SIAM J. NUMER. ANAL.
Vol. 23, No. 4, August 1986

002

A NONMONOTONE LINE SEARCH TECHNIQUE FOR NEWTON'S METHOD*

L. GRIPPO†, F. LAMPARIELLO† AND S. LUCIDI†

Abstract. In this paper a nonmonotone steplength selection rule for Newton's method is proposed, which can be viewed as a generalization of Armijo's rule. Numerical results are reported which indicate that the proposed technique may allow a considerable saving both in the number of line searches and in the number of function evaluations.

Key words. nonlinear programming, unconstrained minimization, Newton's method, line search techniques

AMS(MOS) subject classifications. 90C30, 49D15, 65K05

1. Introduction. We consider the problem of minimizing a twice continuously differentiable function $f(x)$, $x \in R^n$, by means of Newton's method. It is known that this method in its pure form has serious drawbacks, so that several modifications have been proposed in order to ensure global convergence towards local minima (see e.g. [11], [15], [6], [5]). All these schemes require the use of a line search technique which guarantees a monotonical decrease of the objective function. On the other hand it is recognized [5], [8], [3] that enforcing monotonicity of the function values by selecting a stepsize not equal to unity can considerably slow the rate of convergence in the intermediate stages of the minimization process, especially in the presence of narrow curved valleys. Therefore it appears that an ideal line search strategy for Newton's method should allow an increase in the function value at each step, while retaining global convergence.

A line search technique which exhibits, to some extent, these features is the so-called watchdog technique [3], which has been proposed in connection with recursive quadratic programming methods for constrained optimization.

The approach considered in this paper applies, in essence, the same basic idea of the watchdog technique, by relaxing some standard line search condition in a way that preserves a global convergence property. More specifically we impose that the function value of each new iterate satisfy an Armijo's condition with respect to the maximum function value of a prefixed number of previous iterates. Thus the proposed technique can be viewed as a generalization of Armijo's rule [1] in the sense that it allows an increase in the function values without affecting the convergence properties.

2. Steplength selection for Newton's method. Let $f: R^n \to R$ be a twice continuously differentiable function and denote by $g(x)$ the gradient of f and by $H(x)$ the Hessian matrix.

Let $\{x_k\}$ be the sequence defined by:

$$x_{k+1} = x_k - \alpha_k H_k^{-1} g_k, \qquad k = 0, 1, \cdots \tag{1}$$

where $g_k \triangleq g(x_k)$ and $H_k \triangleq H(x_k)$.

It is well known that, if at a stationary point x^* of f the matrix $H(x^*)$ is nonsingular, then there is a neighborhood Ω of x^* such that for any $x_0 \in \Omega$ the Newton iterates (i.e. (1) with $\alpha_k = 1$) are well defined, remain in Ω and converge to x^*. The use of a line

* Received by the editors January 22, 1985, and in revised form August 9, 1985.

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

search technique along the Newton direction has two motivations: the difficulty of providing a starting point x_0 belonging to Ω and the need of avoiding convergence towards local maxima. It can be shown however that, if H is nonsingular and $g_k' H_k^{-1} g_k \neq 0$, convergence to local maxima can be prevented by simply ensuring that the direction of movement is a descent direction. This fact is stated in the following proposition.

PROPOSITION. *Assume that the sequence* $\{x_k\}$ *produced by* (1) *with*

$$\alpha_k = \operatorname{sgn}[g_k' H_k^{-1} g_k]$$

is well defined and remains in a set S where H is nonsingular. Moreover assume that

(i) $\{x_k\}$ *converges to* $x^* \in S$;

(ii) $g_k' H_k^{-1} g_k \neq 0$, *for all k.*

Then x^* *is a stationary point of f in S which cannot be a local maximum.*

Proof. Since by (i): $\|x_{k+1} - x_k\| = \|H_k^{-1} g_k\| \to 0$ and $H(x^*)$ is nonsingular, we have $g(x^*) = 0$. Assume that x^* is a local maximum, so that there exists a closed sphere $B(x^*)$ where $y'H(x)y < 0$ for all $x \in B(x^*)$, $y \neq 0$ and a number $\eta > 0$ such that $\lambda_{\max}[H(x)] \leqq -\eta$ for all $x \in B(x^*)$, where $\lambda_{\max}[H]$ is the maximum eigenvalue of H. Then, by (i) and (ii), there exists a $\bar{k}$ such that $\alpha_k = -1$ for all $k \geqq \bar{k}$. Therefore we can write:

$$\begin{aligned} f(x_{k+1}) &= f(x_k) + g_k'[x_{k+1} - x_k] + \tfrac{1}{2}[x_{k+1} - x_k]' H_k [x_{k+1} - x_k] + r(\|x_{k+1} - x_k\|) \\ &= f(x_k) + \tfrac{3}{2}[x_{k+1} - x_k]' H_k [x_{k+1} - x_k] + r(\|x_{k+1} - x_k\|) \\ &\leqq f(x_k) - \tfrac{3}{2}\eta \|x_{k+1} - x_k\|^2 + r(\|x_{k+1} - x_k\|) \quad \text{for } k \geqq \bar{k}, \end{aligned}$$

where $\lim_{\varepsilon \to 0^+} (r(\varepsilon)/\varepsilon^2) = 0$.

It follows that, for sufficiently large k, $f(x_{k+1}) < f(x_k)$ and this contradicts the assumption that x^* is a local maximum. □

In practice the use of descent directions also prevents convergence to saddle points, although this cannot be excluded in principle.

Consider, as an example, the minimization of Wood's function (see Problem 2 of § 4). In this case Newton's method with unit stepsize converges to a saddle point near $[-1, 1, -1, 1]'$. However if we take $\alpha_k = \operatorname{sgn}[g_k' H_k^{-1} g_k]$ as suggested in the preceding proposition we obtain convergence to the required optimal solution $x^* = [1, 1, 1, 1]'$ in 31 iterations with $f(x^*) < 10^{-38}$. The same behaviour occurs even if we assume as a starting point $[-1, 1, -1, 1]'$.

This device however does not ensure in general convergence of the iterates. For example, on the helical valley function (see Problem 6 of § 4) the algorithm diverges rapidly.

It can be argued that the adoption of a line search technique along the Newton's direction has essentially the objective of ensuring convergence of the sequence $\{x_k\}$. On the other hand the use of a stepsize selection rule which enforces monotonicity can considerably slow the rate of convergence. This can be illustrated by considering the minimization of Rosenbrock's function (see Problem 1 of § 4 with $n = 2$). In this case Newton's method with unit stepsize converges to the optimal solution in 7 iterations. It can be observed that, although the search directions are descent directions, the sequence of the function values is not monotonic as shown in Table 1. On the same problem, a monotone line search technique, such as Armijo's method, requires for the same accuracy 22 iterations and 30 function evaluations.

The foregoing discussion gives evidence for the potential usefulness of a non-monotonic line search technique.

TABLE 1

Iteration number	x_1	x_2	f
0	-1.2	1	24.2
1	-1.175	1.381	4.73188
2	0.7631	-3.175	1.41×10^4
3	0.7634	0.5828	0.05596
4	1.000	0.944	0.31319
5	1.000	1.000	1.85×10^{-11}
6	1.000	1.000	3.43×10^{-20}
7	1	1	$<10^{-38}$

3. A nonmonotone line search technique. In this section we define an acceptability criterion for the stepsize which can be viewed as a generalization of the Armijo rule [1]. In particular, we show that the condition which implies a monotonic decrease of $f(x_k)$ can be relaxed and yet global convergence can be established.

We state the following result.

THEOREM. *Let $\{x_k\}$ be a sequence defined by*

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

Let $a>0$, $\sigma\in(0,1)$, $\gamma\in(0,1)$ and let M be a nonnegative integer. Assume that:

(i) *the level set $\Omega_0 \triangleq \{x: f(x)\leqq f(x_0)\}$ is compact*;

(ii) *there exist positive numbers c_1, c_2 such that*:

$$g_k' d_k \leqq -c_1\|g_k\|^2, \tag{2}$$

$$\|d_k\|\leqq c_2\|g_k\|; \tag{3}$$

(iii) *$\alpha_k=\sigma^{h_k}a$, where h_k is the first nonnegative integer h for which*:

$$f(x_k+\sigma^h a d_k)\leqq \max_{0\leqq j\leqq m(k)}[f(x_{k-j})]+\gamma\sigma^h a g_k' d_k, \tag{4}$$

where $m(0)=0$ and $0\leqq m(k)\leqq \min[m(k-1)+1, M]$, $k\geqq 1$.

Then:

(a) *the sequence $\{x_k\}$ remains in Ω_0 and every limit point $\bar{x}$ satisfies $g(\bar{x})=0$*;

(b) *no limit point of $\{x_k\}$ is a local maximum of f*;

(c) *if the number of the stationary points of f in Ω_0 is finite, the sequence $\{x_k\}$ converges.*

Proof. Let $l(k)$ be an integer such that:

$$k-m(k)\leqq l(k)\leqq k, \tag{5}$$

$$f(x_{l(k)})=\max_{0\leqq j\leqq m(k)}[f(x_{k-j})].$$

By (iii) we have that the sequence $\{f(x_{l(k)})\}$ is nonincreasing. In fact, taking into account that $m(k+1)\leqq m(k)+1$, one can write:

$$f(x_{l(k+1)})=\max_{0\leqq j\leqq m(k+1)}[f(x_{k+1-j})]$$

$$\leqq \max_{0\leqq j\leqq m(k)+1}[f(x_{k+1-j})]=\max[f(x_{l(k)}), f(x_{k+1})]=f(x_{l(k)}).$$

Moreover we obtain from (4), for $k>M$:

$$
\begin{aligned}
f(x_{l(k)}) &= f(x_{l(k)-1}+\alpha_{l(k)-1}d_{l(k)-1}) \\
&\leqq \max_{0\leqq j\leqq m(l(k)-1)} [f(x_{l(k)-1-j})]+\gamma\alpha_{l(k)-1}g'_{l(k)-1}d_{l(k)-1} \qquad (6)\\
&= f(x_{l(l(k)-1)})+\gamma\alpha_{l(k)-1}g'_{l(k)-1}d_{l(k)-1}.
\end{aligned}
$$

Now, since $f(x_k)\leqq f(x_0)$ for all k, $\{x_k\}\subset\Omega_0$ so that $\{f(x_{l(k)})\}$ admits a limit for $k\to\infty$. Also since $\alpha_k>0$ and $g'_kd_k<0$, it follows from (6) that:

$$\lim_{k\to\infty}\alpha_{l(k)-1}g'_{l(k)-1}d_{l(k)-1}=0. \tag{7}$$

By (ii) we have $\alpha_k g'_k d_k\leqq -c_1\alpha_k\|g_k\|^2\leqq-(c_1/c_2^2)\alpha_k\|d_k\|^2$ for all k, and thus, since $\alpha_k\leqq a$, (7) implies:

$$\lim_{k\to\infty}\alpha_{l(k)-1}\|d_{l(k)-1}\|=0. \tag{8}$$

We prove now that $\lim_{k\to\infty}\alpha_k\|d_k\|=0$. Let

$$\hat{l}(k)\triangleq l(k+M+2).$$

First we show, by induction, that for any given $j\geqq 1$

$$\lim_{k\to\infty}\alpha_{\hat{l}(k)-j}\|d_{\hat{l}(k)-j}\|=0 \tag{9}$$

and

$$\lim_{k\to\infty}f(x_{\hat{l}(k)-j})=\lim_{k\to\infty}f(x_{l(k)}). \tag{10}$$

(Here and in the sequel we assume, without loss of generality, that the iteration index k is large enough to avoid the occurrence of negative subscripts, that is $k\geqq j-1$.) If $j=1$, since $\{\hat{l}(k)\}\subset\{l(k)\}$, (9) follows from (8). This in turn implies $\|x_{\hat{l}(k)}-x_{\hat{l}(k)-1}\|\to 0$, so that (10) holds for $j=1$, since $f(x)$ is uniformly continuous on Ω_0. Assume now that (9) and (10) hold for a given j. Then by (4) one can write:

$$f(x_{\hat{l}(k)-j})\leqq f(x_{l(\hat{l}(k)-j-1)})+\gamma\alpha_{\hat{l}(k)-j-1}g'_{\hat{l}(k)-j-1}d_{\hat{l}(k)-j-1}.$$

Taking limits for $k\to\infty$, we have, by (10):

$$\lim_{k\to\infty}\alpha_{\hat{l}(k)-(j+1)}g'_{\hat{l}(k)-(j+1)}d_{\hat{l}(k)-(j+1)}=0.$$

Using the same arguments employed for deriving (8) from (7), we obtain:

$$\lim_{k\to\infty}\alpha_{\hat{l}(k)-(j+1)}\|d_{\hat{l}(k)-(j+1)}\|=0.$$

Moreover this implies $\|x_{\hat{l}(k)-j}-x_{\hat{l}(k)-(j+1)}\|\to 0$, so that by (10) and the uniform continuity of f on Ω_0:

$$\lim_{k\to\infty}f(x_{\hat{l}(k)-(j+1)})=\lim_{k\to\infty}f(x_{\hat{l}(k)-j})=\lim_{k\to\infty}f(x_{l(k)}).$$

Thus we conclude that (9) and (10) hold for any given $j\geqq 1$.

Now for any k:

$$x_{k+1}=x_{\hat{l}(k)}-\sum_{j=1}^{\hat{l}(k)-k-1}\alpha_{\hat{l}(k)-j}d_{\hat{l}(k)-j}. \tag{11}$$

By (5) we have: $\hat{l}(k)-k-1=l(k+M+2)-k-1\leqq M+1$, so that (11) implies, by (9):

$$\lim_{k\to\infty}\|x_{k+1}-x_{\hat{l}(k)}\|=0. \tag{12}$$

Since $\{f(x_{l(k)})\}$ admits a limit, it follows from the uniform continuity of f on Ω_0:

$$\lim_{k\to\infty} f(x_k)=\lim_{k\to\infty} f(x_{\hat{l}(k)}). \tag{13}$$

By (4) we have:

$$f(x_{k+1})\leqq f(x_{l(k)})+\gamma\alpha_k g_k' d_k.$$

Taking limits for $k\to\infty$, by (13) we obtain: $\lim_{k\to\infty}\alpha_k g_k' d_k=0$ which implies, as noted before:

$$\lim_{k\to\infty}\alpha_k\|d_k\|=0 \tag{14}$$

and

$$\lim_{k\to\infty}\alpha_k\|g_k\|^2=0. \tag{15}$$

Now let $\bar{x}$ be any limit point of $\{x_k\}$ and relabel $\{x_k\}$ a subsequence converging to $\bar{x}$. Then by (15) either $\lim_{k\to\infty}\|g_k\|=0$, which implies, by continuity, $g(\bar{x})=0$, or there exists a subsequence $\{x_k\}_K\subset\{x_k\}$ such that:

$$\lim_{k\to\infty, k\in K}\alpha_k=0.$$

In this case, by assumption (iii) there exists an index $\bar{k}$ such that, for all $k\geqq\bar{k}$, $k\in K$:

$$f\left(x_k+\frac{\alpha_k}{\sigma}d_k\right)>\max_{0\leqq j\leqq m(k)}[f(x_{k-j})]+\gamma\frac{\alpha_k}{\sigma}g_k' d_k\geqq f(x_k)+\gamma\frac{\alpha_k}{\sigma}g_k' d_k.$$

By the Theorem of the Mean we can find, for any $k\geqq\bar{k}$, $k\in K$, a point $u_k=x_k+\omega_k(\alpha_k/\sigma)d_k$, with $\omega_k\in(0,1)$, such that:

$$g'(u_k)d_k\geqq\gamma g_k' d_k. \tag{16}$$

Let now $\{x_k\}_{K_1}\subset\{x_k\}_K$ be a subsequence such that:

$$\lim_{k\to\infty, k\in K_1}x_k=\bar{x},\qquad \lim_{k\to\infty, k\in K_1}\frac{d_k}{\|d_k\|}=\bar{d}.$$

By (14), $u_k\to\bar{x}$ for $k\to\infty$, $k\in K_1$, so that, dividing both members of (16) by $\|d_k\|$ and taking limits we obtain:

$$(1-\gamma)g'(\bar{x})\bar{d}\geqq 0.$$

Since $1-\gamma>0$ and $g_k' d_k<0$ for all k, we have:

$$g'(\bar{x})\bar{d}=0$$

which implies, by (ii), $g(\bar{x})=0$, and this completes the proof of (a).

As regards (b), assume that there exists a limit point $\bar{x}$ which is a local maximum of f. By (12) there must exist a subsequence $\{x_{l(k)}\}_K\subset\{x_{l(k)}\}$ converging to $\bar{x}$. On the other hand, recalling that $\{f(x_{l(k)})\}$ is nonincreasing and admits a limit, we have: $\lim_{k\to\infty}f(x_{l(k)})=f(\bar{x})$ and $f(x_{l(k)})\geqq f(\bar{x})$ for all k. Moreover, by (6) we have:

$$f(x_{l(k')})<f(x_{l(k)})\quad\text{for } k'\geqq k+M,$$

so that, for sufficiently large $k\in K$, we can find in any neighborhood of $\bar{x}$ a point $x_{l(k)}$ such that $f(x_{l(k)})>f(\bar{x})$. This contradicts the assumption that $\bar{x}$ is a local maximum.

The last assertion of the theorem follows from known results [1, p. 476] taking (14) into account. □

Remark. We observe that the nonmonotone steplength selection rule considered here in connection with Armijo's method could also be defined in association with other acceptability criteria for the stepsize, such as the so-called Goldstein's conditions [11, p. 490]. □

The preceding theorem can be easily specialized to Newton's method, or more generally to algorithms of the Newton class, by imposing suitable conditions which ensure satisfaction of (2) and (3). In fact, let the search direction be defined by:

$$d_k = -B_k^{-1} g_k$$

and assume that $\{B_k\}$ is a sequence of symmetric positive definite matrices with uniformly bounded eigenvalues $\lambda_i(B_k)$, i.e. that there exist λ, Λ such that, for all k:

$$0 < \lambda \leqq \lambda_i(B_k) \leqq \Lambda.$$

Then

$$g_k' d_k \leqq -\frac{1}{\Lambda}\|g_k\|^2, \qquad \|d_k\| \leqq \frac{1}{\lambda}\|g_k\|.$$

Also note that if we set $a = 1$ in condition (iii) and assume that the sequence $\{x_k\}$ converges to a strong local minimum x^* and that:

$$\lim_{k\to\infty} \frac{\|[B_k^{-1} - H^{-1}(x^*)]g_k\|}{\|g_k\|} = 0$$

then, for sufficiently large k, $\alpha_k = 1$ and $\{x_k\}$ converges superlinearly. This follows directly from a known result established for Armijo's rule (see e.g. [2, p. 36]).

4. Numerical results. In this section we report the numerical results obtained for a set of standard test problems, by means of the following algorithm, which constitutes a modified version of Newton's method and uses the acceptability criterion for the stepsize defined in § 2.

ALGORITHM MODEL.

Data: x_0, integer $M \geqq 0$, $c_1 > 0$, $c_2 > 0$, $\gamma \in (0, 1)$, $\sigma \in (0, 1)$.

Step 1. Set $k = 0$, $m(0) = 0$ and compute $f_0 \triangleq f(x_0)$.

Step 2. Compute g_k; if $g_k = 0$ stop, else compute H_k; if H_k is singular, set $d_k = -g_k$, $m(k) = 0$ and go to Step 5.

Step 3. Compute $d_k = -H_k^{-1} g_k$. If $|g_k' d_k| < c_1\|g_k\|^2$ or $\|d_k\| > c_2\|g_k\|$, set $d_k = -g_k$, $m(k) = 0$ and go to Step 5.

Step 4. If $g_k' d_k > 0$, set $d_k = -d_k$.

Step 5. Set $\alpha = 1$.

Step 6. Compute $f_\alpha = f(x_k + \alpha d_k)$. If

$$f_\alpha \leqq \max_{0 \leqq j \leqq m(k)} [f_{k-j}] + \gamma \alpha g_k' d_k$$

set $f_{k+1} = f_\alpha$, $x_{k+1} = x_k + \alpha d_k$, $k = k + 1$, $m(k) \leqq \min[m(k-1)+1, M]$ and go to Step 2.

Step 7. Set $\alpha = \sigma\alpha$ and go to Step 6.

Note that the algorithm cannot cycle indefinitely between Steps 6 and 7. In fact, since $g_k' d_k < 0$ and $\gamma < 1$, we must have for sufficiently small values of α:

$$f_\alpha \leqq f_k + \gamma \alpha g_k' d_k \leqq \max_{0 \leqq j \leqq m(k)} [f_{k-j}] + \gamma \alpha g_k' d_k.$$

Typical values for the parameters are: $c_1 = 10^{-5}$, $c_2 = 10^5$, $\gamma = 10^{-3}$, $\sigma = 0.5$.

It can be easily verified that the sequence produced by the algorithm satisfies conditions (2), (3), (4).

The algorithm has been tested on the following set of problems.

Problem 1. *Extended Rosenbrock function* [13].

$$f(x)=\sum_{i=1}^{n-1}[100(x_{i+1}-x_i^2)^2+(1-x_i)^2],$$

$$x_0=[-1.2, 1, \cdots, -1.2, 1]', \quad x^*=[1, 1, \cdots, 1, 1]', \quad f(x^*)=0.$$

Problem 2. *Wood function* [4].

$$f(x)=100(x_1^2-x_2)^2+(x_1-1)^2+(x_3-1)^2+90(x_3^2-x_4)^2$$
$$+10.1[(x_2-1)^2+(x_4-1)^2]+19.8(x_2-1)(x_4-1),$$

$$x_0=[-3, -1, -3, -1]', \quad x^*=[1, 1, 1, 1]', \quad f(x^*)=0.$$

Problem 3. *Powell singular function* [12].

$$f(x)=(x_1+10x_2)^2+5(x_3-x_4)^2+(x_2-2x_3)^4+10(x_1-x_4)^4,$$

$$x_0=[3, -1, 0, 1]', \quad x^*=[0, 0, 0, 0]', \quad f(x^*)=0.$$

Problem 4. *Cube* [9].

$$f(x)=100(x_2-x_1^3)^2+(1-x_1)^2,$$

$$x_0=[-1.2, -1]', \quad x^*=[1, 1]', \quad f(x^*)=0.$$

Problem 5. *Trigonometric function* [14].

$$f(x)=\sum_{i=1}^{n}\left[n+i(1-\cos x_i)-\sin x_i-\sum_{j=1}^{n}\cos x_j\right]^2,$$

$$x_0=\left[\frac{1}{5n}, \cdots, \frac{1}{5n}\right]', \qquad x^*=[0, \cdots, 0]', \quad f(x^*)=0.$$

Problem 6. *Helical valley function* [7].

$$f(x)=100[(x_3-10\theta)^2+(\sqrt{x_1^2+x_2^2}-1)^2]+x_3^2$$

where

$$2\pi\theta=\begin{cases}\arctan(x_2/x_1) & \text{if } x_1>0,\\ \pi+\arctan(x_2/x_1) & \text{if } x_1<0,\end{cases}$$

$$x_0=[-1, 0, 0]', \quad x^*=[1, 0, 0]', \quad f(x^*)=0.$$

The algorithm has been employed by assuming at Step 6:

$$m(k)=0 \quad \text{for } k<N,$$
$$m(k)=\min[m(k-1)+1, M] \quad \text{for } k\geqq N,$$

where $N\geqq 1$ is the number of initial steps in which standard Armijo's rule is used. For $M=0$ the algorithm reduces to Armijo's method for all k.

The numerical experiments have been performed for various values of N and M.

In Table 2 the results obtained for Problems 1-5 with $N=1$ and $M=10$ are compared with those obtained in the case $M=0$.

In particular we report, for each problem, the number n of variables, the number n_l of line searches, the number n_f of function evaluations and the value $f(\hat{x})$ of the objective function at the solution found $\hat{x}$.

TABLE 2
Results for Problems 1-5.

Problem	n	Nonmonotone line search ($N=1, M=10$) n_l	n_f	$f(\hat{x})$	Armijo's line search ($M=0$) n_l	n_f	$f(\hat{x})$
1. Rosenbrock	2	12	17	$<10^{-38}$	22	30	$<10^{-38}$
	10	30	31	$<10^{-38}$	39	47	$<10^{-38}$
	20	44	45	$<10^{-38}$	52	61	$<10^{-38}$
2. Wood	4	31	35	$<10^{-38}$	40	70	$<10^{-38}$
3. Powell	4	34	35	$.2 \cdot 10^{-21}$	34	35	$.2 \cdot 10^{-21}$
4. Cube	2	11	17	$.2 \cdot 10^{-33}$	28	40	$.5 \cdot 10^{-26}$
5. Trigonometric	20	6	8	$<10^{-38}$	6	8	$<10^{-38}$
	60	6	8	$<10^{-38}$	6	8	$<10^{-38}$

The computations have been performed in double precision arithmetic on DIGITAL VAX-11/780.

We observe that in most of the problems considered, the numbers n_l and n_f for $M=10$ are considerably smaller than those required for $M=0$. In particular, for Problem 1 ($n=2$) and for Problem 4, the saving is of the order of 50%. In some cases (Problems 3 and 5) the full Newton step satisfies Armijo's conditions, so that the results obtained are independent of M.

Note that in Problem 5, both for $M=10$ and for $M=0$, it was necessary to take a starting point x_0 closer to $x^*=0$ than that suggested in [14], in order to avoid convergence to different local minima.

As regards the dependence on the parameters N and M, we experienced that in Problems 1-5 (where Newton's method converges) quite satisfactory results can be obtained for $1 \leqq N \leqq 3$ and $5 \leqq M \leqq 10$. As an example, we report, for Problem 2, in Table 3 the results obtained with $N=1$ and $M=0$, 1, 5, 10, 15, 20 and in Table 4 those obtained with $M=10$ and $N=1$, 2, 5, 10. In all cases $f(\hat{x})<10^{-38}$.

Problem 6 is an example where, starting from the given initial point, Newton's method does not converge.

In this case the choice $N=1$ still ensures convergence of our algorithm but the results are worse than those obtained with Armijo's method, especially for large values of M. This is shown in Table 5, where $N=1$, $M=0$, 1, 5, 10 and $f(\hat{x})<10^{-38}$.

However, if we adopt Armijo's method during the initial steps, the behaviour of the algorithm can be considerably improved even for relatively large values of M. In

TABLE 3
Results for Problem 2 *with* $N=1$.

($N=1$) M	(Armijo's method) 0	1	5	10	15	20
n_l	40	38	30	31	44	49
n_f	70	67	40	35	47	51

TABLE 4
Results for Problem 2 with $M = 10$.

($M=10$) N	1	2	3	5	10	(Armijo's method)
n_l	31	29	30	32	36	40
n_f	35	33	40	49	70	70

TABLE 5
Results for Problem 6 with $N = 1$.

($N=1$) M	(Armijo's method) 0	1	5	10
n_l	16	17	22	56
n_f	20	43	28	87

Table 6 we compare the results obtained for $M=10$ and $N=1$, 2, 3, 5 with those corresponding to $M=0$.

We note, in particular, that the nonmonotone line search technique is beneficial especially during the intermediate stages of the minimization process. In fact, for $N=2,3$ we obtain the best results, whereas for $N \geqq 5$ the behaviour is the same as that exhibited by Armijo's method.

TABLE 6
Results for Problem 6 with $M = 10$.

($M=10$) N	1	2	3	5	(Armijo's method)
n_l	56	13	13	16	16
n_f	87	16	16	20	20

5. Conclusions. The use of a nonmonotone line search technique in connection with Newton's method appears to be particularly valuable especially in the intermediate and in the final stages of the minimization process.

We remark that, although the algorithm employed constitutes a rough implementation of Newton's method, the results obtained compare quite favourably with those given in the literature (see e.g. [15], [6]) in correspondence to more sophisticated modifications of Newton's method. This suggests the possibility of applying a nonmonotone line search technique to more recent implementations of Newton-type methods such as the model-trust region method [6], [5], [10].

Acknowledgment. The authors would like to thank the anonymous referees for their useful suggestions.

REFERENCES

[1] L. ARMIJO, *Minimization of functions having Lipschitz-continuous first partial derivatives*, Pacific J. Math., 16 (1966), pp. 1-3.

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

[3] R. M. CHAMBERLAIN, M. J. D. POWELL, C. LEMARECHAL AND H. C. PEDERSEN, *The watchdog technique for forcing convergence in algorithms for constrained optimization*, Math. Programming Stud., 16 (1982), pp. 1-17.

[4] A. R. COLVILLE, *A comparative study of non-linear programming codes*, IBM New York Scientific Centre T.R. 320-2925, 1968.

[5] J. E. DENNIS AND R. B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[6] R. FLETCHER, *Practical Methods of Optimization, Vol. 1*, John Wiley, New York, 1980.

[7] R. FLETCHER AND M. J. D. POWELL, *A rapidly convergent descent method for minimization*, Comput. J., 6 (1963), pp. 163-168.

[8] M. HESTENES, *Conjugate Direction Methods in Optimization*, Springer-Verlag, New York, 1980.

[9] A. LEON, *A Comparison Among Eight Known Optimizing Procedures*, in Recent Advances in Optimization Techniques, A. Lavi and T. Vogl, eds., John Wiley, New York, 1966, pp. 28-46.

[10] J. J. MORÉ AND D. C. SORENSEN, *Newton's method*, in Studies in Numerical Analysis, G. H. Golub, ed., The Mathematical Association of America, Washington, DC, 1984, pp. 29-82.

[11] J. M. ORTEGA AND W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[12] M. J. D. POWELL, *An iterative method for finding stationary values of a function of several variables*, Comput. J., 5 (1962), pp. 147-151.

[13] H. H. ROSENBROCK, *An automatic method for finding the greatest or least value of a function*, Comput. J., 3 (1960), pp. 175-184.

[14] E. SPEDICATO, *Computational experience with quasi-Newton algorithms for minimization problems of moderately large size*, in Towards Global Optimization 2, L. C. W. Dixon and G. P. Szegö, eds., North-Holland, New York, 1978, pp. 209-219.

[15] M. A. WOLFE, *Numerical Methods for Unconstrained Optimization*, Van Nostrand Reinhold, New York, 1978.