0$ such that, for all x with $M(x) > 0$, we have

$$\|d(x)\| \leq cM(x).$$

It can be shown that these assumptions imply our assumptions. Let $A_k(x_k, d_k) = \epsilon M(x_k)$, for $\epsilon \in (0, 1)$. Assumption (A1) is then immediate. Assumption (A2) follows from Ref. 15, Theorem 6.1, since it is shown that there exists an $\bar{\alpha} > 0$ such that, for all $\alpha \in (0, \bar{\alpha}_k)$,

$$M(x_k + \alpha d_k) - M(x_k) \leq -\epsilon \alpha M(x_k),$$

for any $\epsilon < 1$. Assumption (A3) follows immediately from (B2). For Assumption (A5), it is proven in Ref. 16 that M is subdifferentially regular, and hence (A5) is equivalent to (A2).

Finally, we consider the results proposed in Ref. 4. In this paper, the following method is described. Let

$$M(x) = (1/2)\|H(x)\|^2,$$

and choose the search direction to satisfy

$$H(x_k) + G(x_k, d_k) = 0$$

at each iteration. It is not assumed that H is a smooth function, but that $G(x_k, d)$ is an appropriate approximation of the directional derivative of H in the direction d at x_k. The motivation behind this analysis comes from nonlinear complementarity problems, and the difficulties associated with the inequalities present in these problems are replaced by the nonsmoothness of H. The assumptions made in Ref. 4 are essentially equivalent to the ones we make in Section 2. This can be seen by defining

$$A_k(x_k, d_k) := 2M(x_k).$$

In the same paper (Ref. 4), a Gauss–Newton method is also proposed. The same merit function is used. In this case, the direction is calculated by solving

$$\min_{d \in \mathbb{R}^n} [H(x_k)^T G(x_k, d) + (1/2)R_k(d)].$$

The assumptions made to prove convergence are essentially equivalent to (A1), (A2), (A4), and (C). Particular instances of functions R_k which are considered are:

(i) $R_k(d) = d^T B_k d$, where B_k is a symmetric positive definite $n \times n$ matrix;
(ii) $R_k(d) = \|G(x_k, d)\|^2 + \epsilon_k \|d\|^2$, where ϵ_k is a nonnegative scalar.

In Ref. 4, the conditions on R_k require in the first case that the sequence $\{B_k\}$ have eigenvalues which are bounded away from zero and in the second case that $\{\epsilon_k\}$ be bounded away from zero. In the model that we propose, we can relax these conditions by essentially using the following forms:

(i) $R_k(d) = d^T B_k d + c_1 \mathscr{M}_k \|d\|^2$, where B_k is a symmetric positive semidefinite $n \times n$ matrix;
(ii) $R_k(d) = \|G(x_k, d)\|^2 + c_1 \mathscr{M}_k \|d\|^2$.

This example shows that our analysis can lead to weaker requirements than those already cited in the literature.

Three further examples of the use of the framework presented in this paper are described in Ref. 17. In particular, these examples exploit the full generality of the auxiliary function depending on k, x_k, d_k.

References

1. BURDAKOV, O. P., *Some Globally Convergent Modifications of Newton's Method for Solving Systems of Nonlinear Equations*, Soviet Mathematics Doklady, Vol. 22, pp. 376–379, 1980.
2. STOER, J., and BULIRSCH, R., *Introduction to Numerical Analysis*. Springer-Verlag, New York, New York, 1980.
3. POLAK, E., *On the Global Stabilization of Locally Convergent Algorithms*, Automatica, Vol. 12, pp. 337–349, 1976.
4. HAN, S. P., PANG, J. S., and RANGARAJ, N., *Globally Convergent Newton Methods for Nonsmooth Equations*, Mathematics of Operations Research, Vol. 17, pp. 586–607, 1992.
5. HARKER, P. T., and XIAO, B., *Newton's Method for the Nonlinear Complementarity Problem: A B-Differentiable Equation Approach*, Mathematical Programming, Vol. 48, pp. 339–358, 1990.
6. PANG, J. S., HAN, S. P., and RANGARAJ, N., *Minimization of Locally Lipschitzian Functions*, SIAM Journal on Optimization, Vol. 1, pp. 57–82, 1991.
7. GRIPPO, L., LAMPARIELLO, F., and LUCIDI, S., *A Nonmonotone Line Search Technique for Newton's Method*, SIAM Journal on Numerical Analysis, Vol. 23, pp. 707–716, 1986.
8. GRIPPO, L., LAMPARIELLO, F., and LUCIDI, S., *A Class of Nonmonotone Stabilization Methods in Unconstrained Optimization*, Numerische Mathematik, Vol. 59, pp. 779–805, 1991.
9. DENNIS, J. E., and SCHNABEL, R. B., *Numerical Methods for Unconstrained Optimizations and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.
10. SHOR, N. Z., *Minimization Methods for Nondifferentiable Functions*, Springer-Verlag, Berlin, Germany, 1985.
11. ZOWE, J., *Nondifferentiable Optimization*, Computational Mathematical Programming, Edited by K. Schittkowski, Springer-Verlag, Berlin, Germany, pp. 321–356, 1985.
12. BERTSEKAS, D. P., *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, New York, 1982.
13. ORTEGA, J. M., and RHEINBOLDT, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, New York, 1970.
14. CLARKE, F. H., *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, New York, 1983.
15. PSHENICHNY, B. N., and DANILIN, YU. M., *Numerical Methods in Extremal Problems*, MIR Publishers, Moscow, Russia, 1978.
16. KIWIEL, K. C., *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics, Springer-Verlag, Berlin, Germany, Vol. 1113, 1985.
17. FERRIS, M. C., and LUCIDI, S., *Globally Convergent Methods for Nonlinear Equations*, Technical Report 1030, Computer Sciences Department, University of Wisconsin, 1991.