DOI: 10.21276/sb.2016.2.10.6

Scholars Bulletin
(A Multidisciplinary Journal)
An Official Publication of "Scholars Middle East Publishers",
Dubai, United Arab Emirates
Website: http://scholarsbulletin.com/

ISSN 2412-9771 (Print)
ISSN 2412-897X (Online)

A non-monotone Wolf-Type line search strategy for unconstrained optimization

Yuzhu Li, Qinghua Zhou*
College of Mathematics and Information Science, Hebei University, Baoding, China

***Corresponding Author:**
Qinghua Zhou
Email: zhouqinghua@hbu.edu.cn

Abstract: In this work, we propose and analyze a new line search method for solving unconstrained optimization problems. Actually, we combine a non-monotone strategy into a modified Wolf rule and design a new algorithm that possibly chooses a larger step length. Generally, the global convergence is analyzed under some suitable conditions.
Keywords: unconstrained optimization; line search method; non-monotone strategy; global convergence.

INTRODUCTION

Consider the following unconstrained optimization problem:

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

where $f : R^n \to R$ is continuously differentiable function. Through out this paper, we use the following notation:

- $\|\bullet\|$ is the Euclidean norm.
- $f_k = f(x_k)$, $g_k = \nabla f(x_k)$, $G_k = \nabla^2 f(x_k)$ and B_k be a symmetric approximation of G_k.

Traditional iterative methods for solving (1) are either line search method or trust region method (see [1]. Line search methods proceed as follows: given a point x_k, find a descent direction d_k such that $d_k^T \nabla f(x_k) < 0$, a suitable step length α_k and construct the new point as follows:

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

In order to find a suitable step length α_k, we must solve the following one-dimensional minimization problem

$$\min_{\alpha\ge 0} \phi(\alpha) = f(x_k + \alpha d_k). \tag{3}$$

Some conditions proposed for the acceptance of α_k, where the monotone Wolfe-rule is stated as follows:

$$f(x_k + \alpha_k d_k) \le f(x_k) + \sigma \alpha_k \nabla f(x_k)^T d_k, \tag{4}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \delta g_k^T d_k. \tag{5}$$

And $\sigma \in (0,1/2)$, $\delta \in (0,1/2)$.

In [2] Grippo *et al*. presented the famous non-monotone line search. They defined the acceptance rule for α_k as follows:

$$f(x_k + \alpha_k d_k) \le f_{l(k)} + \sigma \alpha_k g_k^T d_k, \tag{6}$$

and

$$f_{l(k)} = \max_{0\le j\le m(k)} \{f_{k-j}\}, \quad k = 0,1,2,\cdots, \tag{7}$$

where $N \geq 0$ is an integer constant, $m(0)=0$ and for all $k \geq 1$. We have $0 \leq m(k) \leq \min\{m(k-1)+1, N\}$. Although this non-monotone technique has many advantages, it contains some drawbacks [3-5]. Therefore, there are also some proposals to overcome these disadvantages, see [6-10]. Among these works, the authors recommended several formulate to substitute for $f_{l(k)}$ appeared in (6). And the experiments proved the effectiveness of those strategies.

The rest of this article is organized as follows. In Section 2, we introduce a new wolf-type line search condition. In Section 3, we describe a new non-monotone line search algorithm and present its properties. The global convergence of the proposed algorithm is proved in Section 4. Finally, we draw some conclusions in Section 5.

New Wolfe-Type Line Search Condition

In this part, followed the idea of [6], where the author introduced a new kind of non-monotone Armijo line search. We revise it into Wolf-Type ones.

Actually, we choose the automatically adjustable initial step length s_k as follows:

$$s_k = -\frac{g_k{}^T d_k}{d_k{}^T B_K d_k}. \tag{8}$$

Like in [4], if $d_k{}^T B_k d_k \leq 0$, then we set $B_k = B_k + iI$ where i is the smallest non-negative integer such that

$$i > -\frac{d_k{}^T B_k d_k}{\|d_k\|^2}.$$

We also choose the step length α_k to be the largest α in $\{s_k, \rho s_k, \rho^2 s_k, \cdots\}$ satisfying the next modified non-monotone Wolfe-type conditions:

$$f\left(x_k + \alpha_k d_k\right) \leq R_k + \sigma\alpha_k\left[g_k{}^T + \gamma\|g_k\|^2\right], \tag{9}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \delta g_k{}^T d_k. \tag{10}$$

where $\gamma > 0$ is a constant, and

$$R_k = \eta_k f_{l(k)} + (1-\eta_k) f_k, \tag{11}$$

where $\eta_{\min} \in [0,1)$, $\eta_{\max} \in [\eta_{\min}, 1)$ and $\eta_k \in [\eta_{\min}, \eta_{\max}]$.

New Non-monotone Line Search Algorithm

We firstly state some basic and standard assumptions for the convenience of analyzing the convergence properties of the method which would be introduced in later part of this section.

(H1) The level set $L(x_0) = \{x \in R^n \mid f(x) \leq f(x_0), x_0 \in R^n\}$ is bounded.

(H2) There exist constants $0 < m \leq M$ such that for all k

$$m\|d_k\|^2 \leq d_k{}^T B_k d_k \leq M\|d_k\|^2. \tag{12}$$

(H3) The gradient $g(x)$ of $f(x)$ is Lipschitz continuous over an open convex set S that contains $L(x_0)$, i.e., there exists a positive constant L such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \text{ for all } x, y \in S$$

It follows from (H3) that there exists a constant $\gamma_1 > 0$ such that

$$\|g(x)\| \leq \gamma_1 \quad \forall x \in S. \tag{13}$$

Furthermore, due to guarantee the global convergence of the iterative schema (2), we need that a direction d_k satisfies the following sufficient descent conditions

$$g_k^T d_k \le -c_1 \|g_k\|^2 \tag{14}$$

and

$$\|d_k\| \le c_2 \|g_k\|. \tag{15}$$

where c_1 and c_2 are two positive real-valued constants.

The new algorithm can be described as follows:

Algorithm 3.1 (A new non-monotone algorithm)

Step 1 An initial point $x_0 \in R^n$ and a symmetric positive matrix $B_0 \in R^{n\times n}$ are given. The constants $0 < \rho < 1$, $0 < \sigma < 1$, $\gamma > 0$, $0 < s < \gamma_2$, $\varepsilon > 0$, $0 < \delta < 1$, $0 \le \eta_{\min} \le \eta_0 \le \eta_{\max} < 1$ are also given. Compute $R_0 = f(x_0)$ and set $k = 0$.

Step 2 Compute g_k. If $\|g_k\| \le \varepsilon$, then stop.

Step 3 Choose the descent direction d_k satisfying in (14) and (15). Then find the step length $\alpha_k \in \{s_k, \rho s_k, \rho^2 s_k, \cdots\}$ that is the largest among these number satisfying the conditions (9) and (10). Set $x_{k+1} = x_k + \alpha_k d_k$.

Step 4 Compute R_k by (11), update B_{k+1} by a quasi-Newton formula and choose $\eta_{\min} \le \eta_k \le \eta_{\max}$. Set $k = k+1$ and go to Step2.

Lemma 3.1 Suppose that the condition (14) holds and the sequence $\{x_k\}$ is generated by Algorithm 3.1. Then the sequence $\{f_{l(k)}\}$ is non-increasing.

Proof. Using the definition R_k and $f_{l(k)}$, we have

$$R_k = \eta_k f_{l(k)} + (1-\eta_k) f_k \le \eta_k f_{l(k)} + (1-\eta_k) f_{l(k)} = f_{l(k)}. \tag{16}$$

This leads to

$$f(x_k + \alpha_k d_k) \le R_k + \sigma\alpha_k \left[g_k^T d_k + \gamma \|g_k\|^2 \right] \le f_{l(k)+}\sigma\alpha_k \left[g_k^T d_k + \gamma \|g_k\|^2 \right]. \tag{17}$$

The preceding inequality and the descent condition $g_k^T d_k < 0$ indicate that

$$f_{k+1} \le f_{l(k)}. \tag{18}$$

On the other hand, from (7), we get

$$f_{l(k+1)} = \max_{0\le j\le m(k+1)} \{ f_{k+1-j} \} \le \max_{0\le j\le m(k)+1} \{ f_{k+1-j} \} = \max\{ f_{l(k)}, f_{k+1} \}.$$

This fact together with (18) complete the proof.

Lemma 3.2(See Lemma5 in [11] Suppose that the sequence $\{x_k\}$ is generated by Algorithm 3.1, then we have

$$f_{k+1} \le R_{k+1}, \quad \forall k \in N. \tag{19}$$

Furthermore, if (H1) holds, then the sequence $\{x_k\}$ is contained in $L(x_0)$.

Corollary 3.1 (See Corollary6 in [11]) Suppose that (H1) holds and the sequence $\{x_k\}$ be generated by Algorithm 3.1. Then the sequence $\{f_{l(k)}\}$ is convergent.

Lemma 3.3 Suppose that the sequence $\{x_k\}$ is generated by Algorithm 3.1. If $\tilde{\alpha}$ and α are step lengths which satisfy the standard Armijo rule and Algorithm 3.1, respectively, then $\tilde{\alpha} \le \alpha$.

Proof. If $\tilde{\alpha}$ and α are the step lengths which satisfy the standard Armijo rule and the new Wolf-type line search method (inequality (9)), respectively, then we have

$$f\left(x_k+\tilde{\alpha}_k d_k\right)-R_k \le f\left(x_k+\tilde{\alpha}_k d_k\right)-f_k \le \sigma\tilde{\alpha}_k\left[g_k{}^T d_k+\gamma\|g_k\|^2\right].$$

This implies that $\tilde{\alpha} \le \alpha$. From (15) and (13), we have

$$\|d_k\| < \infty \quad \forall k \in N \cup \{0\}. \tag{20}$$

Now, by Taylor's theorem, (9) and (19) we obtain

$$\begin{aligned}
&\lim_{\alpha\to0^+}\frac{R_k-f\left(x_k+\alpha d_k\right)+\sigma\alpha\left[g_k{}^T d_k+\gamma\|g_k\|^2\right]}{\alpha}\\
&\ge \lim_{\alpha\to0^+}\frac{f_k-f\left(x_k+\alpha d_k\right)+\sigma\alpha\left[g_k{}^T d_k+\gamma\|g_k\|^2\right]}{\alpha}\\
&= \lim_{\alpha\to0^+}\frac{f_k-f\left(x_k+\alpha g_k{}^T d_k+\circ\left(\alpha\|d_k\|\right)\right)+\sigma\alpha\left[g_k{}^T d_k+\gamma\|g_k\|^2\right]}{\alpha}\\
&=\left[-\left(1-\sigma\right)g_k{}^T d_k+\gamma\sigma\|g_k\|^2\right]>0.
\end{aligned}$$

So, there exists an $\hat{\alpha}_k > 0$ such that

$$f\left(x_k+\alpha d_k\right) \le R_k+\sigma\alpha\left[g_k{}^T d_k+\gamma\|g_k\|^2\right] \quad \forall \alpha \in [0,\hat{\alpha}_k].$$

Therefore, the new non-monotone line search is well-defined.

Global Convergence Analysis

In this section, we discuss some convergence properties of the proposed Wolfe-type line search algorithm and prove the global convergence to first-order critical points under some suitable conditions.

For the convenience of expression, we let $K_1=\{k\in K|\alpha_k=s_k\}$ and $K_2=\{k\in K|\alpha_k<s_k\}$. Obviously, $K=K_1\cup K_2$ is an infinite subset of the set $\{0,1,2,\cdots\}$.

Lemma 4.1 Suppose that (H1) and (H2) hold, the direction d_k satisfies (14) and (15) and the sequence $\{x_k\}$ be generated by Algorithm 3.1. Then we have

$$\lim_{k\to\infty} f_{l(k)} = \lim_{k\to\infty} f\left(x_k\right). \tag{21}$$

Proof. From (7), (9) and (16), for $k > N$, we obtain

$$\begin{aligned}
f\left(x_{l(k)}\right) &= f\left(x_{l(k)-1}+\alpha_{l(k)-1}d_{l(k)-1}\right)\\
&\le R_{l(k)-1}+\sigma\alpha_{l(k)-1}\left[g_{l(k)-1}^T d_{l(k)-1}+\gamma\left\|g_{l(k)-1}\right\|^2\right]\\
&\le f\left(x_{l(l(k)-1)}\right)+\sigma\alpha_{l(k)-1}\left[g_{l(k)-1}^T d_{l(k)-1}+\gamma\left\|g_{l(k)-1}\right\|^2\right].
\end{aligned}$$

The preceding inequality together with Corollary 3.1, $\alpha_k > 0$ and $g_k^T d_k < 0$ imply that

$$\lim_{k\to\infty}\alpha_{l(k)-1}\left[g_{l(k)-1}^T d_{l(k)-1}+\gamma\left\|g_{l(k)-1}\right\|^2\right]=0. \tag{22}$$

Using (14) and (15), we have

$$\alpha_k\left[g_k^T d_k+\gamma\|g_k\|^2\right] \le -c_1\alpha_k\|g_k\|^2+\alpha_k\gamma\|g_k\|^2 \le -\left(\left(c_1-\gamma\right)/c_2{}^2\right)\alpha_k\|d_k\|^2$$

for all k. This fact along with $\alpha_k < \gamma_2$ and (22) suggest that

$$\lim_{k\to\infty}\alpha_{l(k)-1}\left\|d_{l(k)-1}\right\|=0. \tag{23}$$

We now prove that $\lim_{k\to\infty}\alpha_k\|d_k\|=0$. Let $\hat{l}_k=l(k+N+2)$. Firstly, we show that, for any $j\geq 1$, we have

$$\lim_{k\to\infty}\alpha_{\hat{l}(k)-j}\left\|d_{\hat{l}(k)-j}\right\|=0. \tag{24}$$

and

$$\lim_{k\to\infty} f\left(x_{\hat{l}(k)-j}\right)=\lim_{k\to\infty} f\left(x_{l(k)}\right). \tag{25}$$

If $j=1$, since $\{\hat{l}_k\}\subseteq\{l(k)\}$, the relation (24) directly follows from (23). The condition (24) indicates that $\left\|x_{\hat{l}(k)}-x_{\hat{l}(k)-1}\right\|\to 0$.This fact along with the fact that $f(x)$ is uniformly continuous on L_0 imply that (25) holds, for $j=1$. Now, we assume that (24) and (25) hold, for a given j. Then, using (9) and (16), we obtain

$$f\left(x_{\hat{l}(k)-j}\right)\leq R_{\hat{l}(k)-j-1}+\sigma\alpha_{\hat{l}(k)-j-1}\left[g^T_{\hat{l}(k)-j-1}d_{\hat{l}(k)-j-1}+\gamma\left\|g_{\hat{l}(k)-j-1}\right\|^2\right]$$
$$\leq f\left(x_{l(\hat{l}(k)-j-1)}\right)+\sigma\alpha_{\hat{l}(k)-j-1}\left[g^T_{\hat{l}(k)-j-1}d_{\hat{l}(k)-j-1}+\gamma\left\|g_{\hat{l}(k)-j-1}\right\|^2\right].$$

Following the same arguments employed for deriving (23), we deduce

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

This means that

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

This fact together with uniformly continuous property of $f(x)$ on $L(x_0)$ and (25) indicate that

$$\lim_{k\to\infty} f\left(x_{\hat{l}(k)-(j+1)}\right)=\lim_{k\to\infty} f\left(x_{\hat{l}(k)-j}\right)=\lim_{k\to\infty} f\left(x_{l(k)}\right). \tag{26}$$

Thus, we conclude that (24) and (25) hold for any $j\geq 1$.

The rest of the proof is similar to lemma2 in [12].

Corollary 4.1 Suppose that (H1) and (H2) hold, d_k satisfies (14) and (15) and the sequence $\{x_k\}$ be generated by Algorithm 3.1. From (16) and (19) and Lemma 4.1,then we have

$$\lim_{k\to\infty} R_k=\lim_{k\to\infty} f(x_k). \tag{27}$$

Lemma 4.2 Assume that Algorithm 3.1 generates an infinite sequence $\{x_k\}$ and (H3) hold, then for all $k\in K_2$, there exists a constant $\overline{\alpha}>0$ such that $\alpha_k>\overline{\alpha}$.

Proof. From (10) and (H3), we have

$$L\alpha_k\|d_k\|^2\geq\left(g(x_k+\alpha_k d_k)-g_k\right)^T d_k\geq(\delta-1)g_k^T d_k>0. \tag{28}$$

Thus, we can conclude that

$$\alpha_k\geq\frac{(\delta-1)g_k^T d_k}{L\|d_k\|^2}=\frac{(\delta-1)\left|g_k^T d_k\right|}{L\|d_k\|^2}. \tag{29}$$

This inequality together with (14) and (15) lead us to have

$$\alpha_k\geq\frac{(1-\delta)c_1\|g_k\|^2}{Lc_2{}^2\|g_k\|^2}=\frac{c_1(1-\delta)}{Lc_2{}^2}. \tag{30}$$

Let $\overline{\alpha}=\dfrac{c_1(1-\delta)}{Lc_2{}^2}$, we complete the proof.

Lemma 4.3 Assume that (H2) holds, $0 < \gamma < c_1$, and the sequence $\{x_k\}$ is generated by Algorithm 3.1, then there exists a positive constant β such that

$$R_k - f_{k+1} \geq \beta \|g_k\|^2 \quad \forall k \in N. \tag{31}$$

Proof. If $k \in K_1$, then (12) and (14) together with (8), (9) and (15) imply that

$$\begin{aligned} R_k - f_{k+1} &\geq -\sigma\alpha_k \left[g_k^T d_k + \gamma \|g_k\|^2 \right] \\ &= \sigma \frac{-g_k^T d_k}{d_k^T B_k d_k} \left[-g_k^T d_k - \gamma \|g_k\|^2 \right] \\ &\geq \frac{c_1 \sigma \|g_k\|^2}{M \|d_k\|^2} \left[c_1 \|g_k\|^2 - \gamma \|g_k\|^2 \right] \\ &\geq \frac{c_1 \sigma}{c_2{}^2 M} (c_1 - \gamma) \|g_k\|^2 = \beta_1 \|g_k\|^2 \quad \forall k \in K_1, \end{aligned} \tag{32}$$

Where $\beta_1 = \frac{c_1 \sigma}{c_2{}^2 M}(c_1 - \gamma)$.

On the other hand, if $k \in K_2$, we have $\alpha_k < s_k$. We define $\hat{\alpha}_k = \alpha_k / \rho$, so, by definition of the new Wolfe-type condition, we get

$$L\hat{\alpha}_k \|d_k\|^2 \geq \left(g(x_k + \hat{\alpha}_k d_k) - g_k\right)^T d_k \geq (\delta - 1) g_k^T d_k > 0.$$

If setting $k = \frac{\rho(1-\delta)}{L}$, then we have

$$\alpha_k = -k \frac{g_k^T d_k}{\|d_k\|^2} \quad \forall k \in K_2. \tag{33}$$

Using (12) and (14) together with (9), (15) and (33), we obtain

$$\begin{aligned} R_k - f_{k+1} &\geq -\sigma\alpha_k \left[g_k^T d_k + \gamma \|g_k\|^2 \right] \\ &= -k\sigma \frac{g_k^T d_k}{\|d_k\|^2} \left[-g_k^T d_k - \gamma \|g_k\|^2 \right] \\ &\geq c_1 k\sigma \frac{\|g_k\|^2}{\|d_k\|^2} \left[c_1 \|g_k\|^2 - \gamma \|g_k\|^2 \right] \\ &\geq \frac{c_1 k\sigma}{c_2{}^2} (c_1 - \gamma) \|g_k\|^2 = \beta_2 \|g_k\|^2 \quad \forall k \in K_2, \end{aligned} \tag{34}$$

Where $\beta_2 = \frac{c_1 k \sigma}{c_2{}^2}(c_1 - \gamma)$. Now, let

$$\beta = \max\{\beta_1, \beta_2\}. \tag{35}$$

Then, (32), (34) together with (35) conclude the result.

Theorem 4.1 Suppose that (H1) and (H2) hold, and the sequence $\{x_k\}$ is generated by Algorithm 3.1. Then we have

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{36}$$

Proof. Using (31), we know that there is a positive constant such that

$$R_k - f_{k+1} \geq \beta \|g_k\|^2, \quad \forall k \in N.$$

Then from Corollary 4.1, we can complete the proof.

CONCLUSION

In this work, we introduce a new non-monotone Wolfe-type line search strategy for unconstrained optimization problems. Compared with (4), we relax the right-hand side of it, so the new rule accepts a larger step-size especially whenever iterates are far from the optimum. This can decrease the number of iterations and function evaluations. On the other hand, compared with (6), it is obviously that we fully employ the current objective function value f_k, it can improve the efficiency of the algorithm. We analyzed the properties of the algorithm and proved the global convergence theory under some suitable conditions.

Acknowledgement

The author would be very grateful to the anonymous referees for their valuable comments and suggestions.

REFERENCES

1. Nocedal, Jorge, and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.
2. Grippo, L., Lampariello, F., & Lucidi, S. (1986). A nonmonotone line search technique for Newton's method. *SIAM Journal on Numerical Analysis*, *23*(4), 707-716.
3. Dai, Y. H. (2002). On the nonmonotone line search. *Journal of Optimization Theory and Applications*, *112*(2), 315-330.
4. Shi, Z. J., & Shen, J. (2006). Convergence of nonmonotone line search method. *Journal of Computational and Applied Mathematics*, *193*(2), 397-412.
5. Zhang, H., & Hager, W. W. (2004). A nonmonotone line search technique and its application to unconstrained optimization. *SIAM journal on Optimization*, *14*(4), 1043-1056.
6. Ahookhosh, M., Amini, K., & Bahrami, S. (2012). A class of nonmonotone Armijo-type line search method for unconstrained optimization. *Optimization*, *61*(4), 387-404.
7. Hu, S. L., Huang, Z. H., & Lu, N. (2010). A non-monotone line search algorithm for unconstrained optimization. *Journal of Scientific Computing*, *42*(1), 38-53.
8. Shi, Z., & Wang, S. (2011). Modified nonmonotone Armijo line search for descent method. *Numerical Algorithms*, *57*(1), 1-25.
9. Toint, P. L. (1996). An assessment of nonmonotone linesearch techniques for unconstrained optimization. *SIAM Journal on Scientific Computing*, *17*(3), 725-739.
10. Zhang, H., & Hager, W. W. (2004). A nonmonotone line search technique and its application to unconstrained optimization. *SIAM journal on Optimization*, *14*(4), 1043-1056.
11. Ahookhosh, M., & Amini, K. (2012). An efficient nonmonotone trust-region method for unconstrained optimization. *Numerical Algorithms*, *59*(4), 523-540.
12. Amini, K., Ahookhosh, M., & Nosratipour, H. (2014). An inexact line search approach using modified nonmonotone strategy for unconstrained optimization. *Numerical Algorithms*, *66*(1), 49-78.