JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 112, No. 2, pp. 315–330, February 2002 (©2002)

On the Nonmonotone Line Search[1,2]

Y. H. Dai[3]

Communicated by G. Di Pillo

Abstract. The technique of nonmonotone line search has received many successful applications and extensions in nonlinear optimization. This paper provides some basic analyses of the nonmonotone line search. Specifically, we analyze the nonmonotone line search methods for general nonconvex functions along different lines. The analyses are helpful in establishing the global convergence of a nonmonotone line search method under weaker conditions on the search direction. We explore also the relations between nonmonotone line search and R-linear convergence assuming that the objective function is uniformly convex. In addition, by taking the inexact Newton method as an example, we observe a numerical drawback of the original nonmonotone line search and suggest a standard Armijo line search when the nonmonotone line search condition is not satisfied by the prior trial steplength. The numerical results show the usefulness of such suggestion for the inexact Newton method.

Key Words. Unconstrained optimization, uniform convexity, Armijo line search, nonmonotone line search, R-linear convergence.

1. Introduction

The technique of nonmonotone line search was proposed first in Ref. 1 and has received many successful applications or extensions in both unconstrained optimization and constrained optimization; for example, see Refs. 2–13. Although promising in its current state of development, research

[1]This research was partly supported by China NSF Grant 19801033 and the Innovation Fund of the Chinese Academy of Sciences.

[2]The author thanks two anonymous referees for useful suggestions that improved this paper greatly.

[3]Associate Professor, State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing, China.

on the topic of nonmonotone line search is still very much in infancy, as pointed out in Ref. 12. The purpose of this paper is to provide some basic analyses of the nonmonotone line search.

We consider the unconstrained optimization problem

$$\min f(x), \qquad x \in R^n, \tag{1}$$

where f is a smooth function and where its gradient $g = \nabla f(x)$ is available. Suppose that the current approximation to the solution of (1) is x_k. If $g_k = \nabla f(x_k) \neq 0$, then a line search method defines a search direction d_k in some way, finds a steplength α_k by carrying out some line search along d_k, and computes the next approximation x_{k+1} as follows:

$$x_{k+1} = x_k + \alpha_k d_k. \tag{2}$$

In obtaining the steplength α_k, traditional line searches require the function value to decrease monotonically at every iteration, namely,

$$f(x_{k+1}) < f(x_k). \tag{3}$$

For example, see the Armijo line search (Ref. 14) and the Wolfe line search (Ref. 15). The nonmonotone line search does not impose the condition (3); as a result, it is helpful to overcome the case where the sequence of iterates follows the bottom of a curved narrow valley, a common occurrence in difficult nonlinear problems. For some optimization methods that are not one-step Q-linearly convergent in the objective function, but have some kind of convergence properties, the nonmonotone line search is especially useful in keeping the good properties of the methods, since they do not ensure a descent in the objective function at every iteration. For example, see the sequential quadratic programming method, that is known to be only two-step superlinearly convergent.

Let $0 < \lambda_1 < \lambda_2$, $\sigma \in (0, 1)$, $\delta \in (0, 1)$, and let M be a positive integer; assume that, at the kth iteration, a prior trial steplength $\bar{\alpha}_k \in (\lambda_1, \lambda_2)$ is given. The nonmonotone line search is to choose the first nonnegative integer h_k such that the steplength

$$\alpha_k = \bar{\alpha}_k \sigma^{h_k} \tag{4}$$

satisfies

$$f(x_k + \alpha_k d_k) \leq \max_{0 \leq j \leq m(k)} f(x_{k-j}) + \delta \alpha_k g_k^T d_k, \tag{5}$$

where

$$m(0) = 0 \quad \text{and} \quad 0 \leq m(k) \leq \min[m(k-1)+1, M-1], \qquad k \geq 1.$$

If $m(k) \equiv 0$, the above nonmonotone line search reduces to the Armijo line search.

In this paper, results on the nonmonotone line search are obtained from three aspects. First, we analyze the nonmonotone methods for general nonconvex functions along different lines; see Section 2. The analyses are helpful in establishing the global convergence of a nonmonotone line search method even in the absence of the sufficient descent condition (16) or the condition (17). Next, we explore the relations between nonmonotone line search and R-linear convergence assuming that f is uniformly convex (see Section 3). Then, by taking the inexact Newton method as an example, we observe a numerical drawback of the nonmonotone line search in Ref. 1 and suggest a standard Armijo line search when the condition (5) is not satisfied by the prior trial steplength $\bar{\alpha}_k$; see Section 4. Our numerical results show the usefulness of such a suggestion for the inexact Newton method. Finally, discussions and conclusions are given in Section 5.

2. Analyses of Nonmonotone Methods for General Functions

At first, we have the following lemma for the nonmonotone line search, where $f(x)$ may be a general function.

Lemma 2.1. Suppose that $f(x)$ is bounded below on R^n and that its gradient $\nabla f(x)$ is Lipschitz continuous; namely, there exists $L > 0$ such that $\|\nabla f(y) - f(z)\| \leq L\|y - z\|$, for any $y, z \in R^n$. Consider any iterative method (2), where d_k is a descent direction and α_k is obtained by the nonmonotone line search (4)–(5). Then, for any $l \geq 1$,

$$\begin{aligned} &\max_{1 \leq i \leq M} f(x_{Ml+i}) \\ &\leq \max_{1 \leq i \leq M} f(x_{M(l-1)+i}) + \delta \max_{0 \leq i \leq M-1} [\alpha_{Ml+i} g_{Ml+i}^T d_{Ml+i}]. \end{aligned} \tag{6}$$

Further, we have that

$$\sum_{l \geq 1} \min_{0 \leq i \leq M-1} \{|g_{Ml+i}^T d_{Ml+i}|, (g_{Ml+i}^T d_{Ml+i})^2 / \|d_{Ml+i}\|^2\} < +\infty. \tag{7}$$

Proof. To prove (6), it suffices to show that the following inequality holds for $j = 1, \ldots, M$:

$$f(x_{Ml+j}) \leq \max_{1 \leq i \leq M} f(x_{M(l-1)+i}) + \delta \alpha_{Ml+j-1} g_{Ml+j-1}^T d_{Ml+j-1}. \tag{8}$$

Since the line search conditions (4)–(5) imply

$$f(x_{Ml+1}) \leq \max_{0 \leq i \leq m(Ml)} f(x_{Ml-i}) + \delta \alpha_{Ml} g_{Ml}^T d_{Ml}, \tag{9}$$

it follows from this and

$$m(Ml) \leq M - 1$$

that (8) holds for $j=1$. Suppose that (8) holds for any $1 \le j \le M-1$. With the descent property of d_k, this implies

$$\max_{1 \le i \le j} f(x_{Ml+i}) \le \max_{1 \le i \le M} f(x_{M(l-1)+i}). \tag{10}$$

By the line search conditions, the induction hypothesis,

$$m(Ml+j) \le M-1,$$

and (10), we obtain

$$\begin{aligned} f(x_{Ml+j+1}) &\le \max_{0 \le i \le m(Ml+j)} f(x_{Ml+j-i}) + \delta \alpha_{Ml+j} g_{Ml+j}^T d_{Ml+j} \\ &\le \max\left\{ \max_{1 \le i \le M} f(x_{M(l-1)+i}), \max_{1 \le i \le j} f(x_{Ml+i}) \right\} \\ &\quad + \delta \alpha_{Ml+j} g_{Ml+j}^T d_{Ml+j} \\ &\le \max_{1 \le i \le M} f(x_{M(l-1)+i}) + \delta \alpha_{Ml+j} g_{Ml+j}^T d_{Ml+j}. \end{aligned}$$

Thus, (8) is also true for $j+1$. By induction, (8) holds for $1 \le j \le M$. Therefore, (6) holds.

Since $f(x)$ is bounded below, it follows that

$$\max_{1 \le i \le M} f(x_{Mj+i}) > -\infty.$$

By summing (6) over l, we can get

$$\sum_{l \ge 1} \min_{0 \le i \le M-1} [-\alpha_{Ml+i} g_{Ml+i}^T d_{Ml+i}] < +\infty. \tag{11}$$

Suppose that (5) is false for the prior trial steplength $\bar{\alpha}_k$. Then, we have

$$\begin{aligned} f(x_k + (\alpha_k/\sigma) d_k) &> \max_{0 \le j \le m(k)} f(x_{k-j}) + \delta(\alpha_k/\sigma) g_k^T d_k \\ &\ge f(x_k) + \delta(\alpha_k/\sigma) g_k^T d_k. \end{aligned} \tag{12}$$

By the mean-value theorem and the Lipschitz continuity of ∇f, we can show that

$$\begin{aligned} & f(x_k + \alpha d_k) - f(x_k) \\ &= \alpha g_k^T d_k + \int_0^\alpha [\nabla f(x_k + t d_k) - g_k]^T d_k \, dt \\ &\le \alpha g_k^T d_k + \int_0^\alpha t L \|d_k\|^2 dt \\ &= \alpha g_k^T d_k + (1/2) L \alpha^2 \|d_k\|^2 \\ &\le \delta \alpha g_k^T d_k, \qquad \text{for all } \alpha \in (0, [2(1-\delta)/L] \cdot |g_k^T d_k| / \|d_k\|^2]. \end{aligned} \tag{13}$$

It follows from (12)–(13) that

$$\alpha_k/\sigma \geq [2(1-\delta)/L]|g_k^T d_k|/\|d_k\|^2. \tag{14}$$

If (5) is true for the prior trial steplength $\bar{\alpha}_k$, then $\alpha_k = \bar{\alpha}_k \geq \lambda_1$. So, the following relation holds for some constant $c > 0$:

$$\alpha_k \geq \min\{\lambda_1, c|g_k^T d_k|/\|d_k\|^2\}, \tag{15}$$

which with (11) implies the truth of (7). □

From relation (6), we see that, for any nonmonotone line search method, the sequence $\{\max_{1 \leq i \leq M} f(x_{Ml+i})\}$ is strictly monotonically decreasing and the decrease can be estimated. As a result, this relation plays an important role in the R-linear convergence analyses in Section 3. Since the relation (7) concerns only the search directions $\{d_k\}$, it is useful in analyzing the convergence properties of a nonmonotone line search method, as will be shown subsequently.

Suppose that there exist constants c_1 and c_2 such that, for all k,

$$g_k^T d_k \leq -c_1\|g_k\|^2, \tag{16}$$

$$\|d_k\| \leq c_2\|g_k\|. \tag{17}$$

We can show easily the following convergence result via Lemma 2.1.

Theorem 2.1. Suppose that $f(x)$ is bounded below on R^n and that its gradient $\nabla f(x)$ is Lipschitz continuous. Consider any iterative method (2), where d_k satisfies (16)–(17) and α_k is obtained by the nonmonotone line search (4)–(5). Then, there exists a constant c_3 such that

$$\|g_{k+1}\| \leq c_3\|g_k\|, \qquad \text{for all } k. \tag{18}$$

Further, we have that

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{19}$$

Proof. Noting that $\alpha_k \leq \lambda_2$, by this, (2), and (17) we have that

$$\|x_{k+1} - x_k\| \leq \alpha_k\|d_k\| \leq c_2\lambda_2\|g_k\|; \tag{20}$$

this and the Lipschitz continuity of $\nabla f(x)$ yield

$$\|g_{k+1} - g_k\| \leq c_2\lambda_2 L\|g_k\|. \tag{21}$$

Thus, (18) holds with

$$c_3 = 1 + c_2\lambda_2 L.$$

In addition, it follows by (7), (16), (17) that

$$\lim_{l\to\infty} \|g_{Ml+\phi(l)}\| = 0, \tag{22}$$

where

$$0 \le \phi(l) \le M-1.$$

By (18), we have that

$$\|g_{M(l+1)+i}\| \le c_3^{2M} \|g_{Ml+\phi(l)}\|, \qquad \text{for } i = 0, \ldots, M-1. \tag{23}$$

Therefore, it follows from (22)–(23) that (19) holds. □

In the case where the level set

$$\mathscr{L} = \{x \in R^n : f(x) \le f(x_1)\}$$

is bounded, the relation (19) implies that every cluster point of $\{x_k\}$ is a stationary point of $f(x)$. It follows from (19)–(20) that $x_{k+1} - x_k$ tends to zero as $k \to \infty$. This shows that, if the number of the stationary points of f in $\mathscr{L}$ is finite, the sequence $\{x_k\}$ converges.

Thus, we prove again the global convergence of nonmonotone line search methods for general functions under the conditions (16)–(17). Compared with the one in Ref. 1, here the proof is quite different. The Lipschitz continuity of the gradient is required in our proof. However, with this additional assumption, the useful relation (7) can be established, and hence it makes possible to weaken the conditions (16) and/or (17) on the search direction d_k. For example, if $\|d_k\|^2$ increases at most linearly with k, namely, it there exist some positive constants β and γ such that

$$\|d_k\|^2 \le \beta + \gamma k, \qquad \text{for all } k, \tag{24}$$

and if the sufficient condition (16) still holds, then one can prove that the method converges in the sense that

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{25}$$

Theorem 2.2. Suppose that $f(x)$ is bounded below on R^n and that its gradient $\nabla f(x)$ is Lipschitz continuous. Consider any iterative method (2), where d_k satisfies (16) and (24), and where α_k is obtained by the nonmonotone line search (4)–(5). Then, the method yields the convergence relation (25).

Proof. We proceed by contradiction, assuming that the relation (25) does not hold. Then, there exists some constant $\tau > 0$ such that

$$\|g_k\| \ge \tau, \qquad \text{for all } k \ge 1. \tag{26}$$

This and (16) imply

$$|g_k^T d_k| \geq c_1 \tau^2. \tag{27}$$

Then, it follows from (27) and (24) that

$$\begin{aligned}
&\sum_{l \geq 1} \min_{0 \leq i \leq M-1} \{|g_{Ml+i}^T d_{Ml+i}|, (g_{Ml+i}^T d_{Ml+i})^2 / \|d_{Ml+i}\|^2\} \\
&\geq \sum_{l \geq 1} \min_{0 \leq i \leq M-1} \{c_1 \tau^2, c_1^2 \tau^4 / [\beta + \gamma(Ml + i)]\} \\
&\geq \sum_{l \geq 1} \min\{c_1 \tau^2, c_1^2 \tau^4 / [\beta + \gamma(Ml + M - 1)]\} \\
&= +\infty.
\end{aligned} \tag{28}$$

The above relation contradicts (7). The contradiction shows the truth of (25). □

Reference 16 relaxed further the sufficient descent condition (16). In Ref. 16, a nonmonotone conjugate gradient algorithm for solving (1) was proposed, for which only the following relations are shown for all k:

$$-g_k^T d_k \geq \min\{\beta, \gamma / \sqrt{k}\}, \tag{29}$$

$$(g_k^T d_k)^2 / \|d_k\|^2 \geq \tau / k, \tag{30}$$

where β, γ, τ are some positive constants. It is obvious that the above relations are weaker than the conditions (16) and (17), respectively. However, similarly to Theorem 2.2, we can prove without difficulty via the relations (7), (29), (30) that the algorithm yields the convergence relation (25).

3. Relations between Nonmonotone Line Search and *R*-Linear Convergence

In this section, we explore the relations between nonmonotone line search and R-linear convergence assuming that f is uniformly convex and d_k satisfies (16)–(17). More precisely, we assume that there exist positive constants η_1 and η_2 such that the following relation holds for any $y, z \in R^n$:

$$\eta_1 \|y - z\|^2 \leq (y - z)^T [\nabla f(y) - \nabla f(z)] \leq \eta_2 \|y - z\|^2. \tag{31}$$

In this case, let x^* be the unique minimizer of f, and let γ and $\bar{\gamma}$ be positive constants such that

$$\gamma \|\nabla f(x)\|^2 \leq f(x) - f(x^*) \leq \bar{\gamma} \|\nabla f(x)\|^2, \qquad \text{for all } x \in R^n. \tag{32}$$

The following theorem shows that any iterative method using the nonmonotone line search is R-linear convergent for uniformly convex functions.

Theorem 3.1. Suppose that $f(x)$ is a smooth and uniformly convex function. Consider any iterative method (2), where d_k satisfies (16)–(17) and α_k is obtained by the nonmonotone line search (4)–(5). Then, there exist constants $c_4 > 0$ and $c_5 \in (0, 1)$ such that

$$f(x_k) - f(x^*) \leq c_4 c_5^k [f(x_1) - f(x^*)]. \tag{33}$$

Proof. First, we have by (18) and (32) that

$$f(x_{k+1}) - f(x^*) \leq b[f(x_k) - f(x^*)], \qquad \text{for all } k \geq 1, \tag{34}$$

where

$$b = c_3^2 \bar{\gamma}^2 / \gamma^2 > 1.$$

For any $l \geq 0$, let $\psi(l)$ be any index in $[Ml+1, M(l+1)]$ for which

$$f(x_{\psi(l)}) = \max_{1 \leq i \leq M} f(x_{Ml+i}). \tag{35}$$

By the definition of $\psi(l)$, and by (6) and (15)–(17), we can get that

$$f(x_{\psi(l)}) \leq f(x_{\psi(l-1)}) - c_6 \min_{0 \leq i \leq M-1} \|g_{Ml+i}\|^2, \tag{36}$$

where

$$c_6 = \delta \min\{\lambda c_1, c_1^2 c / c_2^2\}$$

is a positive constant. Let $\xi(l)$ and $\zeta(l)$ be any indices in $[Ml+1, M(l+2)]$ for which

$$\|g_{\xi(l)}\| = \min_{1 \leq i \leq 2M} \|g_{Ml+i}\|, \tag{37}$$

$$f(x_{\zeta(l)}) = \max_{1 \leq i \leq 2M} f(x_{Ml+i}), \tag{38}$$

and denote by c_7 the constant

$$c_7 = [c_6 + \bar{\gamma} c_3^{8M}]^{-1}. \tag{39}$$

Now, we define an infinite subsequence $\{k_i : i \geq 0\} \subset \{1, 2, 3, \ldots\}$ as follows. Pick $k_0 = \psi(0)$. Suppose that $k_i = \psi(\bar{l})$ is chosen for some $\bar{l}$. If

$$\|g_{\xi(\bar{l}+1)}\|^2 \leq c_7 [f(x_{\psi(\bar{l})}) - f(x^*)], \tag{40}$$

we define

$$k_{i+1} = \zeta(\bar{l}+1);$$

otherwise, we set

$$k_{i+1} = \psi(\bar{l}+3).$$

For the subsequence $\{k_i\}$ defined as above, it is obvious that

$$k_{i+1} - k_i \leq 4M. \tag{41}$$

In addition, there must exist a constant $c_8 \in (0, 1)$ such that

$$f(x_{k_{i+1}}) - f(x^*) \leq c_8[f(x_{k_i}) - f(x^*)], \qquad \text{for all } i \geq 1. \tag{42}$$

In fact, if (40) holds, it follows by (32), (18), the definition of k_{i+1}, and (40)–(41) that

$$\begin{aligned} f(x_{k_{i+1}}) - f(x^*) &\leq \bar{\gamma}\|g_{k_{i+1}}\|^2 \\ &\leq \bar{\gamma} c_3^{8M} \|g_{\xi(\bar{l})}\|^2 \\ &< \bar{\gamma} c_3^{8M} c_7 [f(x_{k_i}) - f(x^*)]. \end{aligned} \tag{43}$$

If

$$\|g_{\xi(\bar{l}+1)}\|^2 > c_7[f(x_{\psi(\bar{l})}) - f(x^*)], \tag{44}$$

we have by this and (36)–(37) that

$$f(x_{k_{i+1}}) - f(x^*) \leq (1 - c_6 c_7)[f(x_{k_i}) - f(x^*)]. \tag{45}$$

Therefore by (45), (43), and the choice (39) of c_7, we know that (42) holds with

$$c_8 = 1 - c_6 c_7.$$

For any $k \geq 1$, assume that

$$k \in [k_i, k_{i+1}), \qquad \text{for some } i.$$

Then, we have from (41) that

$$k - k_i \leq 4M. \tag{46}$$

By (41), we can get also

$$k_i \leq k_0 + 4Mi; \tag{47}$$

Combination of (47) with $1 \leq k_0 \leq M$ and (46) shows that

$$\begin{aligned} i &\leq (k_i - k_0)/4M \\ &\geq (k - 4M - k_0)/4M \\ &= k/4M - 5/4. \end{aligned} \tag{48}$$

Thus, by (34), (42), (46), (48), we obtain

$$\begin{aligned} f(x_k) - f(x^*) &\leq b^{k-k_i}[f(x_{k_i}) - f(x^*)] \\ &\leq b^{4M} c_8^i [f(x_{k_0}) - f(x^*)] \\ &\leq b^{4M} c_8^{k/4M - 5/4} [f(\dot{x}_{k_0}) - f(x^*)] \\ &\leq b^{5M} c_8^{k/4M - 5/4} [f(x_1) - f(x^*)]. \end{aligned} \tag{49}$$

Therefore, (33) holds with

$$c_4 = b^{5M} c_8^{-5/4}, \qquad c_5 = c_8^{1/4M}. \qquad \square$$

From the above proof, we see that, when a nonmonotone line search method is applied to minimize a uniformly convex function, there must exist a Q-linear convergent subsequence $\{x_{k_i}: i \geq 0\}$ satisfying

$$k_{i+1} - k_i \leq M_1, \qquad \text{for all } i \geq 0, \tag{50}$$

where M_1 is some fixed positive integer.

Generally, a R-linearly or R-superlinearly method is not necessarily such that the nonmonotone line search conditions are satisfied. For example, we consider the gradient method $d_k = -g_k$ for the 1-dimensional function

$$f(x) = (1/2)x^2, \qquad x \in R^1. \tag{51}$$

If $x_1 \neq 0$ and α_k is chosen to be

$$\alpha_k = \begin{cases} 1 - 2^{-k}, & \text{if } k = i^2 \text{ for some integer } i, \\ 2, & \text{otherwise,} \end{cases} \tag{52}$$

it is easy to show that $\{x_k\}$ is R-superlinearly convergent to the point $x^* = 0$, but that condition (5) does not hold for any fixed M. Nevertheless, if there exists some Q-linearly subsequence $\{x_{k_i}\}$ satisfying

$$k_{i+1} - k_i \leq M_1, \qquad \text{for some } M_1,$$

it can be shown that condition (5) must hold for the given steplengths.

Theorem 3.2. Suppose that $f(x)$ is a smooth, uniformly convex function, and that x_1 is given. Consider some iterative method (2), where d_k satisfies (16)–(17) and α_k satisfies

$$\lambda_2 \geq \alpha_k \geq \min\{\lambda_1, c|g_k^T d_k|/\|d_k\|^2\}. \tag{53}$$

Assume that there exists an infinite subsequence $\{k_i: i \geq 0\}$ satisfying (42) and (50), where M_1 is some integer and $c_8 \in (0, 1)$. Then, for any $\delta \in (0, 1)$, there must exist an integer M and a sequence $m(k)$ satisfying

$$m(0) = 0 \quad \text{and} \quad m(k) \leq \min\{m(k-1) + 1, M - 1\}$$

such that

$$f(x_k + \alpha_k d_k) \leq \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \delta \alpha_k g_k^T d_k, \qquad \text{for all } k \geq M. \tag{54}$$

Proof. Assume without loss of generality that

$$k_0 \leq M_1; \tag{55}$$

otherwise, set

$$M_1 = \max\{k_0, M_1\}.$$

Note that (34) still holds due to (17) and (53). Similarly to the third inequality in (49), we can prove by (34), (42), (50) that there exist constants $c_9 > 0$ and $c_{10} \in (0, 1)$ such that

$$f(x_{k_i+j}) - f(x^*) \leq c_9 c_{10}^j [f(x_{k_i}) - f(x^*)], \qquad \text{for all } i \geq 0 \text{ and } j \geq 1. \tag{56}$$

In addition, it follows from (17) that

$$g_k^T d_k \geq -c_2 \|g_k\|^2, \tag{57}$$

which with (16), (53), (32) implies

$$\alpha_k g_k^T d_k \geq -c_{11}(f(x_k) - f(x^*)), \qquad \text{for all } k, \tag{58}$$

where

$$c_{11} = c_2 \lambda_2 / \gamma.$$

Then, for any $\delta \in (0, 1)$, since $c_{10} \in (0, 1)$, there must exist some integer M_2 such that

$$c_9 c_{10}^j [c_{10} + \delta c_{11}] \leq 1, \qquad \text{for all } j \geq M_2. \tag{59}$$

Let

$$M = M_1 + M_2.$$

For any $k \geq M$, we know by (50) and (55) that there must be some integer $\hat{i} \geq 0$ such that

$$k_{\hat{i}} \in [k - M + 1, k - M_2]. \tag{60}$$

From (60), it follows that

$$M_2 \leq k - k_{\hat{i}} \leq M - 1. \tag{61}$$

By (56) and (58)–(60), we obtain

$$\begin{aligned} f(x_{k+1}) - f(x^*) &\leq c_9 c_{10}^{M_2+1} [f(x_{k_{\hat{i}}}) - f(x^*)] \\ &\leq (1 - \delta c_9 c_{11} c_{10}^{M_2}) [f(x_{k_{\hat{i}}}) - f(x^*)] \\ &\leq f(x_{k_{\hat{i}}}) - f(x^*) + \delta \alpha_k g_k^T d_k. \end{aligned} \tag{62}$$

Therefore, if we further choose

$$m(k) = \min\{k, M - 1\},$$

the relation (54) must hold due to (61)–(62). This completes our proof. □

Theorem 3.2 tells us for what optimization methods the nonmonotone line search could be applied, whereas the good properties of the methods may be still preserved. For example, see the sequential quadratic programming method, which is known to be two-step superlinearly convergent. Another example is the Barzilai and Borwein gradient method (Ref. 17), which is of the form

$$x_{k+1} = x_k - [\|x_k - x_{k-1}\|^2/(g_k - g_{k-1})^T(x_k - x_{k-1})]\, g_k, \tag{63}$$

where x_1, x_2 are given initial points. For convex quadratic functions, such a method is shown to be R-superlinearly convergent if the dimension is $n = 2$ (see Ref. 17). For convex quadratic functions of any dimension n, the method is shown to be R-linearly convergent (see Ref. 18); specifically, it is shown that there must exist some Q-linearly convergent subsequence $\{f(x_{k_i})\}$ satisfying

$$k_{i+1} - k_i \le M_1, \qquad \text{for some } M_1.$$

Thus, if a nonmonotone line search is applied, every point defined by the Barzilai and Borwein gradient method may be accepted by the line search as $x_k \to x^*$, where x^* satisfies $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) > 0$, as is observed in Ref. 10.

4. Modified Nonmonotone Line Search

As mentioned once in Section 1, the nonmonotone line search is helpful to overcome the case where the sequence of iterates follows the bottom of a curved narrow valley. However, in practical computations, it is always the case that users do not know where there are curved narrow valleys. For functions whose nonlinearity is not strong, one would prefer a monotone line search to a nonmonotone one. For example, consider the Brown and Dennis function. To satisfy the inequality (68), the inexact Newton method with the nonmonotone line search (4)–(5) requires 22 iterations with 301 function evaluations. In contrast, the method using the standard Armijo line search needs only 14 iterations and 90 function evaluations; see below for details.

Suppose that, for some optimization method, the prior trial steplength at the kth iteration is $\bar{\alpha}_k$. For example, in the Newton method, $\bar{\alpha}_k = 1$. Since this prior trial steplength has some good properties, it is reasonable to accept it relatively easily; in other words, provided that the prior trial steplength $\bar{\alpha}_k$ satisfies the condition (5), it seems reasonable for one to accept it. However, in the case where the prior trial steplength $\bar{\alpha}_k$ does not satisfy (5), it seems more reasonable to require the second candidate steplength $\hat{\alpha}_k$ to satisfy the standard Armijo condition

$$f(x_k + \hat{\alpha}_k d_k) \le f(x_k) + \delta \hat{\alpha}_k g_k^T d_k, \tag{64}$$

since $\hat{\alpha}_k$ does not possess the good properties that the prior steplength $\bar{\alpha}_k$ has.

The new nonmonotone strategy can be expressed as follows:

set $\alpha_k = \bar{\alpha}_k$, if (5) holds for $\bar{\alpha}_k$; otherwise, obtain α_k by doing a standard Armijo line search along d_k.

Since the Armijo condition is stronger than (5), the results obtained in Sections 2–3 still hold for the above modified nonmonotone line search. In addition, it is obvious that the modified nonmonotone line search can also to a great extent overcome the case where the sequence of iterates follows the bottom of some curved narrow valley. Further, for some optimization methods, the modified nonmonotone line search is superior to the old nonmonotone line search, as will be shown below.

To show the usefulness of the new nonmonotone line search, we tested the inexact Newton method,

$$d_k = -H_k^{-1} g_k, \tag{65}$$

where H_k is an approximation to the Hessian of f at the current point x_k. In our tests, we used a central difference technique to obtain H_k, namely,

$$(H_k)_i = [\nabla f(x_k + \gamma e_i) - \nabla f(x_k - \gamma e_i)]/2\gamma. \tag{66}$$

Here $(H_k)_i$ and e_i mean the ith column of H_k and the unit matrix I_n, and γ is the difference stepsize given by

$$\gamma = \min\{10^{-3}, \max\{10^{-3}\|g_k\|, 10^{-6}\}\}. \tag{67}$$

This choice for the difference stepsize is typical, for the numerical results with smaller γ provide the same conclusion. We also used the safeguard in Ref. 1; namely, we set

$$d_k = -g_k, \text{ if } H_k \text{ is singular, or } |g_k^T d_k| < 10^{-5}\|g_k\|^2, \text{ or } \|d_k\| > 10^5\|g_k\|.$$

In addition,

$$d_k = -d_k, \qquad \text{if } g_k^T d_k > 0.$$

The initial guess $\bar{\alpha}_k$ is set to 1, and the constants δ and σ are 10^{-3} and 0.5, respectively. The value of M is set to 10 and

$$m(k) = \min[m(k-1), M].$$

We tested the above algorithm using the two nonmonotone line searches with double precision in an SGI Indigo workstation. The codes were written in the FORTRAN language. Our test problems and the initial

Table 1. Numerical comparisons.

		Method					
		Armijo		Old		New	
Problem	n	N_f	N_g	N_f	N_g	N_f	N_g
MGH5	2	8	16	19	27	19	27
MGH11	3	23	38	32	41	22	35
MGH14	4	38	55	29	32	34	54
MGH16	4	14	90	22	301	12	85
MGH20	9	12	13	12	13	12	13
MGH21	16	21	29	11	16	16	22
MGH21	100	21	29	11	16	16	22
MGH23	8	34	43	22	23	22	23
MGH23	100	36	106	48	205	31	98
MGH23	200	62	143	>999	>999	55	136
MGH24	3	31	39	11	12	11	12
MGH24	20	50	63	33	34	33	34
MGH25	20	5	76	>999	>999	5	76
MGH25	50	11	254	>999	>999	11	254
MGH26	20	7	12	9	13	9	13
MGH26	50	13	35	12	23	15	35
MGH26	100	36	80	20	58	20	44
MGH35	8	7	11	8	11	7	11
MGH35	20	17	30	28	46	18	26

points used are drawn from Ref. 19. For each problem, the limiting number of function evaluations is set to 999, and the stopping condition is

$$\|g_k\| \leq 10^{-6}. \tag{68}$$

We report our numerical results in Table 1, where "Armijo", "Old", and "New" stand for the Armijo line search, the nonmonotone line search (5), and the modified nonmonotone line search. The symbols n, N_f, N_g mean the dimension of the problem, the number of iterations, and the number of function evaluations.

The unconstrained optimization problems are numbered in the same way as in Ref. 19. For example, MGH5 means Problem 5 in Ref. 19. From Table 1, we can see that, for some problems, both the old nonmonotone line searchand the modified nonmonotone line search require fewer iterations and function evaluations than the Armijo line search. However, for some problems, the old nonmonotone line search performs much worse than the Armijo line search, whereas the modified nonmonotone line search performs as well or better than the Armijo line search. Therefore, our numerical results show that the use of the modified nonmonotone line search in the inexact Newton method is superior to that of the old nonmonotone line search.

5. Discussion and Conclusions

In this paper, we have provided some basic analyses of the nonmonotone line search in Ref. 1. First, we analyzed the nonmonotone methods for general nonconvex functions along different lines. The analyses are helpful in establishing the global convergence of a nonmonotone line search method even in the absence of the sufficient descent condition (16) or the condition (17). Next, we have explored the relations between nonmonotone line search and R-linear convergence in the case where f is uniformly convex. From Theorem 3.2, we know for what optimization methods the nonmonotone line search could be applied, while preserving the good properties of the methods. Then, by taking the inexact Newton method as an example, we have observed a numerical drawback of the nonmonotone line search in Ref. 1 and suggested a standard Armijo line search when the condition (5) is not satisfied by the prior trial steplength $\bar{\alpha}_k$. Numerical results have been reported, which showed that the use of the modified nonmonotone line search in the inexact Newton method is superior on average to that of the original nonmonotone line search.

We should say that our numerical experiments for the modified nonmonotone line search are limited. For some optimization methods, if the nonmonotone line search condition (5) is not satisfied by the prior trial steplength, it may be better to require the second candidate steplength to meet (5) with a relatively small $M > 0$. Anyway, we believe that both the theoretical and numerical analyses will be helpful in future studies on the technique of nonmonotone line search.

References

1. 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.
2. Bonnans, J. F., Panier, E., Tits, A., and Zhou, J. L., *Avoiding the Maratos Effect by Means of a Nonmonotone Line Search, II: Inequality Constrained Problems—Feasible Iterates*, SIAM Journal on Numerical Analysis, Vol. 29, pp. 1187–1202, 1992.
3. Deng, N. Y., Xiao, Y., and Zhou, F. J., *Nonmonotonic Trust-Region Algorithm*, Journal of Optimization Theory and Applications, Vol. 26, pp. 259–285, 1993.
4. Grippo, L., Lampariello, F., and Lucidi, S., *A Truncated Newton Method with Nonmonotone Line Search for Unconstrained Optimization*, Journal of Optimization Theory and Applications, Vol. 60, pp. 401–419, 1989.

5. GRIPPO, L., LAMPARIELLO, F., and LUCIDI, S., *A Class of Nonmonotone Stabilization Methods in Unconstrained Optimization*, Numerische Mathematik, Vol. 59, pp. 779–805, 1991.
6. KE, X., and HAN, J., *A Nonmonotone Trust Region Algorithm for Equality Constrained Optimization*, Science in China, Vol. 38A, pp. 683–695, 1995.
7. KE, X., LIU, G., and XU, D., *A Nonmonotone Trust-Region Algorithm for Unconstrained Optimization*, Chinese Science Bulletin, Vol. 41, pp. 197–201, 1996.
8. LUCIDI, S., ROCHETICH, F., and ROMA, M., *Curvilinear Stabilization Techniques for Truncated Newton Methods in Large-Scale Unconstrained Optimization*, SIAM Journal on Optimization, Vol. 8, pp. 916–939, 1998.
9. PANIER, E., and TITS, A., *Avoiding the Maratos Effect by Means of a Nonmonotone Line Search, I: General Constrained Problems*, SIAM Journal on Numerical Analysis, Vol. 28, pp. 1183–1195, 1991.
10. RAYDAN, M., *The Barzilai and Borwein Gradient Method for the Large-Scale Unconstrained Minimization Problem*, SIAM Journal on Optimization, Vol. 7, pp. 26–33, 1997.
11. TOINT, P. L., *An Assessment of Nonmonotone Line Search Techniques for Unconstrained Optimization*, SIAM Journal on Scientific Computing, Vol. 17, pp. 725–739, 1996.
12. TOINT, P. L., *A Nonmonotone Trust-Region Algorithm for Nonlinear Optimization Subject to Convex Constraints*, Mathematical Programming, Vol. 77, pp. 69–94, 1997.
13. ZHOU, J. L., and TITS, A., *Nonmonotone Line Search for Minimax Problems*, Journal of Optimization Theory and Applications, Vol. 76, pp. 455–476, 1993.
14. ARMIJO, L., *Minimization of Functions Having Lipschitz Continuous First Partial Derivatives*, Pacific Journal of Mathematics, Vol. 16, pp. 1–3, 1966.
15. WOLFE, P., *Convergence Conditions for Ascent Methods*, SIAM Review, Vol. 11, pp. 226–235, 1969.
16. DAI, Y. H., *A Nonmonotone Conjugate Gradient Algorithm for Unconstrained Optimization*, Research Report, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 2000 (accepted by Journal of Systems Science and Complexity).
17. BARZILAI, J., and BORWEIN, J. M., *Two-Point Stepsize Gradient Methods*, IMA Journal of Numerical Analysis, Vol. 8, pp. 141–148, 1988.
18. DAI, Y. H., and LIAO, L. Z., *R-Linear Convergence of the Barzilai and Borwein Gradient Method*, Research Report 99-039, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1999 (accepted by IMA Journal of Numerical Analysis).
19. MORÉ, J. J., GARBOW, B. S., and HILLSTROM, K. E., *Testing Unconstrained Optimization Software*, ACM Transactions on Mathematical Software, Vol. 7, pp. 17–41, 1981.