Algorithmic Operations Research Vol.4 (2009) 70–75

General Form of a Nonmonotone Line Search Technique for Unconstrained Optimization

Zhensheng Yu and Ji Lin

College of Science, University of Shanghai for Science and Technology, Shanghai, 200093, PR China

Abstract

By using the forcing function, we propose a general form of nonmonotone line search technique for unconstrained optimization. The technique includes some well known nonmonotone line search as special cases while independent on the nonmonotone parameter case. We establish the global convergence of the method under weak conditions and we report numerical test results with a modified BFGS method to show the effectiveness of the proposed method.

Key words: Unconstrained optimization; Nonmonotone F-rule; Global convergence; Modified BFGS method.

1. Introduction

In this paper, we consider the following unconstrained optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ is a continuously differentiable function with gradient function $g(x)=\nabla f(x)$.

Line search method is one of the most well known methods for solving (1): for a given x_k, the line search generate the next point by:

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

where $\alpha_k>0$ is a step size and d_k is a search direction.

The traditional line searches require the function value descent monotonically at every iteration, namely:

$$f(x_{k+1})\le f(x_k) \tag{2}$$

Recent research [4,9,11,16] indicates that the monotone line search technique may considerably reduce the rate of convergence when the iteration is trapped near a narrow curved valley, which can result in very short steps or zigzagging. The nonmonotone line search technique does not impose the condition (2), as a result, it is helpful to overcome this drawback. Serval numerical tests show that the nonmonotone line search technique for unconstrained optimization and constrained optimization is efficient and competitive.

The first nonmonotone line search technique was proposed by L.Grippo, F.Lampariello and S.Lucidi [6] for unconstrained optimization, where the next iteration satisfies

$$f(x_{k+1})\le \max_{0\leqslant j\leqslant min\{k-1,M\}} f(x_{k-j}), \tag{3}$$

where positive integer M is a nonmonotone parameter.

Many numerical experiments have suggested that the nonmonotone line search technique is efficient and practical for solving some nonlinear large-scale optimization problems [2,13]. However, in some cases the numerical performance is very dependent on the choice of the nonmonotone parameter (see [6,13,16,19]).

Zhang and Hager [19] proposed a new nonmonotone line search algorithm without the nonmonotone parameter, and proved global convergence under the following direction assumptions:

$$g_k^T d_k\leqslant -c_1\|g_k\|^2 \tag{4}$$

and

$$\|d_k\|\leqslant c_2\|g_k\| \tag{5}$$

where c_1 and c_2 are two positive constants. Numerical results show the new nonmonotone line search technique uses fewer function and gradient evaluations, on average, than those of the monotone or the traditional nonmonotone scheme.

⋆ This work is supported by the National Natural Science Foundation of China (No.10671126) and Shanghai Leading Academic Discipline Project(S30501)
Email: Zhensheng Yu [zhsh-yu@163.com].

Sun et al. [14] combined forcing function with the nonmonotone line search technique and proposed a general line search rule, called the nonmonotone F-rule. They proved that the nonmonotone Armijo line search rule, the nonmonotone Goldstein line search rule and the nonmonotone Wolfe line search rule are special cases of the nonmonotone F-rule (see Proposition 2.4 in [14]). To obtain the global convergence, Sun et al. required the direction d_k to satisfy the following conditions:

$$\left|\frac{-g_k^T d_k}{\|d_k\|}\right| \geqslant \sigma(\|g_k\|),\ k = 1, 2, \cdots, \qquad (6)$$

and (5), where $\sigma(\cdot)$ is a forcing function.

Noted that the condition (3) may prevent large step where the gradient is small, as in the neighborhood of saddle point and at the bottom of the valley. To overcome this drawback, Yu and Pu [18] proposed a new nonmonotone line search technique and remove the condition (3), moreover, they established the strong convergence property under conditions weaker than those of the existed traditional nonmonotone line search techniques. However, their algorithms still dependent on the nonmonotone parameter.

In this paper, we propose an algorithm model by combining the F-rule in [14] with the nonmonotone line search technique in [19]. The algorithm model includes many well known nonmonotone line searches as special cases but independence on the nonmonotone parameter. We establish the global convergence of the algorithm without the direction assumption (5) and we implement our algorithm model with a modified BFGS method [10] to show the efficiency of the algorithm.

The remainder of our paper is organized as follows. In Section 2, the algorithm model is stated. In Section 3, the global convergence is established. In Section 4, we present computational results, and give numerical comparisons.

2. Algorithm model

The following assumption is imposed throughout the paper.

Assumption 1 *$f(x)$ is bounded below on the level set $\mathcal{L} = \{x \in \mathbb{R}^n | f(x) \leqslant f(x_1)\}$, and the gradient function $g(x)$ is uniformly continuous in $\mathcal{L}$.*

Definition 1 *The function $\sigma : [0, +\infty] \to [0, +\infty]$ is a forcing function(F-function), if for any sequence $\{t_i\} \subset [0, +\infty]$*

$$\lim_{i\to\infty} \sigma(t_i) = 0 \quad \textit{implies} \quad \lim_{i\to\infty} t_i = 0. \qquad (7)$$

Now, we state our algorithm model.

Algorithm 1

Step 0: Given $x_1 \in \mathbb{R}^n$, $0 \leqslant \eta_{min} \leqslant \eta_{max} < 1$, $\delta \in (0, 1)$. Set $V_1 = f(x_1)$, $Q_1 = 1$, $k = 1$.

Step 1: If $\|g_k\|$ sufficiently small, then stop.

Step 2: Compute the search direction d_k that satisfies condition (6).

Step 3: Set $\alpha = 1$.

Step 4: If

$$f(x_k + \alpha d_k) \leqslant V_k - \sigma(-g_k^T d_k/\|d_k\|) \qquad (8)$$

does not hold, set $\alpha = \delta\alpha$, repeat Step 4.

Step 5: Define $\alpha_k = \alpha$, $x_{k+1} = x_k + \alpha_k d_k$.

Step 6: Choose $\eta_k \in [\eta_{min}, \eta_{max}]$, and set

$$Q_{k+1} = \eta_k Q_k + 1, \quad V_{k+1} = (\eta_k Q_k V_k + f_{k+1})/Q_{k+1}. \qquad (9)$$

$k = k + 1$, and go to Step 1.

Remark: Similar to proposition 2.4 in [19], it is easy to see that our nonmonotone line search contains nonmonotone Arimijo rule, the nonmonotone Goldstein rule and the nonmonotone Wolfe rule as the special case.

3. Global convergence

To establish the global convergence of Algorithm 1, we first prove two lemmas.

Lemma 1 *If $\sigma(-g_k^T d_k/\|d_k\|) \geqslant 0$ for each k, then for the iterates $\{x_k\}$ generated by Algorithm 1, we have $f_k \leqslant V_k$ for all k.*

Proof. Defining $D_k : R \to R$ by

$$D_k(t) = \frac{tV_k + f_k}{t + 1},$$

we have

$$D_k'(t) = \frac{V_{k-1} - f_k}{(t + 1)^2}.$$

Since $\sigma(-g_k^T d_k/\|d_k\|) \geqslant 0$, it follows from (8) that $f_k \leqslant V_{k-1}$, which implies that $D_k'(t) \geqslant 0$. Hence, D_k is nondecreasing. In particular, taking $t = \eta_{k-1} Q_{k-1} \geqslant 0$ gives

$$f_k = D_k(0) \leqslant D_k(\eta_{k-1} Q_{k-1}) = V_k. \ \square$$

Lemma 2 *If Q_k $(k = 1, 2, \dots)$ are generated by Algorithm 1, then*

$$Q_{k+1} \leqslant \frac{1}{1-\eta_{max}}. \tag{10}$$

Proof. From $Q_1 = 1$, $Q_{k+1} = \eta_k Q_k + 1$, and the fact that $\eta_k \in [\eta_{min}, \eta_{max}]$, we have

$$\begin{aligned} Q_{k+1} &= \eta_k(\eta_{k-1}Q_{k-1}+1)+1 \\ &= \eta_k\eta_{k-1}Q_{k-1}+\eta_k+1 \\ &\leqslant \eta_{max}^2 Q_{k-1}+\eta_{max}+1 \\ &\leqslant \cdots \\ &\leqslant \eta_{max}^k Q_1 + \eta_{max}^{k-1}+\cdots+\eta_{max}^2+\eta_{max}+1 \\ &= \sum_{j=0}^{k}\eta_{max}^j. \end{aligned}$$

Since $\eta_{max} < 1$, we deduce

$$Q_{k+1} \leqslant \sum_{j=0}^{k}\eta_{max}^j \leqslant \sum_{j=0}^{\infty}\eta_{max}^j = \frac{1}{1-\eta_{max}}. \quad \square$$

Now we establish a global convergence theorem for Algorithm 1.

Theorem 1 *Let function $f : \mathbb{R}^n \to \mathbb{R}$ satisfy Assumption 1, if the search direction d_k satisfies (6). Then the iterates $\{x_k\}$ generated by Algorithm 1 contained in the level set $\mathcal{L}$ and*

$$\lim_{k\to\infty}\|g_k\| = 0. \tag{11}$$

Proof. Combining (8) and (9),

$$\begin{aligned} V_{k+1} &= \frac{\eta_k Q_k V_k + f_{k+1}}{Q_{k+1}} \\ &\leqslant \frac{\eta_k Q_k V_k + V_k - \sigma(-g_k^T d_k/\|d_k\|)}{Q_{k+1}} \\ &= V_k - \frac{\sigma(-g_k^T d_k/\|d_k\|)}{Q_{k+1}}, \end{aligned}$$

which means $V_{k+1} \leqslant V_k$ for all k, since $f_{k+1} \leqslant V_k$, we have

$$f_{k+1} \leqslant V_k \leqslant V_{k-1} \leqslant \cdots \leqslant V_1 = f(x_1),$$

which implies the sequence $\{x_k\}$ is contained in the level set $\mathcal{L}$.

On the other hand, since f is bounded below and $f_k \leqslant V_k$ for all k, we conclude that V_k is bounded from below. Hence,

$$\sum_{k=1}^{\infty}\frac{\sigma(-g_k^T d_k/\|d_k\|)}{Q_{k+1}} < \sum_{k=1}^{\infty}(V_k - V_{k+1}) < \infty.$$

By (10),

$$(1-\eta_{max})\sum_{k=1}^{\infty}\sigma\left(\frac{-g_k^T d_k}{\|d_k\|}\right) < \infty.$$

Therefore,

$$\lim_{k\to\infty}\sigma\left(\frac{-g_k^T d_k}{\|d_k\|}\right) = 0,$$

which means from Definition 1 that

$$\lim_{k\to\infty}\frac{-g_k^T d_k}{\|d_k\|} = 0.$$

Using condition (6), we deduce

$$\lim_{k\to\infty}\sigma(\|g_k\|) = 0,$$

which implies (11) holds. $\square$

As an application of the nonmonotone line search, we consider the nonmonotone quasi-Newton method:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{12}$$

where α_k is obtained by the nonmonotone line search (8), and

$$d_k = -B_k^{-1}g_k, \tag{13}$$

where B_k is a $n \times n$ symmetric positive definite matrix and obtained by some quasi-Newton formulas.

Let θ_k be the angle of $-g_k$ and d_k, in what follows, we assume that there exists a positive constant τ such that

$$cos\theta_k = \frac{-g_k^T d_k}{\|g_k\|\|d_k\|} \geqslant \tau. \tag{14}$$

Theorem 2 *Let function $f : \mathbb{R}^n \to \mathbb{R}$ satisfy Assumption 1, consider the nonmonotone quasi-Newton method (12)-(14). Then the iterates $\{x_k\}$ generated by Algorithm 1 contained in $\mathcal{L}$ and*

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

Proof. Similar to the proof of Theorem 3.1, we have

$$\lim_{k\to\infty}\frac{-g_k^T d_k}{\|d_k\|} = 0.$$

By (14), we know that

$$\frac{-g_k^T d_k}{\|d_k\|} \geqslant \tau\|g_k\|,$$

which implies (15) hold.

4. Numerical Experiences

In what follows, we first implement our algorithm model in the context of Li and Fukushima's BFGS method [10] with slight modification. Here we combine the line search proposed by Byrd and Nocedal[3] with our nonmonotone technique.

The algorithm is described as follows.

Algorithm 2

Step 0: Given $x_1 \in \mathbb{R}^n$, $0 \leqslant \eta_{min} \leqslant \eta_{max} < 1$, $\delta \in (0,1), \tau \in (0,1), \gamma_1 \in (0,1), \gamma_2 \in (0,1)$. Choose $B_1 \in R^{n\times n}$ symmetric positive definite. Set $V_1 = f(x_1)$, $Q_1 = 1$, $k = 1$.

Step 1: If $\|g_k\| = 0$, then stop.

Step 2: Compute $d_k = -B_k^{-1} g_k$, if $-g_k^T d_k < \tau\|g_k\|\|d_k\|$, set $d_k = -g_k$.

Step 3: Set $\alpha = 1$.

Step 4: If both

$$f(x_k + \alpha d_k) \leqslant V_k + \gamma_1 \alpha g_k^T d_k. \tag{16}$$

and

$$f(x_k + \alpha d_k) \leqslant V_k - \gamma_2 \left(\frac{g_k^T d_k}{\|d_k\|}\right)^2. \tag{17}$$

does not hold, set $\alpha = \delta\alpha$, repeat Step 4.

Step 5: Define $\alpha_k = \alpha, x_{k+1} = x_k + \alpha_k d_k$.

Step 6: Choose $\eta_k \in [\eta_{min}, \eta_{max}]$, and set

$$Q_{k+1} = \eta_k Q_k + 1, \quad V_{k+1} = (\eta_k Q_k V_k + f_{k+1})/Q_{k+1}. \tag{18}$$

Step 7: Update B_k using the formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}. \tag{19}$$

where $s_k = x_{k+1} - x_k = \alpha_k d_k$ and $y_k = g_{k+1} - g_k + t_k\|g_k\| s_k$ with $t_k = 1 + max\{0, -(g_{k+1} - g_k)^T s_k / \|s_k\|^2\}$. $k = k+1$, and go to Step 1.

We compare the behavior of the Algorithm 2 with two different implementations of the nonmonotone line search technique in [6] and the method in [19]. In [6], V_k in (16) is replaced by

$$\max_{0\leqslant j \leqslant min\{k-1,M\}} f(x_{k-j}). \tag{20}$$

We choose $M = 5$ and $M = 10$.

The algorithms were coded in Matlab 7.4. The test problems were taken from Moré et al. [12], except " Strictly Convex 1 " and " Strictly Convex 2 " that are provided in [13]. The total number of the test problems is 39. For the numerical experiments we set following initial parameters: $\delta = 0.5$, $\gamma_1 = 10^{-3}$, $\gamma_2 = 10^{-3}$, $B_1 = I$. Although the best convergence results were obtained by dynamically varying η_k in Step 6, using values closer to 1 when the iterates were far from the optimum, and using values closer to 0 when the iterates were near an optimum, the numerical experiments reported here employ a fixed value $\eta_k \equiv 0.85$, which adopted by Zhang and Hager in [19]. To decide when to stop the execution of the algorithms declaring convergence we used the criterion $\|g_k\|_\infty \leqslant 10^{-6}(1 + |f(x_k)|)$.

The numerical results are shown in Table 1, where the test problems from [12] are numbered in the following way: " MGHi " means the i-th problem in [12]. In addition, "Dim" denotes the dimension of the problem, and IT, FE are number of iterations and number of function evaluations respectively. The number of gradient evaluations is equal to that of iterations since no gradient evaluation is required in the line search procedure.

From Table 1, we see that the our line search algorithms and the algorithm in [6] require the same number of iterations and the same number of function evaluations for some problems, whereas for some other problems, the Algorithm 2 be implemented with (16) (when $\eta_k \equiv 0.85$) performs better than with (20) (when $M = 5$ and $M = 10$). The gains are sometimes significant, for example, for MGH19, MGH22 and so on. However, our numerical performance is not better than those of [19] although our method has good theoretical performance. Hence how to improve the method to enjoy both good theoretical and numerical performance deserves further study.

5. Conclusion

In this paper, we consider a general form of the nonmonotone line search technique and some well known nonmonotone line search techniques can be seen as the special case of our technique. But compared with other nonmonotone technique not only independence on the nonmonotone parameter but also remove some restricted condition for the search direction. The numerical tests with a modified BFGS method show the efficiency of the proposed method. How to extend this technique to trust region method deserves further discussing, we leave it as a future work.

Table 1

Numerical comparisons

Problem name	Dim	IT/FE (M=5)	IT/FE (M=10)	IT/FE ($\eta_k \equiv 0.85$)	IT/FE ([19])
MGH2	2	40/51	40/51	40/51	11/31
MGH5	2	19/26	19/26	19/26	19/43
MGH7	3	50/72	50/72	50/72	42/116
MGH8	3	21/31	21/31	21/31	36/110
MGH9	3	6/9	6/9	6/9	5/11
MGH12	3	51/61	51/61	51/61	31/69
MGH13	4	55/78	55/78	53/76	40/99
MGH14	4	92/131	92/131	92/131	32/97
MGH15	4	30/33	30/33	30/33	30/61
MGH16	4	98/149	98/149	98/149	24/100
MGH18	6	23/27	23/27	23/27	54/111
MGH19	11	312/350	309/343	143/168	$>$1000/$>$1000
MGH20	6	79/101	79/101	70/94	36/91
MGH21	8	112/154	98/141	95/136	104/245
	16	156/251	184/282	158/249	162/409
	32	247/445	248/462	221/413	301/800
	64	321/751	339/785	312/714	355/1124
	128	487/1322	487/1387	474/1269	494/1902
	256	714/2534	776/2718	719/2295	750/3448
MGH22	8	134/174	143/183	106/143	58/145
MGH25	9	30/51	30/51	30/51	14/47
MGH26	10	26/27	26/27	26/27	$>$1000/$>$1000
MGH30	4	24/38	24/38	23/37	20/52
	6	28/53	28/53	20/47	24/77

References

[1] L. Armijo, Minimization of function having Lipschitz-continuous first partial derivatives, *Pacific J. Math.* 16 (1996) 1–3.

[2] E.G. Birgin, J.M. Martínez , M. Raydan, Nonmonotone spectral projected gradient methods for convex sets, *SIAM J. Optim.* 10 (2000) 1196–1211.

[3] R.H.Byrd and J.Nocedal, A tool for the analysis of quasi-Newton method with application to unconstrained minimization, *SIAM J. Numer. Anal.* 26 (1989) 727-739.

[4] R.M. Chamberlain, M.J.D. Powell, C. Lemarechal, H.C. Pedersen, The watchdog technique for forcing convergence in algorithms for constrained optimization, *Math. Programming Stud.* 16 (1982), 1–17.

[5] Y. Dai, On the Nonmonotone Line Search, *J. Optim. Theory Appl.* 112 (2001) 315–330.

[6] L. Grippo, F. Lampariello, S. Lucidi, A nonmonotone line search technique for Newtons methods, *SIAM J. Numer. Anal.* 23 (1986) 707–716.

[7] J. Han, G. Liu, General form of stepsize selection rules of line search and relevant analysis of global convergence of BFGS algorithm, Acta Math. Appl. Sinica 18 (1995) 112–122.

[8] J. Han, G. Liu, Global convergence analysis of a new nonmonotone BFGS algorithm on convex objective functions, *Computational Optimization and Applications.* 7 (1997) 277–289.

[9] M. Hestenes, Conjugate direction methods in optimization, *Springer-Verlag, New York.* 1980.

[10] D. Li, M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, *J. Comput. Appl. Math.* 129 (2001) 15–35.

[11] G. Liu, J. Han, D. Sun, Global convergence of the BFGS algorithm with nonmonotone linesearch, *Optimization.* 34 (1995) 147–159.

[12] J.J. Moré, B.S. Garbow, K.E. Hillstrom, Testing unconstrained optimization software, *ACM Transactions on Mathematical Software.* 7 (1981) 17–41.

[13] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, *SIAM J. Optim.* 7 (1997) 26–33.

[14] W. Sun, J. Han, J. Sun, Global convergence of non-monotone descent methods for unconstrained optimization problems, *J. Comput. Appl. Math.* 146 (2002) 89–98.

[15] W. Sun, Y. Yuan, Optimization theory and methods, *Springer, New York.* 2006.

[16] P.L. Toint, An assessment of nonmonotone line search technique for unconstrained optimization, *SIAM J. on Scientific Comput.* 17 (1996) 725–739.

[17] H. Yin, D. Du, The global convergence of self-scaling BFGS algorithm with nonmonotone line search for unconstrained nonconvex optimization problems, *Acta Mathematica Sinica.* 23 (2007) 1233–1240.

[18] Z. Yu, D. Pu, A new nonmonotone line search technique for unconstrained optimization, *J. Comput. Appl. Math.* 219 (2007), 134-144.

[19] H. Zhang, W.W. Hager, A nonmonotone line search technique and its application to unconstrained optimization, *SIAM J. Optim.* 14 (2004) 1043–1056.

Received 1 June 08; revised 2 Dec 08; accepted 17 Dec 08